Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents

Dhaval Parmar[1] · Stefan Olafsson[1,2] · Dina Utami[1] · Prasanth Murali[1] · Timothy Bickmore[1]

Abstract

Designers of virtual agents have a combinatorically large space of choices for the look and behavior of their characters. We conducted two between-subjects studies to explore the systematic manipulation of animation quality, speech quality, rendering style, and simulated empathy, and its impact on perceptions of virtual agents in terms of naturalness, engagement, trust, credibility, and persuasion within a health counseling domain. In the first study, animation was varied between manually created, procedural, or no animations; voice quality was varied between recorded audio and synthetic speech; and rendering style was varied between realistic and toon-shaded. In the second study, simulated empathy of the agent was varied between no empathy, verbal-only empathic responses, and full empathy involving verbal, facial, and immediacy feedback. Results show that natural animations and recorded voice are more appropriate for the agent's general acceptance, trust, credibility, and appropriateness for the task. However, for a brief health counseling task, animation might actually be distracting from the persuasive message, with the highest levels of persuasion found when the amount of agent animation is minimized. Further, consistent and high levels of empathy improve agent perception but may interfere with forming a trusting bond with the agent.

Keywords Virtual agents · Animation fidelity · Voice quality · Rendering style · Simulated empathy · Agent perception

✉ Dhaval Parmar
   d.parmar@northeastern.edu

Extended author information available on the last page of the article

1 Introduction

A general assumption in the development of virtual humanoid agents is that anything that makes them more natural and lifelike must be desirable: the most naturalistic voice, appearance, rendering style, nonverbal behavior, and animation fluidity available should always be preferred. However, some studies have demonstrated that this is not always true. For example, Ring et al. found that user preferences for rendering style depend on the type of task the user is performing with the agent: participants preferred toon-shaded characters for entertainment, but more photorealistic characters for serious applications such as medical counseling [31]. Maintaining consistency in the channels of realism is important for avoiding the uncanny valley effect [25] thereby leading towards positive perception of characters [16, 22]. In a study comparing the effects of maximizing the realism of each channel (e.g., photorealism, voice, etc.) versus using channels that were matched in realism, Nass and Gong found that matching channel realism had more impact on positive user perceptions [26]. Similarly, the need for consistency in verbal and non-verbal personality cues for agents was emphasized in a study by Isbister and Nass [15]. Mitchell et al. also found that mismatches in voice realism (human vs. synthetic) and appearance (human vs. robot) led to the highest ratings of "eeriness" [24], following predictions of the uncanny valley effect. Together, these studies indicate a complex, often non-additive, relationship between the realism of each channel and user perceptions of realism, humanness, and acceptability.

Although these and other studies have investigated the impact of voice quality and photorealism [23, 31, 45], few have explored the impact of the animation quality of non-verbal behavior on perceptions of and attitudes towards virtual agents. The quality of conversational behaviors may be particularly important for "Embodied Conversational Agents" (ECAs [8]) that simulate face-to-face conversation with users. In one of the few studies to investigate this, Wu, et al. conducted an investigation in virtual reality, finding that increased realism (animated vs. static character) led to significantly greater perceptions of co-presence and greater emotional response [44]. Although some virtual agent researchers use motion capture or rotoscoped animations, most use procedural animation with behaviors (such as hand gestures) indexed from a relatively small "gestuary". We know anecdotally that these lead to perceptions of "repetitive" and "robotic" behavior, but the precise impacts of these less-than-realistic animations on user perceptions and attitudes are unknown.

In addition, user preferences for different agent designs may not always predict task outcomes. For example, users may find a friendly virtual exercise coach likable but may perform better under the guidance of a "drill sergeant" persona. Consistency of realism in the design of virtual characters has only been examined visually and not with other stimuli [16]. Few studies on virtual agent design have explored the impact of channel fidelity or realism on actual task outcomes, such as persuasion, and those that have largely failed to find any effect [45].

