Embedding Conversational Agents into AR: Invisible or with a Realistic Human Body?

Jens Reinhardt
Hamburg University of Applied Sciences
Hamburg, Germany
jens.reinhardt@haw-hamburg.de

Luca Hillen
Hamburg University of Applied Sciences
Hamburg, Germany
luca.hillen@haw-hamburg.de

Katrin Wolf
Hamburg University of Applied Sciences
Hamburg, Germany
katrin.wolf@haw-hamburg.de

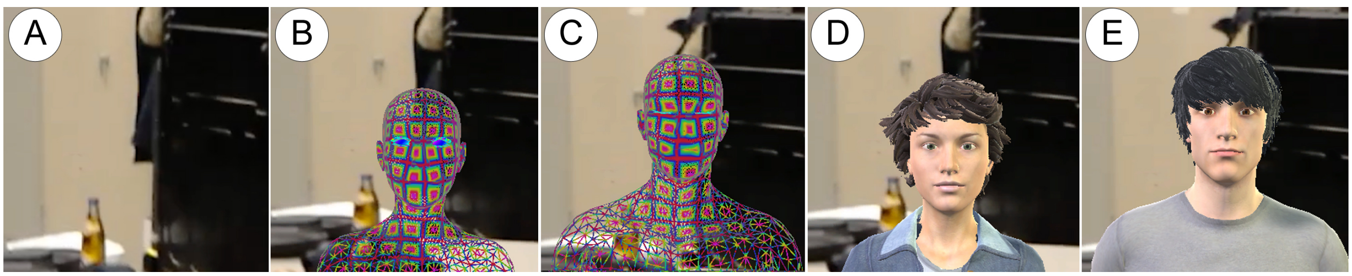

Figure 1: Realism continuum - invisible agent (A), simplified wireframe humanoid agent (female (B), male (C)), fully textured and rich detailed human agent (female (D), male(E))

ABSTRACT

Currently, (invisible) smart speech assistants, such as Siri, Alexa, and Cortana, are used by a constantly growing number of people. Moreover, Augmented Reality (AR) glasses are predicted to become widespread consumer devices in the future. Hence, smart assistants can easily become common applications of AR glasses, which allows for giving the assistant a visual representation as an embodied agent. While previous research on embodied agents found a user preference for a humanoid appearance, research on the uncanny valley suggests that simply designed humanoids can be favored over hyper-realistic humanoid characters. In a user study, we compared agents of simple versus more realistic appearance (seen through AR glasses) versus an invisible state-of-the-art speech assistants (see Figure 1). Our results indicate that a more realistic visualization is preferred as it provides additional communication cues, such as eye contact and gaze, which seem to be key features when talking to a smart assistant. But if the situation requires visual attention, e.g., when being mobile or in a multitask situation, an invisible agent can be more appropriate as they do not distract the visual focus, which can be essential during AR experiences.

TEI '20, February 9–12, 2020, Sydney, NSW, Australia

ACM ISBN 978-1-4503-6107-1/20/02…$15.00
https://doi.org/10.1145/3374920.3374956

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Natural language interfaces**; **Empirical studies in HCI**.

KEYWORDS

embodied conversational agents, intelligent virtual assistants, augmented reality, avatars

ACM Reference Format:
Jens Reinhardt, Luca Hillen, and Katrin Wolf. 2020. Embedding Conversational Agents into AR: Invisible or with a Realistic Human Body?. In *Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '20), February 9–12, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3374920.3374956

1 INTRODUCTION

Intelligent Virtual Assistants are embedded in more and more smartphones, which so far, are relying on digital speech assistants, like *Siri* [1] or *Cortana* [2]. Moreover, ambient voice assistants in the form of smart speakers, such as *Amazon*

[1]https://www.apple.com/siri/
[2]https://www.microsoft.com/en-us/cortana

Echo [3] and *Google Home* [4], have been integrated into millions of homes.

Such assistants are predicted to become more and more widespread as according to Forrester Research, half of American households are expected to have an intelligent loudspeaker device by 2022 [5]. Furthermore, Google claims to sell every second a Google Home device (31.5M devices per year) [6].

At the same time, Augmented Reality (AR) [3] and Mixed Reality (MR) [19, 20] devices in the form of headsets and smart glasses, e.g. *Microsoft Hololens* [7], *Everysight Raptor* [8], *Toshiba dynaEdge AR100 Viewer* [9], *ThirdEye Gen X1* [10], *Everysight Raptor* [11], or *Metavision Meta 2* [12], become both available and affordable. Some of these devices are already equipped with an agent, such as the *HoloLens* with *Cortana*. However, *HoloLens* is an AR device while *Cortana* currently lacks any kind of visual representation.

This paper is motivated by the belief that future AR glasses will, as soon as the computational power will allow for, not limit their smart assistants to be represented by voice only, but will also have a visual representation. As speech-based assistants are (in their auditory appearance) designed to be humanoid using human-like voices, future audio-visual assistants may also be designed as humanoids having a human voice as well as a human body.

Previous work has already shown that embodiment of virtual agents is beneficial [8, 11, 16, 39]. These embodied conversational agents are dialog systems that provide a visual representation of the agent in addition to the natural language interface. They have many of the same characteristics as humans in face-to-face conversations, including the ability to create and respond to verbal and non-verbal communication, such as representational gestures, gaze away and towards one other, and facial expressions [2, 9, 14, 15, 26, 35, 36]. The embodied conversational assistants not only imitate the human voice, but also the visual appearance of people and their behaviour in conversational situations. However, previous work found that a human visualisation of such agents is preferred [38] while earlier research on humanoid visualisation indicated that hyperrealism can cause issues known as Uncanny Valley [22]. Here, a realistic humanoid representation has been shown to possibly cause negative emotional reaction in many fields, such as robotics, animation characters, and digital avatars [5, 18, 29, 30, 34].

Consequently, while the change from a pure voice assistant to an embodied assistant in AR appears to be logical, it remains unclear which degree of realism visually embodies humanoid agents should have.

Research at Uncanny Valley suggests that the degree of realism of the agent's appearance can influence its perception. No previous work has compared the continuum of (ultramodern) voice-controlled conversation agents in Mixed Reality with different degrees of realism of the agent's visualization. This paper aims to reduce this research gap by conducting a user study that compares three different degrees of realism in the visualization of embodied conversational agents.

Our research question is: *How do different realism levels of visual representation influences the usability of embodied conversational agents on AR glasses?*

Results of a user study indicate that a more realistic looking agent leads to significant higher PRAGMATIC QUALITIES than simplified or (state-of-the-art) invisible one, and that the more realistic looking human agent was perceived significantly more ATTRACTIVE than the simplified one. Our findings contribute to the domain of agents design, and we hope to encourage researchers and practitioners to provide intelligent conversational agents with a virtual body to enrich the communication with smart assistants.

2 RELATED WORK

While reviewing previous research, we focus on three categories: the visual representation of virtual agents, the negative effect of hyper-realistic humanoids called the Uncanny Valley, and research around the perception of embodied agents in AR.

Visual Representation of Virtual Agents

A large body of research investigated what visualization type, when designing a virtual agent, is desired by the user interacting with it and how such appearance can foster social interaction, trust, and the belief in the intelligence of the virtual counterpart [8, 11, 16, 25, 39].

Parise et al. found that engagement and collaboration with anthropomorphic agents were significantly higher than with non-human like agents [25]. Furthermore, they found that human-like agents do not need to be photo-realistically represented to engage the user to collaborate with them. Yee et al. assimilated previous empirical studies which compare interfaces with visually embodied agents to interfaces without a visual representation of agents [39]. They found that human-like representations of higher realism produced more

[3]https://www.amazon.de/s?k=echo&i=amazon-devices&ref=nb_sb_noss_2
[4]https://store.google.com/us/product/google_home
[5]https://www.forrester.com/report/Forrester+Data+Smart+Home+Devices+Forecast+2017+To+2022+US/-/E-RES140374
[6]https://www.blog.google/products/assistant/how-google-home-and-google-assistant-helped-you-get-more-done-in-2017/
[7]https://www.microsoft.com/en-us/hololens
[8]https://everysight.com/
[9]http://www.toshiba.eu/generic/dynaedge/
[10]https://www.thirdeyegen.com/x2-smart-glasses-4/
[11]https://everysight.com/
[12]https://www.metavision.com

positive social interactions than representations of lower realism, while behavioral measurements revealed no significant difference between low and high realism representations. Cowell and Stanney investigated design strategies aiming at fostering reliable and trustable dialogues with virtual agents [8]. They found that users generally prefer to interact with a youthful character that matches their ethnicity. McBreen and Jack evaluated human-like and humanoid cartoon-like agents in e-retail applications as conversational sales assistants [16]. Their results showed that the participants expected a high degree of realistic human-like verbal and non-verbal communication behavior from human-like agents. These expectations were reflected in a strong preference for agents who showed natural facial expressions, gestures, and emotions. Further results showed that participants had a preference for 3D instead of 2D cartoon-like agents and a preference to interact with fully embodied agents. Geven et al. explored a system equipped with the cognitive vision that is able to see the user, perform user reactions and react to situations that occur in the environment and interact with the user in the context of personal assistance to help in problem-solving in an office [11]. A traditional GUI, a cartoon-like embodied agent, and a realistic human-like embodied agent were compared in two tasks. The results showed no significant differences between trust and intelligence ratings but showed that a more personal, embodied representation increased social presence, especially in extroverted individuals.

Zibrek et al. investigated whether different rendering styles would directly influence the appeal [40]. They found that this affinity to virtual characters is a complex interaction between the look and personality of the character. Forlizzi et al. [10] investigated the relationship between the visual characteristics of embodied agents and the tasks they carry out. Their results show a clear correlation between agent task and agent appearance and show that people often prefer agents that correspond to the gender stereotypes associated with tasks.

Mousas et al. investigated the influence of appearance and movement of virtual characters on the emotional reactions [23]. They found that the appearance and movement of the virtual characters significantly influenced both dimensions of the emotional response system, valence and reactivity. Previous work has also investigated which gender representation of digital agents is preferred by users. While some works showed a preference to female embodied agents [10, 24, 41], others have found no difference in preference [8, 14, 16] or a preference for male agents [17]. In summary, previous research on agent's appearance shows that users prefer a human visualization over no visualization and an abstract visualization, as well as 3D over 2D, rendered virtual bodies.

The Uncanny Valley

While learning from previous research that human visualizations are recommended for agents, prior research also indicates that hyperrealism in humanoid avatar, character, and robot design can cause eeriness, which is described as the Uncanny Valley [18, 29, 30, 34, 40]. Tinwell [34] explored the relationships between user satisfaction and perceived strangeness, as well as between user satisfaction and human appearance for virtual characters. McDonnell and Breidt studied the relationship between rendering style and perceived trust and provided guidelines for creating plausible virtual characters [18]. Thompson showed that hyper-realistic game characters easily fall into the Uncanny Valley and recommends to opt for a stylized rendering (such as cel-shading) to avoid the uncanny effect [33]. Schwind et al. investigated the negative effects of the Uncanny Valley on the facial characteristics of virtual human faces [29, 30]. They indicated that the most important factors to avoid negative feelings were smooth skin, natural skin color, and human proportions and faces whose characteristics differ from the human norm were perceived as negative.

In summary, prior work on the Uncanny Valley shows that a natural appearance is preferred for avatar and agents design, but the closer the character design gets to the reality as more likely tiny details can cause eeriness. While any visual cue that does not meet the expectations of human features we are used to can cause negative emotions, such as mistrust, a decrease of perceived attractiveness, and lack of sympathy, a common human-like character design leads to a higher degree of social interaction, social presence, and likability.

Agents in Mixed Reality

The works investigating effects on the visual appearance of agents and avatars, we discussed in the two prior sections were not implemented in Mixed Reality (MR). We assume that an agent or avatar that is perceived to be in the same room as the observer might cause different, maybe stronger, effects than an agent or avatar seen on a hand-held device. Hence, in this section, we focus on previous research that has focused on designing and developing agents for virtual and augmented reality.

Balcisoy et al. created one of the first examples of a virtual agent in AR that could play checkers with a person in the real world [4]. The agent had no conversational skills and was unable to respond to the voice and gesture commands of the real user. They found that the presence of the virtual agent in the same room as the real user creates a strong sense of presence. However, no user studies are mentioned that examined how the user felt about the agent. Wagner et al.

investigated virtual characters for AR applications in comparison to text, text and audio, 2D image, and 3D character (not registered in the real world) [37]. Their study shows that 3D virtual characters (AR and non-AR) are generally preferred over other representations. Miyake and Ito investigated a virtual conversational agent in AR controlled by voice and spoken commands [21]. They developed two AR agents that behave like avatars of objects that can be controlled through them. Their results showed that the existence of an agent increased the easiness to talk to the system. Moreover, they found that a system with an agent showed better response accuracy than a system without agents. Anabuki et al. developed *Welbo*, a virtual agent for MR [1]. Their research had a strong focus on the design and implementation of agents in MR. They found that spatial parameters of the virtual agent, including size and location, have a large impact on the perception of the agent. Wang et al. compared different forms of conversational agents for AR: voice-only, non-human, full-size embodied, and a miniature embodied agent, by using a hidden object game with support of the agent [38]. They found that users preferred a miniaturized embodied agent due to the novelty of the size of the agents and the reduced uncanniness compared to the full size embodied agent.