Given the inconsistent findings, general lack of evidence, and a bewildering array of options that designers of virtual agents have, we conducted an empirical study to assess a range of realism options for a virtual agent in a serious health counseling domain. We not only assessed user attitudes toward the agent, but also the effect of the agent's design on its ability to persuade users to commit to obtaining a health care proxy, someone you appoint to make medical decisions on your behalf in the event you are incapacitated and unable to make decisions or communicate with health care providers. Thus,

in our study, we not only manipulate speech realism and rendering style but also the animation quality of nonverbal behavior used by the virtual agent playing the role of a health counselor.

Interactions where agents comfort users through expressed empathy have been shown to be important in alleviating frustration among users and in increasing user affinity towards the agent [3, 19, 27]. Therefore, studying the effects of manipulating different levels of empathy of a persuasive counselor agent on user perceptions may have important health-care applications. This work extends our previous report [28] by further exploring the importance of naturalism on user perceptions of this particular behavior that is essential in education and counseling applications of agents; namely, the simulation of agent empathy for users' affective states.

2 Related work

2.1 Effects of animation fidelity

As virtual characters' visual realism increases, so do users' expectations about the character's behavior. The right balance of animation and visual fidelity is important [38] to avoid the uncanny valley effect [25]. Lane et al. investigated the role of animation fidelity of virtual humans (animated vs. static) in a learning environment for intercultural communication skills and found that learners took significantly longer to analyze and respond to the actions of animated virtual humans, suggesting a deeper engagement [20]. Wu et al. studied the effects of an animated and static virtual human in a medical virtual reality system for educating nurses about the signs and symptoms of patient deterioration. They found that participants in the animated condition exhibited a higher sense of co-presence and greater emotional response, compared to the static condition [44]. Research has also shown that close emulation of the features of human-human face-to-face communication contributes to smoother communication and makes the interaction more stimulating, motivating, and engaging [9, 45]. Thus, although virtual characters have been shown to be effective in the context of health counseling [6, 18], there is a need to systematically study the effects of animation fidelity, as it interacts with the factors of voice and rendering style, in the design of virtual characters.

2.2 Effects of voice realism

Prior research has looked at the social perception of human speech compared to computerized text-to-speech (TTS). Mitchell et al. studied the cross-modal effects of voice (synthetic vs. human-recorded) and embodiment (robot vs. human) and identified that the cross-modal dimensions lead to a feeling of eeriness [24]. Tinwell et al. demonstrated that a visual-auditory mismatch correlates with uncanniness [36]. These results suggest the need for avoiding the uncanny valley by matching the character's visual elements and voice on the continuum between robotic to human-like. Stern et al. conducted a study where listeners were presented with a persuasive argument in either a human or a TTS voice. They found that the human voice was perceived more favorably than the TTS voice and the speaker was perceived more favorably when the voice was human [34]. However, they found no evidence that computerized speech, as compared with the human voice, affected the degree of persuasion. In a study comparing a mix of human and TTS voice versus a TTS voice

alone, Gong et al. showed opposite effects on task performance and attitudinal responses. Users interacting with the TTS-only interface performed the task significantly better, while users interacting with a mixed-voice interface thought they did better and had more positive attitudinal responses [14]. However, the TTS-only voice was preferred due to its consistency and ability to facilitate the users' interaction with the interface. These varied and inconsistent findings show that more research is needed to form a clear picture of the effects of agent voice realism on user perception and task outcomes, and how agent voice interacts with the other channels of realism.