Summary

While previous research on agents in Mixed Reality shows the benefit of using embodied conversational agents to interact with, only Wang et al. investigated different visualizations of agents when the user and the agent share the same space [38]. Wang et al., similar to Parise et al., found that a humanoid visualization is preferred over abstract visualizations of objects [25, 38]. However, while research on the Uncanny Valley suggests that the degree of realism of the agent's appearance can affect their perception, no prior work compared the continuum from (state-of-the-art) voice-only conversational agents in Mixed Reality with different realism degrees of the agent's visualization. This paper aims at reducing that research gap by conducting a user study that compares the perception of agents with different realism degrees in AR.

3 EXPERIMENT

To better understand how the degree of realism of humanoid agents in AR is perceived, we compared in a controlled experiment three different level of realism of agents in AR. We aim at findings that scale across the task as well as across genders. As Forlizzi et al. suggested that the perception of the appearance of avatars depends on the task type [10], and others found effects of gender on user's preference, we varied in our experiment both, task type and gender to explore the influence of the realism degrees we compared [24, 41].

Design

Our study had a 3x3 within subjects design with the independent variables VISUALIZATION (realistic human, simplified human, invisible), and TASK (message, call, weather). The dependent variables were ATTRACTIVENESS, and USABILITY.

Measurements

ATTRACTIVENESS was measured by the AttrakDiff questionnaire [12]. Aiming for an acceptable experiment duration, we needed a short version of the AttrakDiff questionnaire. Hence, we applied the AttrakDiff mini [13] and used as last five items the scales unprofessional/professional, conventional/inventive, isolation/connecting, pleasant/unpleasant, cumbersome, straight from its original version [12] to fit the conversational application.

USABILITY was recorded using the System Usability Scale (SUS) questionnaire [6]. To better understand the quantitative data, we collected additional qualitative feedback in a semi-structured interviews asking:

- What about the just tested visualization type increased the usability of the conversation?
- What about the just tested visualization type decreased the usability of the conversation?

Participants

The experiment was conducted with 18 participants (9 females, 9 males) aged between 23 and 64 years and an average age of 32 years ($SD = 10.55$). Eleven (4 female) of the participants in the experiment already had experience with AR and 13 (4 female) had used a voice assistant before.

Apparatus

The apparatus was implemented in Unity3D [13] presenting three visualization of intelligent conversational agents in AR shown at the Microsoft HoloLens version 1. The interaction was speech-based allowing the user to solve two training tasks as well as three tasks of the experiment. As fully functioning conversational agent, the IBM Watson Assistant [14] was used. Watson Assistant allows for pure conversation based on text and acts on the basis of the IBM-developed agent. In order to respond to and reproduce human speech, additional services needed to be integrated, such as listening and speaking through further services through Text to Speech [15] (TTS) and Speech to Text [16] (STT). The standard voices of the TTS service of IBM were chosen as system output, and the voice gender was chosen according to the gender-specific visualization of the agent.

[13]https://unity.com
[14]https://www.ibm.com/cloud/watson-assistant/
[15]https://cloud.ibm.com/catalog/services/text-to-speech
[16]https://cloud.ibm.com/catalog/services/speech-to-text

To implement the services in Unity3D, the IBM Watson SDK for Unity [17] was used. The avatars were created with the help of Morph3D's [18] Morph Character System (MSC). The Unity Asset SALSA With RandomEyes [19] was integrated, which animated the lips and eyes of the virtual character. Since for one task, the current weather was queried, the seven-day forecast of the Weather Company Data Service [20] was implemented. Finally, Gaze, Input Manager, and Spatial Mapping were taken from the Mixed Reality Toolkit (MRTK) [21] to place and move the character in the environment.

Visual Appearance and Behavior

We used across each VISUALIZATION the female and male voice of the IBM TTS. The conditions REALISTIC HUMAN and SIMPLIFIED HUMAN of the independent variable VISUALIZATION were based on the female and male character models of MCS Female [22] and MCS Male [23], see Fig. 1. The avatars used had a young appearance and same ethnicity as participants, which has been recommended by McBreen and Jack [16]. For the condition SIMPLIFIED HUMAN, textures and shapes of the avatar model were created rendering only a wire-frame of the humanoid agent.

For the REALISTIC HUMAN, the eyes and lips were animated, and the eyes follow the position of the user. Lips were animated based on the audio signal of the agent's answers taking research in the field of audio-visual speech perception into account [28, 32]. To create a natural behavior expected in communication situations [7], the REALISTIC and the SIMPLIFIED HUMAN, constantly turned their head and body into the user's direction so that the agent always looks at them. The INVISIBLE agent reflected the current state-of-the-art of voice assistants. While Wang et al. [38] visualized a virtual box representing an Alexa speaker, the speaker in our AR setup was embedded into the Hololens. Hence, no visual representation was shown when interacting with the INVISIBLE agent.

Experimental Tasks

The apparatus had five tasks implemented, two training tasks, and three tasks that participants were asked to solve during the experiment.

[17]https://assetstore.unity.com/packages/tools/ai/ibm-watsonsdk-for-unity-108831
[18]https://www.morph3d.com/
[19]https://assetstore.unity.com/packages/tools/animation/salsa-with-randomeyes-16944
[20]https://cloud.ibm.com/catalog/services/weather-company-data
[21]https://github.com/Microsoft/MixedRealityToolkit-Unity
[22]https://www.assetstore.unity3d.com/en/#!/content/45807
[23]https://www.assetstore.unity3d.com/en/#!/content/45805

Figure 2: Visualization of the CALL task for realistic humanoid (left) and simplified human (right) agent

Training Task. To become familiar with the agent, the participant performed two training tasks before the experiment started. For the training tasks, we prepared two requests to the agent. To reduce erroneous input by vague pronunciations or slip of the tongue, the agent did not react to the completely spoken sentence but to the words marked in the descriptions of the tasks. During the first training task, the participant asked the agent *"What **time** is it?"*. If the agent understood the request correctly, it told the actual time. For the second training task, the participants had to ask the agent *"Can you tell me a **joke**?"*. If understood, the agent responded with *"Seriously, you want me to tell you a joke?"*. In case the participant said *"Yes"*, the agent told a joke. If a request was not understood, such as *"I didn't understand you, can you try it again?"* or *"Oh, I have no clue what you are talking about, sorry."*.

Call Task. For the experimental task CALL, the participant had to say *"Please, **call** name"*, and *name* was replaced by any self-chosen name. If the agent did not recognize the name, they asked *"Who do you want to call?"*. Then, the participant could tell the name again or choose another one. If the agent recognized the name, the system replied with *"Wait, I will call name for you"* and replaced *name* with the mentioned one. Then, the agent simulated a call, and the VISUALIZATION of the agent for REALISTIC and SIMPLIFIED HUMANOID took a cell phone to make the call (see Figure 2). The participants could hear how the other person's phone rings and after three rings a busy earcon was played. Finally, the agent ended the call and returned *"Maybe just busy"*. Then, the task was completed.

Message Task. During the SMS task, the participant had to say *"Send a **message** to mom"*. The system processed the instructions and answered *"What do you want to write to your mom?"*. The participant was free to chose any short message and tell it the system. The system again processed the message and responded with *"Did I understood this correctly, your message was ..."* - repeating the previously spoken message of the participant. The participants could confirm saying *"yes"*

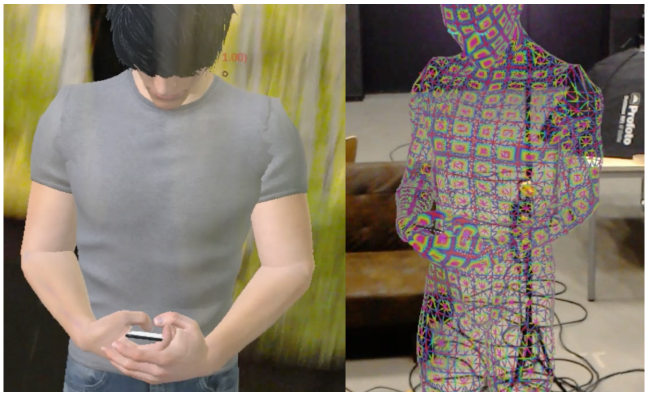

Figure 3: Visualization of SMS task for realistic human (left) and simplified (right) humanoid

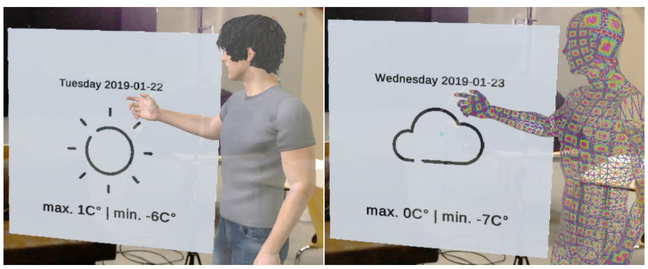

Figure 4: Visualisation of WEATHER task for the realistic (left) and simplified (right) agent

or decline with *"no"* to repeat the recording of the message. After confirmation, the REALISTIC and SIMPLIFIED HUMAN typed a messages into a smart phone (see Figure 3). After that animation, a typical sound of sending a message was played. Finally, the agent said *"OK, I sent it to your mom"*.

Weather Task. For the WEATHER task, the participant could choose between three requests. They had the opportunity to ask *"How is the **weather** on weekday / today / tomorrow?"*, *"Is it **sunny** on weekday / today / tomorrow?"*, or *"Is it **rainy** on weekday / today / tomorrow?"*. The phrases *"on weekday"*, *"today"*, or *"tomorrow"* could be selected by the participants themselves. The phrase *"on weekday"* had to be replaced by a freely chosen day of the week. The agent responded with a verbal explanation of the current local weather forecast, i.e. *"The weather on Wednesday in [city] at [time] is sunny. The lowest temperature is minus six to minus four Celsius. The highest is zero to one Celsius."*. Additionally, for the REALISTIC and SIMPLIFIED visualization, an info-graphics with name of city, weekday and date, a graphical representation of the weather, and also the minimum and maximum temperature was displayed (see Figure 4). The answering agent was animated and showed the spoken information of the weather forecast by a pointing at the info-graphics.

Procedure

After welcoming them, participants read the general instructions, signed a consent form, and filled in a demographic questionnaire. Here, we also collected information about their previous experience with speech and home assistants.

At the beginning of the experiment, the participants got an introduction about the HoloLens device. If there were no open questions, the participants equipped with the Hololens stood at a start position marked with a cross on the floor of the experimental room. This should ensure that all participants started at the same position and under same tracking conditions for each condition. Moreover, all participants were asked to look into the same direction. Then, the experimenter started the prototype via a Remote Desktop PC. Afterwards, the agent greeted the participant. Depending on the actual time, the agent said: *"Hi there, it is a beautiful morning / afternoon / evening. How can I help you?"*, or *"Good morning / afternoon / evening. How can I help you?"*.

To inform the participants about the task they should solve, a card with the exact wording of the command had been prepared for each task. Thus, nine cards were prepared for each participant plus a card for the training task, which always was solved first. The instruction cards for each task were handed to participants for each task. Participants were asked to read the text of the card aloud and clearly and to react on the prompts of the agent. While the agent spoke, participants were asked to be quiet. While participants read the commands, a text displayed what the agent had understood providing feedback about the system state.

To avoid sequence effects, the order of the conditions followed a Latin square design. Both appearances of the agent, the female and the male one, were shown in random order for each condition. Hence, in each condition a task was solved twice, once talking to a male and once to a female agent.

After finishing each condition, a pop-up window was displayed asking the participant to fill in the questionnaires (see Figure 5).

After participants completed one condition, participants removed the HoloLens in order to answer the semi-structured interview and fill in the individual questionnaires. After answering the questionnaires, the participant put on the HoloLens again and gets a new card with the next assigned task. The participant chooses with his eyes an "OK" button, which is confirmed by experimenter (see Figure 5). Then the pop-up disappears and the agent listens again. Now the process is repeated as described above for all other conditions of the experiment.

Each participant explored the agent nine times to fulfill all conditions of the study.

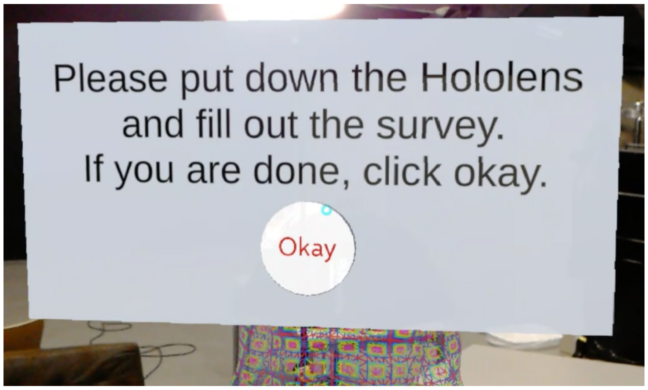

Figure 5: Pop-up window visualizing the end of each task

4 RESULTS

We analyzed the quantitative and qualitative data with the aim of (1) identifying significant effects in our quantitative data and (2) gaining a deeper understanding of the quantitative results through analyzing the qualitative feedback.