2.3 Effects of rendering style

Changes in the appearance of the agent can contribute to positive or negative attitudes regarding the character. Welch et al. demonstrated that visual realism is necessary for human cooperation in a virtual environment [41]. However, Dai and MacDorman showed that realism had the contrary effect where the less realistic agent increased treatment adherence intention and consultation enjoyment [12]. Ring et al. found a toon-shaded agent to be more likable and caring compared to a realistic one when having social dialogue, whereas the more realistic one was found to be more appropriate for serious tasks like medical counseling [31]. McDonnell et al. found that though toon-shaded characters were overall considered more appealing, friendly, and trustworthy, the highest quality realistic rendering scored as favorably as the toon-shaded ones [23]. Zibrek et al. found toon-style characters as having a more agreeable personality in a 2D-screen-based experiment [49], but when the experiment was run within 3D virtual reality the rendering styles did not vary significantly in appeal [48]. The only consistent finding between the two studies was that the introverted personality of the agent was rated significantly lower on appeal than other personalities. Zell et al. [47] found that visual realism alone is a bad predictor for appeal, eeriness, or attractiveness, which diverges from the uncanny valley theory. Zanbaka et al. showed that the visual realism of the agents did not influence the degree of persuasion. In a study comparing virtual humans, virtual characters, and real actors giving persuasive information, they found no difference in persuasion based on the realism of the persuasion source [46]. These inconsistent findings prompted us to investigate this space further, particularly in a serious task-oriented domain, such as health counseling.

2.4 Effects of simulated empathy

Klein et al., conducted one of the first studies on the calming effects of empathic feedback by agents on users, demonstrating significant reductions in (induced) user frustration. While they compared different interaction and feedback techniques, the agent had a text-based representation in all conditions [19]. Bickmore and Schulman explored different approaches for simulating agent empathy for users, contrasting unconstrained user input (high expressivity) and vague agent feedback (low empathic accuracy) with constrained user input (low expressivity) and agent feedback that accurately reflected the user's emotional state (high empathic accuracy), finding that empathic accuracy was more important to users [3]. Bickmore et al., compared and contrasted several media channels for agent conveyance of affect, including physical touch via haptics, and found that among facial display, prosody, and haptics, that agent facial display was the most important channel [4]. Nguyen and Masthoff compared empathic agents that used visual representations versus text only, finding that the animated agent representation led to significantly higher user

ratings of the agent being empathic, trustworthy, enjoyable, caring and likeable compared to the non-visual representation [27]. Kim et al., demonstrated that prosody can play a significant role in agent simulated empathy, showing that a speech-only agent that used appropriate prosody in its empathic messages resulted in significantly better user attitudes towards itself [17]. Since conveying empathy is important in serious settings such as education or healthcare counseling, we wanted to study how various levels of empathic feedback interact with the different agent channels of realism and how it affects user perception.

3 Virtual agent design

In our current effort, we evaluate the effect of varying the animation fidelity, speech quality, rendering style, and simulated empathy of a virtual agent on user perceptions and persuasion following a health counseling conversation on the use of a health care proxy—a legally appointed person who makes medical decisions on behalf of someone unable to do so themselves. We use a single agent in order to systematically vary the features studied in this research, while tightly controlling all other factors such as appearance, gender, race, etc., which is consistent with prior studies [23, 39, 48].

The virtual counselor we created makes the case for obtaining a health care proxy and attempts to persuade the user to commit to obtaining a health care proxy by the end of the dialogue. The script for the agent dialogue was developed in collaboration with a physician and performed by a trained health counselor in a recorded mock-counseling session.

3.1 ECA system design

The ECA system was developed using the Unity game engine [37] and was rendered in a web browser using WebGL. The system used a hierarchical task-network-based dialogue manager to drive the ECA dialogue. It presented users with a multiple-choice response menu at each turn of the conversation (Fig. 1). The system utilized programmatic triggers within the dialogue script to drive agent animations.

3.2 Animation

Following work on the effects of animation fidelity of virtual humans in medical settings [39, 40, 44], we varied the animation quality on three levels. The first level (Static) was non-animated except for lip synchronization to the human voice or TTS.

The second level (Gestuary) utilized a library of gesture and posture shift animations previously developed for conversational health counseling agents. Gestures were based on the reference video of the trained health counselor, which was interpreted and annotated by two independent raters providing descriptions of hand gestures and posture shifts following the coding description in Table 1 with an inter-rater reliability of Cohen's $\kappa = 0.7$. Some examples are shown in Fig. 2. Mapping library features to annotated movements, the animated behaviors were generated by an automatic nonverbal behavior generator [10] synchronized with speech. Gestuary represents the most common procedural animation approach used in the virtual agents research community.

The third level (Manual) represents the highest fidelity of animation. The agent's hand gestures and posture changes were created entirely by a human animator directly following the reference video, an approach sometimes referred to as rotoscoping. Lip synchronization was

Fig. 1 The counseling agent screen as seen by the participants. The agent has her arms ready in gesture space

still performed algorithmically. Compared to the Gestuary agent, the Manual agent was naturally nuanced (see E and F in Fig. 2) and varied (see G vs. C and H vs. D). Looking at Fig. 2-H, a gesture contrasting two items, the Manual agent encodes additional information of one being lower than the other.

3.3 Voice

Similar to the work by Mitchell et al. [24] and Nass and Gong [26], our voice quality manipulation had two levels: recorded human voice (Human) and synthesized (Synth). For the human voice condition, we used audio captured during the scripted mock-counseling session. The recording was split into audio clips for each agent turn of dialogue. We then aligned the script with the recordings using the SPPAS toolkit [7]. In this process SPPAS performed: (1) Inter-Pausal Units (IPUs) segmentation, segmenting the audio signal into units of speech bounded with pauses of at least 200 milliseconds length; (2) tokenization of the text to remove punctuation, converting numbers and symbols to written forms, and segmenting text into words; and (3) conversion of words into phonemes aligned with the speech signal using the Julius speech recognition engine [21] and HTK acoustic models trained from 16,000 Hz audio samples. The phonemes and timing markers were used to generate visemes for lip synchronization. We then combined the output from SPPAS, and the script now annotated with nonverbal behaviors, to create the final instructions sent to the Unity client and executed at runtime. In the synthesized voice condition, we used the Katherine voice from the Cereproc TTS engine [11] to generate the speech audio, the phonemes, and timing markers used by the Unity client to animate the speech.

Table 1 Nonverbal behavior coding description from the rating of animations from the reference footage