Quantitative Analyses

Independent Kruskal-Wallis H tests were used to indicate significant effects on the ordinal data attractiveness and usability and post-hoc analysis with Mann-Whitney U tests were conducted with a Bonferroni correction applied, resulting in a significance level set at .017.

Attractiveness. The three sub-scales of the AttrakDiff-Mini questionnaire, which describe usability, are pragmatic qualities (PQ), hedonic quality (HQ), and attractiveness (ATT). All three scales have as minimum 1 and as maximum 7, while high values refer to high qualities.

Descriptive statistics led to following PQ sub-scores: invisible$_{mean}$ = 4.7833 (SD = 1.0211), simplified $_{mean}$ = 4.4537 (SD = 1.3210), and realistic$_{mean}$ = 5.5093 (SD = 1.1295), and for task: call$_{mean}$ = 4.9167 (SD = 1.3441), SMS$_{mean}$ = 4.8056 (SD = 1.40137), and weather$_{mean}$ = 4.8963 (SD = 1.0968), for HQ: invisible$_{mean}$ = 4.6111 (SD = 1.2635), simplified $_{mean}$ = 4.9136 (SD = 1.2203), and realistic$_{mean}$ = 5.1420 (SD = 1.1485), and for task: call$_{mean}$ = 4.9259 (SD = 1.2263), SMS$_{mean}$ = 4.7530 (SD = 1.3130), and weather$_{mean}$ = 4.9877 (SD = 1.1363), and for ATT: invisible$_{mean}$ = 4.8491 (SD = 1.4211), simplified $_{mean}$ = 4.1572 (SD = 1.8369), and realistic$_{mean}$ = 5.5577 (SD = 1.1785), and for task: call$_{mean}$ = 4.8951 (SD = 1.6759), SMS$_{mean}$ = 5.0185 (SD = 1.5906), and weather$_{mean}$ = 4.7222 (SD = 1.5551) (see Figure 6).

While Kruskal-Wallis H test did not indicate a statistically significant difference in the HQ score ratings between the different visualizations (χ^2 = 2.2157, p = .3303), a Kruskal-Wallis H test showed a statistically significant difference for visualizations in the PQ score ratings (χ^2 = 13.2465, p = .0013) as well as in the ratings of the ATT score (χ^2 = 8.209, p = .0165).

Mann-Whitney U tests showed for the independent variable visualization that a realistic human caused in significantly higher PQ score than a simplified one (U = 54.0, z = −3.4201, p = .0006) and also than the invisible visualization (U = 79.5, z = −2.6122, p = .009), while no significant difference regarding their PQ score could be found between the simplified human and the invisible visualization (U = 127.500, z = −1.0936, p = .2741).

Mann-Whitney U tests also showed that a realistic human caused in a significantly higher ATT score than a simplified visualization (U = 75.5.0, z = −2.7396, p = .0062), while neither a significant difference could be found between the realistic and the invisible version (U = 99.0, z = −2.6122, p = .0461) nor between a simplified and invisible one (U = 135.0, z = −0.855, p = .3926).

Independent Kruskal-Wallis H tests did not indicate a statistically significant difference between the different tasks in any sub-scale (PQ: χ^2 = .01252, p = .9938, HQ: χ^2 = .7433, p = .6896, ATT: χ^2 = .807, p = .668).

Usability. Descriptive statistics led to following SUS scores for visualization: invisible$_{mean}$ = 76.574 (SD = 17.083), simplified$_{mean}$ = 71.944 (SD = 19.779), and realistic$_{mean}$ = 77.731 (SD = 17.343), and for task: call$_{mean}$ = 74.907 (SD = 17.81), SMS$_{mean}$ = 73.935 (SD = 20.215), and weather$_{mean}$ = 77.407 (SD = 16.447) (see Figure 7).

Kruskal-Wallis H tests did neither show any significant difference between the visualizations regarding their usability ratings (χ^2 = 1.610, p = .4471) nor between the three different tasks (χ^2 = .971, p = .6153).

Qualitative Analyses

The qualitative data collected during semi-structured interviews was analyzed through closed coding. The categories were structured according the significant results identified through quantitative analyses to find explanations for the identified effects.

What increased the pragmatic quality of our realistic human visualization? Participants named for their positive responses to the realistic human visualization the following pragmatic reasons: authentic and nice appearance, the desire to talk to a (simulated) human being, and the benefit of visual feedback through facial expressions, especially gaze.

- *"Having a humanoid talking about human related topics like weather made it more reliable that they behaved and kind of looked like humans as well."* (P1, task: weather),

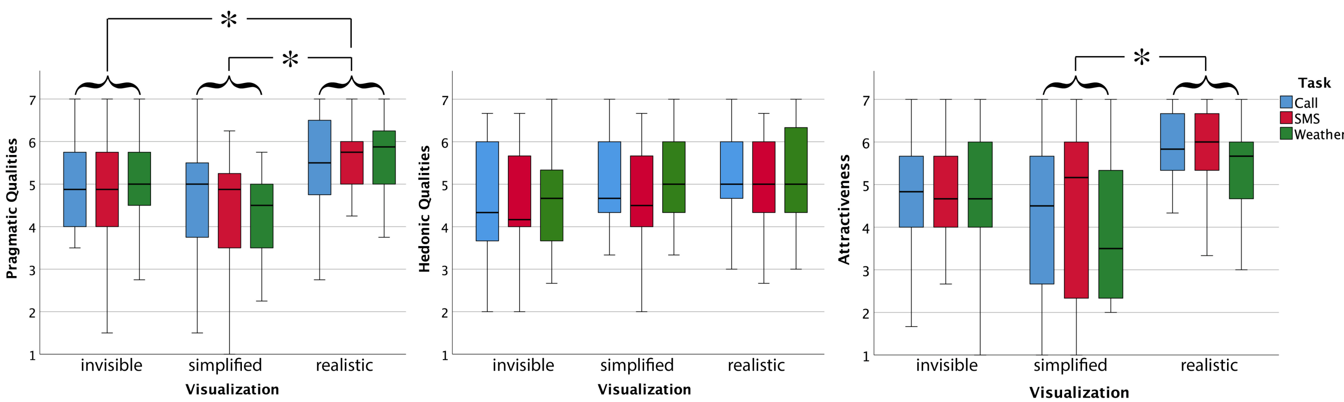

Figure 6: AttrakDiff scales: pragmatic quality, hedonic quality, global attractiveness to rate the three levels of realism from 1=low rating to 7=high rating

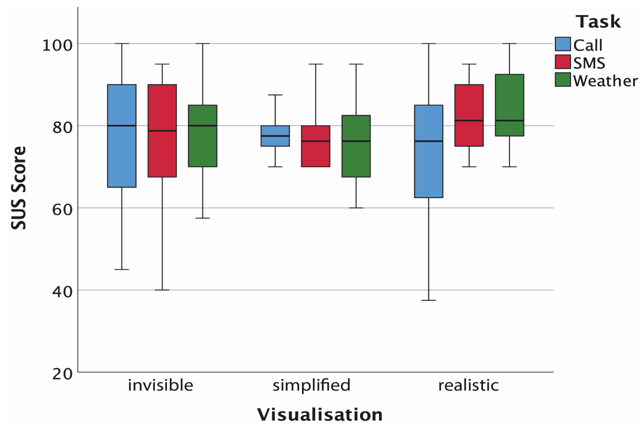

Figure 7: SUS Scores

- *"The nice appearance makes the conversation very pleasant."* (P16, TASK: WEATHER)
- *"Very pleasant to have a human person in front of you"* (P18, TASK: CALL)
- *"Interacting with a human made me feel more connected to the outside world in contrast to my previous experiences with VR headsets and AR applications which all shielded me from my surrounding "* (P1, TASK: SMS)
- *"Existing visual focus, I knew who I am talking to and who gives me answers."* (P6, TASK: WEATHER)

Moreover, communication with the human agent created a sense of social interaction:

- *"Ease of use increases mainly because it feels more like a social interaction, as it is not natural to speak with "the air". Especially the visualization of the call with the speech assistant let it seem much more real and natural. The fact that giving a "person" commands is probably easier and more intuitive than speaking into the empty space."* (P17, TASK: CALL)

Some participants even noted that the human visualization creates a kind of (emotional) connection:

- *"Very humanoid visualization, somehow creates something like connection ..."* (P2, TASK: SMS)
- *"Familiar type of visualization (human) makes it easy to connect to the speech assistant."* (P5, TASK: CALL)
- *"I had a caregiver I talked to."* (P6, TASK: CALL)
- *"The visualization is interesting because one imagines that a kind of butler does a job for a person and that this person is human."* (P12, TASK: SMS)
- *"Seeing a person to speak to helps a lot to feel comfortable and to feel understood."* (P14, TASK: WEATHER)

What decreased the quality of the simplified versus the realistic humanoid? Participants stated the appearance of the SIMPLIFIED VISUALIZATION was distracting and unpleasant:

- *"The visual style was very distracting and in some way made the whole process of the interaction with the assistant very unpleasant. I did not want to keep looking at the speech assistant."* (P5, TASK: CALL)
- *"Perhaps another way of visualizing the less dynamic one is easier for communication, since one is distracted by the bright colors and the animation."* (P17, TASK: CALL)

The humanoid representation lacked realism and was found to be too technical:

- *"Gradient has distracted, overall it was too human-like, but then no real human"* (P6, TASK: CALL),
- *"The technical appearance has a negative effect on me."* (P6, TASK: WEATHER).

What decreased the quality of the invisible agent versus the realistic humanoid? The missing embodiment of the agent triggered a feeling of social isolation and being in an unsocial or unpleasant situation:

- *"Especially due to the contrast of the humanoid voice assistant shown before, this seems extremely odd and empty when there is no visualization. Without it [the human avatar, t.a.], it feels very isolated."* (P17, TASK:SMS),
- *"I feel more comfortable when generally speaking with something visible, because I have the feeling that a body I see is listening to me more than one I don't see"* (P13, TASK: CALL),
- *"When I don't see anything or anyone, It's a bit unsocial."* (P14, weather),
- *"The voice feels like it's a voice coming out of your head. This is unpleasant."* (P13, TASK: SMS).

The absence of a non-visual communication, such as body language and gaze was missed as it was found harder and unpleasant to communicate with the agent and to interact with the system:

- *"Only voice and tone of speech as a connection between participant and speech assistant. No other connection like body language."* (P5, TASK: CALL),
- *"for me the invisible interface didn't improve the interaction, there was no clear point to look at, and it was unclear where the voice came from spatially. As I am in a visual medium, I need visual feedback. "* (P6, TASK: CALL).

Similar to the simplified humanoid, the invisible agent was found too technical:

- *"really too technical compared to a visualized assistant - although it's not true..."* (P2, TASK: CALL).

When would an invisible agent be more useful? However, quantitative as well as qualitative data indicated a true benefit of a realistic rendered conversational agent, the absence of a visual appearance of the agent was also named as an advantage in some cases. There is no distraction from other visual information, which was in particular emphasized as an advantage in situations where full attention is required:

- *"Certainly helpful if you have to focus on other visible things."* (P2, TASK: SMS),
- *"It's amazing how weird it feels when you can't see, but then you concentrate more on the content."* (P2, TASK: WEATHER),
- *"No distraction from a visual component."* (P5, TASK: CALL),
- *"I feel less distracted without a visualization."* (P7, TASK: CALL).

5 DISCUSSION

In this section, we discuss the results of our study and relate them to results from previous work. Afterwards we give design recommendations for agents in AR.

Advantages of a Realistic Humanoid Agent

Our qualitative data indicate that a realistic human design of the agent is perceived appealing and makes the interaction with the system a social and quasi-natural conversational experience. That is supported by the significantly higher perceived PRAGMATIC QUALITIES compared to the simplified and invisible agent, but also makes the realistic agent significantly more ATTRACTIVE than the simplified one. These results are in line with previous work on embodied agents done in other fields than AR. Both, Parise et al. [25] and Yee et al. [39] found as well that users prefer agents with human-like representations of higher realism. Moreover, Wagner et al. [37] showed that embodied agents are perceived to be more sympathetic and likable than agents without embodiment, which also corresponds to our finding.

From a pragmatic point of view, it seems desirable for users that agents have social cues. The interviews show that the lack of social cues was criticized for both the simplified and the invisible agent. Especially gaze was appreciated as communication and social connection support. This is in line with the results of Schrammel et al. [27] who found that the gaze direction of the agent was recognized, and also in line with the findings of Koda and Maes [14] who found that users were not distracted by the presence of a face or a facial expression. Moreover, these results show that users expect agents to behave socially, which is reflected, among other things, in eye contact. Gaze cues as an instrument of non-verbal communication and the lack of gaze for the simplified and invisible agent may also have contributed to a significantly lower rating of their PRAGMATIC QUALITIES. This, again, is consistent with prior research [2, 35, 36] who found that non-verbal communication benefits the interaction with the agent.