| Behavior	Description	Tags
Beat	Bi-phasic movement of the hand to emphasize parts of the speech	BEAT_L, BEAT_R, BEAT_BOTH
Contrast	Movement of the arm indicating one of two objects in the discourse being contrasted	CONTRAST_L, CONTRAST_R
Palms-down push	In gesture space, palms are down, fingers outstretched, and a movement of the elbow pushes the hands down or out slightly	PALMS_DOWN_L, PALMS_DOWN_R, PALMS_DOWN_BOTH
Posture shift	A gradual or sudden shift of weight from one leg to the other	POSTURE
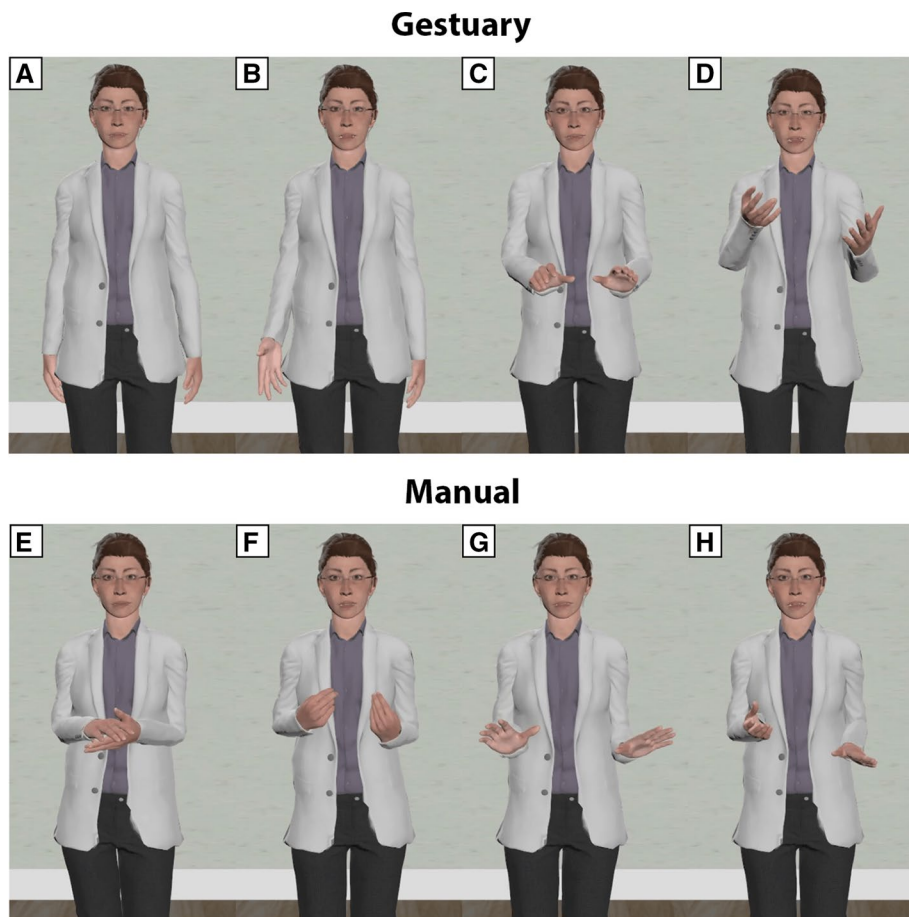
Gestuary

Manual

Fig. 2 Animation frames from the Gestuary (**a**–**d**) and Manual (**e**–**h**) conditions, with examples of nonverbal behaviors such as beat (**b**), contrast (**d**, **h**), and palms-down (**c**, **g**)

3.4 Rendering styles

Following Ring et al. [31], we created two versions of a virtual agent model to evaluate the effect of rendering style, as shown in Fig. 3. First, we created the 3D model using Adobe Fuse CC character creation software [1]. Then, for the Realistic version, we applied the detailed diffuse, normal, and ambient occlusion maps generated by Fuse to the model for a high level of detail and realistic shadows. For the Toon version, we applied an average blur effect to the default body diffuse map of the model in Adobe Photoshop [1], imported the model into Unity, and applied the Unity Toon Basic Outline shader to the diffuse material.

Fig. 3 The agent with the realistic rendering style on the left, and the toon-shaded rendering style on the right

3.5 Simulated empathy

Communicating empathy is an essential behavior in healthcare counseling. Empathy can be conveyed verbally by acknowledging the user's emotions at appropriate times in the dialogue, visually through facial display of affect, and physically through body movements such as leaning closer to the user or physical touch. Though the virtual agent is unable to touch the user's shoulder or hold their hand, the agent can provide the nonverbal immediacy of empathy by appearing to move closer to the user [33, 35].

Following prior work in simulated agent empathy by Nguyen and Masthoff [27] and others, we tested three levels of agent empathy. We first identified all "empathic opportunities" in the agent's dialogue with the user for providing agent reflection of both positive and negative user affective states. We then created three versions of agent empathic response: NO-EMPATHY, in which the agent does not acknowledge the user's affective state at all; VERBAL-EMPATHY, in which the agent provides a verbal acknowledgment of the user's affect, but the agent's visual representation is not changed (i.e., no change to agent facial display or apparent distance to the user); and FULL-EMPATHY, in which the agent provides the same verbal feedback as in VERBAL-EMPATHY, but also provides a concordant facial display and decreases its apparent distance to the user by moving the virtual camera closer to the agent, in order to increase perceived immediacy and body-perception [30, 35].