One benefit of seeing the lips of an agent might have positive effects, even though it has not been mentioned by our participants. Research in the field of audio-visual speech perception showed that the sight of the face of a talking person helps in understanding the spoken word [31], and seeing the speaker's lips enables the listener to hear and thus to understand better [28].

Disadvantages of a Simplified Humanoid Agent

In general, the simplified agent was perceived to be disturbing and not very appealing. Research on negative emotions caused through character design could lead to the assumption that a hyper-realistic character can have an uncanny effect rather than a stylized character [33]. In our study the stylized version was perceived unpleasant, which means that (1) the realistic humanoid does not fall into the Uncanny Valley and (2) that the simplified humanoid is simply neither appealing nor nice.

A lack of gaze and eye contact was stated, although the simplified agent had the same possibilities as the realistic human agent to turn and look at the user.

Thus, it seems as if the missing rendered eyes are key in knowing whom an agent is talking to and for creating a sort of social connection and enjoyable communication situation.

Disadvantages of an Invisible Humanoid Agent

However, current speech assistants do not have a visual representation and are invisible (except for the physical device that embeds the hardware). Our results show that the invisible visualization led to a significantly lower score in PRAGMATIC QUALITIES compared to the human agent. As participants highlighted the value of visual cues in communication, such as gaze, the absence of visual cues can explain the decreased pragmatic quality.

Advantages of an Invisible Humanoid Agent

However, even though the quantitative results indicate that an invisible agent lowers usability, the qualitative feedback gave interesting insights into possible use cases when an invisible agent might be better than a visible one.

For tasks and activities where visual attention is important, a visible agent can be perceived as a distraction. Therefore, it is not desirable to use an agent in such cases when the focus is required for other tasks, may it be driving a car, pedestrian navigation, or storing bought items in the fridge.

While most of our findings confirm previous work identifying the favor of realistically rendered humanoid agents (versus simplified or invisible ones), we can learn from our qualitative feedback that such preference clearly depends on the situation, context, and cognitive, in particular visually, demanding primary task.

Design Recommendations

For the design of embodied conversational agents in AR, we recommend a richly detailed and textured humanoid representation if the primary task is to communicate with the agent. Then, we highly recommend the implementation of social cues, such as turning the agent in the direction of the user and following the user with the eyes of the agent when the user moves. We recommend to carefully design such behavior with high degree of realism, being aware that photo-realism may cause uncanny emotions.

When the user is mobile or busy with a primary task, making the agent invisible should seriously be taken into account.

Limitations

Even though a controlled study truly has the advantage to isolate a factor and to understand how the independent variables affect the dependent ones, such experiment design is always limited to a limited number of variations.

We tested three visualization types and three tasks, and both numbers limit our findings. The simplified visualization using a wire-frame rendering of the character is only one way to reduce the degree of detail of the realistic humanoid avatar. In terms of pragmatic quality, a different, perhaps more appealing design would probably have received a better rating of attractiveness. Also, the animation of the eyes would probably have led to a better evaluation of the attractiveness concerning the pragmatic quality. Participants were standing during the scenarios in which the tasks were solved. In mobile scenarios, for example, when walking, cycling, or driving, the pragmatic quality of the invisible agent most probably would have been rated much better. At least this is what our qualitative results suggest.

We showed female and male agents to avoid a gender bias on our results. We did not define gender as variable as otherwise, the study would have taken too long. Such investigation would be worth exploring in future work.

Hence, further research that compares different stylized agents would lead to a better understanding how to design simplified agents; and a study design with multitasks and cognitive demanding situations would help to better understand under which conditions a visible versus an invisible agent is appropriate.

6 CONCLUSION

In this paper, we investigated how different realism levels of visual representation influence the usability of intelligent conversational agent on AR glasses. Results of a user study show that a realistic humanoid agent increases the pragmatic qualities compared to a simplified version and an invisible agent, and that our realistic rendered agent was perceived to be more attractive than its simplified version. While these results can be generalized for conversational tasks solved when standing, we also found that in situations when the environment or situation required the visual attention of the user, an invisible agent can be more appropriate.

Through these design recommendation, we hope to help researchers and practitioners to design and develop better, useful, and adaptive embodied conversational agents for future Augmented and Mixed Reality.

7 ACKNOWLEDGMENTS

This work is supported by the German Ministry of Education and Research (BMBF) within the GEVAKUB project (01JKD1701B).

REFERENCES

[1] Mahoro Anabuki, Hiroyuki Kakuta, Hiroyuki Yamamoto, and Hideyuki Tamura. 2000. Welbo: An Embodied Conversational Agent Living in Mixed Reality Space. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems (CHI EA '00)*. ACM, New York, NY, USA, 10–11. https://doi.org/10.1145/633292.633299

[2] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

[3] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoper. Virtual Environ.* 6, 4 (Aug. 1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

[4] Selim Balcisoy, Remy Torre, Michal Ponder, Pascal Fua, and Daniel Thalmann. 2000. Augmented Reality for Real and Virtual Humans. In *Proceedings of the International Conference on Computer Graphics (CGI '00)*. IEEE Computer Society, Washington, DC, USA, 303–. http://dl.acm.org/citation.cfm?id=792759.793057

[5] Harry Brenton, Marco Gillies, Daniel Ballin, and David Chatting. 2005. D.: The uncanny valley: does it exist. In *In: 19th British HCI Group Annual Conference: workshop on human-animated character interaction*.

[6] John Brooke. 1996. SUS: A quick and dirty usability scale.

[7] Justine Cassell. 2000. Embodied Conversational Agents. MIT Press, Cambridge, MA, USA, Chapter Nudge Nudge Wink Wink: Elements of Face-to-face Conversation for Embodied Conversational Agents, 1–27. http://dl.acm.org/citation.cfm?id=371552.371554

[8] Andrew J. Cowell and Kay M. Stanney. 2003. Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents. In *Intelligent Virtual Agents*, Thomas Rist, Ruth S. Aylett, Daniel Ballin, and Jeff Rickel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 301–309.

[9] Marc Fabri, Salima Y. Awad Elzouki, and David Moore. 2007. Emotionally Expressive Avatars for Chatting, Learning and Therapeutic Intervention. In *Proceedings of the 12th International Conference on Human-computer Interaction: Intelligent Multimodal Interaction Environments (HCI'07)*. Springer-Verlag, Berlin, Heidelberg, 275–285. http://dl.acm.org/citation.cfm?id=1769590.1769621

[10] Jodi Forlizzi, John Zimmerman, Vince Mancuso, and Sonya Kwak. 2007. How Interface Agents Affect Interaction Between Humans and Computers. In *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces (DPPI '07)*. ACM, New York, NY, USA, 209–221. https://doi.org/10.1145/1314161.1314180

[11] Arjan Geven, Johann Schrammel, and Manfred Tscheligi. 2006. Interacting with Embodied Agents That Can See: How Vision-enabled Agents Can Assist in Spatial Tasks. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (NordiCHI '06)*. ACM, New York, NY, USA, 135–144. https://doi.org/10.1145/1182475.1182490

[12] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer*.