3.6 Realism study stimuli

Based on the levels of animation fidelity, voice realism, and rendering styles, 12 different versions of the system were created for the Realism Study, as described in Table 2.

Table 2 The 12 Realism Study conditions obtained from the combinations of animation fidelity, speech realism, and rendering styles, and the participant counts for each condition

Condition	Animation	Speech	Rendering	Participants
1	Manual	Human	Realistic	22
2	Manual	Human	Toon	19
3	Manual	Synth	Realistic	35
4	Manual	Synth	Toon	23
5	Gestuary	Human	Realistic	21
6	Gestuary	Human	Toon	33
7	Gestuary	Synth	Realistic	28
8	Gestuary	Synth	Toon	32
9	Static	Human	Realistic	29
10	Static	Human	Toon	19
11	Static	Synth	Realistic	35
12	Static	Synth	Toon	23

Greeting:
Agent: Hi, my name is Catherine. It's great to meet you.
User: Hi Catherine.
A: How are you doing?
U: Good.

Topic introduction (having a healthcare proxy):
A: So today, I would like to talk to you about, what you would like to have happen, if you got so sick that you could no longer communicate with your health care providers.
U: What do you mean?
A: Say you are in a situation where you are paralyzed or have stopped breathing and the doctors have to place you in a breathing machine. In order for your family and doctors to protect your wishes, they first need to know what your wishes are.
U: I see.
A: Alright, the first step is choosing a person to communicate with the doctors on your behalf. I am going to refer to this person as your health care agent.
U: Okay.
...
A: You could start by making a list of people who help you make important decisions in your life. Is there anyone that you trust to make decisions with you when you need help?
U: Yes.
A: I'm glad to hear that. Do you think one of these people could be your health care agent? Remember that, the person you choose, will have legal authority to make medical decisions for you.

Addressing barriers and decision-making:
A: This can be a hard situation. You could reach out to a family member or a health provider that you feel understands you best to have a conversation about this.
U: Okay.
A: That's okay. This can be hard to talk about. So, after you talk with the person you chose, you should talk with your doctor about this. Does that sound like something you can do?
U: Yes.
A: That is wonderful! So, there are also some documents you have to fill out to finalize that process. You should also talk with your doctor about that.

Emphasizing importance in the face of resistance:
A: That's okay, but just know that they cannot protect your rights and wishes if they are not your legal health care agent.
U: I understand.

Wrap-up and farewell:
A: Thanks for answering those questions. I appreciate you sharing your thoughts with me.
U: I am glad we had this discussion.
A: Alright, have a great day. Goodbye.

Fig. 4 Samples from the counseling dialogue between the user (U) and the agent (A) during various segments of the conversation

The counseling dialogue script was the same in each condition, i.e., about 15–20 turns long and lasting approximately 10 min (Fig. 4). At the end of the conversation, the system automatically redirected the user to a post-interaction questionnaire website.

4 Realism study

4.1 Realism study: method

To evaluate the effect of animation fidelity, voice quality, and rendering style on user perception, we conducted a 3 (Animation: Manual vs. Gestuary vs. Static) x 2 (Voice: Human vs. Synthetic) x 2 (Rendering style: Realistic vs. Toon-shaded) factorial between-subjects study on the Amazon Mechanical Turk (AMT) platform [2] (Table 2). Following enrollment, participants interacted with the agent over the web and then filled out self-report questionnaires. A between-subjects approach was chosen to effectively handle the array of manipulations for the multiple factors leading to the 12 study conditions, avoiding any difficulties in counter-balancing each condition while controlling for ordering bias. Such a controlled study, with agent interaction for each condition, would take a long time to complete for each participant in a within-subjects design. Further, running such a long study would not be practical in the mTurk environment, as the mTurk participants may lose interest in completing the long study and may not provide sincere responses.