[13] Marc Hassenzahl and Andrew Monk. 2010. The Inference of Perceived Usability From Beauty. *Human-Computer Interaction* 25, 3 (2010), 235–260. https://doi.org/10.1080/07370024.2010.500139

[14] T. Koda and P. Maes. 1996. Agents with faces: the effect of personification. In *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*. 189–194. https://doi.org/10.1109/ROMAN.1996.568812

[15] Nicole C. Krämer, Nina Simons, and Stefan Kopp. 2007. The Effects of an Embodied Conversational Agent's Nonverbal Behavior on User's Evaluation and Behavioral Mimicry. In *Intelligent Virtual Agents*, Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 238–251.

[16] Helen McBreen, Paul Shade, Mervyn Jack, and Peter Wyard. 2000. Experimental Assessment of the Effectiveness of Synthetic Personae for Multi-modal e-Retail Applications. In *Proceedings of the Fourth International Conference on Autonomous Agents (AGENTS '00)*. ACM, New York, NY, USA, 39–45. https://doi.org/10.1145/336595.336968

[17] H. M. McBreen and M. A. Jack. 2001. Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 31, 5 (Sep. 2001), 394–405. https://doi.org/10.1109/3468.952714

[18] Rachel McDonnell and Martin Breidt. 2010. Face Reality: Investigating the Uncanny Valley for Virtual Faces. In *ACM SIGGRAPH ASIA 2010 Sketches (SA '10)*. ACM, New York, NY, USA, Article 41, 2 pages. https://doi.org/10.1145/1899950.1899991

[19] Paul Milgram. 2009. A Taxonomy of (Real and Virtual World) Display and Control Interactions. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST '09)*. ACM, New York, NY, USA, 10–10. https://doi.org/10.1145/1643928.1643932

[20] P. Milgram and F. Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems* E77-D, 12 (dec 1994), 1321–1329. Available at http://vered.rose.utoronto.ca/people/paul_dir/IEICE94/ieice.html.

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

[22] M. Mori, K. F. MacDorman, and N. Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (June 2012), 98–100. https://doi.org/10.1109/MRA.2012.2192811

[23] Christos Mousas, Dimitris Anastasiou, and Ourania Spantidi. 2018. The effects of appearance and motion of virtual characters on emotional reactivity. *Computers in Human Behavior* 86 (2018), 99 – 108. https://doi.org/10.1016/j.chb.2018.04.036

[24] A. Niculescu, D. Hofs, B. van Dijk, and A. Nijholt. 2010. How the agent's gender influence users' evaluation of a QA system. In *2010 International Conference on User Science and Engineering (i-USEr)*. 16–20. https://doi.org/10.1109/IUSER.2010.5716715

[25] Salvatore Parise, Sara Kiesler, Lee Sproull, and Keith Waters. 1996. My Partner is a Real Dog: Cooperation with Social Agents. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 399–408. https://doi.org/10.1145/240080.240351

[26] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. 2015. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Comput. Graph. Forum* 34, 6 (Sept. 2015), 299–326. https://doi.org/10.1111/cgf.12603

[27] Johann Schrammel, Arjan Geven, Reinhard Sefelin, and Manfred Tscheligi. 2007. "Look!": Using the Gaze Direction of Embodied Agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1187–1190. https://doi.org/10.1145/1240624.1240803

[28] Jean-Luc Schwartz, Frédéric Berthommier, and Christophe Savariaux. 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, 2 (2004), B69 – B78. https://doi.org/10.1016/j.cognition.2004.01.006

[29] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the Uncanny Valley in Virtual Character Design. *Interactions* 25, 5 (Aug. 2018), 45–49. https://doi.org/10.1145/3236673

[30] Valentin Schwind, Katrin Wolf, Niels Henze, and Oliver Korn. 2015. Determining the Characteristics of Preferred Virtual Faces Using an

Avatar Generator. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. ACM, New York, NY, USA, 221–230. https://doi.org/10.1145/2793107.2793116

[31] W. H. Sumby and Irwin Pollack. 1954. Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America* 26, 2 (1954), 212–215. https://doi.org/10.1121/1.1907309 arXiv:https://doi.org/10.1121/1.1907309

[32] Quentin Summerfield, Vicki Bruce, Alan Cowey, Andrew W. Ellis, and D. I. Perrett. 1992. Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 335, 1273 (1992), 71–78. https://doi.org/10.1098/rstb.1992.0009

[33] Clive Thompson. 2005. Monsters of photorealism. *Wired. com* (2005).

[34] Angela Tinwell. 2009. Uncanny As Usability Obstacle. In *Proceedings of the 3D International Conference on Online Communities and Social Computing: Held As Part of HCI International 2009 (OCSC '09)*. Springer-Verlag, Berlin, Heidelberg, 622–631. https://doi.org/10.1007/978-3-642-02774-1_67

[35] Ivo van Es, Dirk Heylen, Betsy van Dijk, and Anton Nijholt. 2002. Gaze Behavior of Talking Faces Makes a Difference. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 734–735. https://doi.org/10.1145/506443.506571

[36] Ivo van Es, Dirk Heylen, Betsy van Dijk, and Anton Nijholt. 2002. Making Agents Gaze Naturally - Does It Work?. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '02)*. ACM, New York, NY, USA, 357–358. https://doi.org/10.1145/1556262.1556320

[37] Daniel Wagner, Mark Billinghurst, and Dieter Schmalstieg. 2006. How Real Should Virtual Characters Be?. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '06)*. ACM, New York, NY, USA, Article 57. https://doi.org/10.1145/1178823.1178891

[38] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 281, 12 pages. https://doi.org/10.1145/3290605.3300511

[39] Nick Yee, Jeremy N Bailenson, and Kathryn Rickertsen. 2007. A Meta-analysis of the Impact of the Inclusion and Realism of Human-like Faces on User Experiences in Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/1240624.1240626

[40] Katja Zibrek, Elena Kokkinara, and Rachel Mcdonnell. 2018. The Effect of Realistic Appearance of Virtual Characters in Immersive Environments - Does the Character's Personality Play a Role? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1681–1690. https://doi.org/10.1109/TVCG.2018.2794638

[41] John Zimmerman, Ellen Ayoob, Jodi Forlizzi, and Mick McQuaid. 2005. Putting a Face on Embodied Interface Agents. (1 2005). https://doi.org/10.1184/R1/6470366.v1