4.1.1 Participants

The study was conducted on the Amazon Mechanical Turk (AMT) platform. All participants were required to have a 90% or higher approval rating on AMT, be located in the US, and use either Mozilla Firefox or Google Chrome with WebGL 2.0 support as their web browser.

4.1.2 Measures

In addition to socio-demographics, the participants completed the following questionnaires:

Manipulation check To assess user perceptions of our manipulations, we developed three composite measures for each factor, shown in Table 3. Chronbach's alpha showed that two of the measures had high internal consistency, i.e., the animation fidelity measure ($\alpha = 0.93$) and the voice quality measure ($\alpha = 0.96$). These measures were administered after the interaction with the agent.

Trust in the agent The 15-item Wheeless trust inventory [42] was adapted to measure participants' trust in the agent, administered after the agent interaction.

Information credibility A 6-item measure adapted from the web credibility research questionnaire [13] to measure participant perception of the credibility of the information provided by the agent, administered after the agent interaction, as shown in Table 4.

Agent satisfaction An 18-item, 7-point scale measure assessing different perceptions of the agent, including satisfaction, likability, friendliness, and caring, adapted from [31], as shown in Table 4.

Persuasion In our study, persuasion is the change in participants' intent to obtain a health care proxy. This intent is assessed at the beginning and end of the conversation on a 10-point scale via dialogue by the agent. Our persuasion outcome is this pre-post change in intent.

Table 3 The items and anchors for the measures of voice, animation, and appearance quality of the agent

Voice quality
(Disagree completely ↔Agree completely)
The agent sounded like a person
The agent's voice sounded natural
The agent's voice sounded robotic
The agent's voice was smooth
The agent's voice was awkward
The agent's voice sounded comforting
The agent's voice was eerie
The agent's voice sounded mechanical
The agent's voice sounded artificial
The agent's voice sounded weird
Animation quality
(Disagree completely ↔Agree completely)
The character's movements seemed natural
The character acted robotic
The character's behavior was smooth
The character's behavior was awkward
The character's behavior was repetitive
The character's behavior was eerie
The character's behavior was mechanical
The character's movements were human-like
The character's behaviors felt artificial
The character was stiff
Appearance quality
(Disagree completely ↔ Agree completely)
The character looked realistic
The character looked appealing
The character looked familiar
The character looked eerie

4.1.3 Procedure

All participants via Amazon's Mechanical Turk indicated their willingness to participate after being presented with a description of the study and consent information. Before interacting with the agent, they completed questionnaires on personal demographics, health-literacy, and medical mistrust. Participants interacted with one of twelve conditions (Table 2), speaking with the corresponding agent for ten minutes. Importantly, the agent defined the concept of health care proxy and asked the participant twice, once at the beginning and once at the end, regarding their commitment to obtaining a health care proxy. Lastly, participants answered questionnaires regarding the agent (voice, animation, appearance, general perception), interpersonal trust, and the perceived credibility of the agent (Table 3 and Table 4).

	General agent ratings
Table 4 The items and anchors for the general agent ratings and information credibility	*(Disagree completely ↔ Agree completely)*
	I could easily understand the character
	I felt comfortable interacting with the character
	The character had an appropriate body language
	The character was engaging
	The character was charismatic
	The character was warm
	I had fun interacting with the character
	The character was boring
	I felt awkward talking to the character
	I paid close attention to the character
	I felt like I was talking face-to-face with a person
	The character looked appropriate for her job
	(Not at all ↔ Very much)
	How friendly was the character?
	How trustworthy was the character?
	How easy was talking to the character?
	How much would you like to continue working with the character?
	How much do you like the character?
	How much do you feel that the character cares about you?
	Information credibility
	(Not at all ↔ very much)
	How believable was the information?
	How trustworthy was the information?
	How competent was the information?
	How credible was the information?
	How unbiased was the information?
	How expert was the information?

4.2 Realism study: results

A total of 305 participants (160 Male, 145 Female) aged 19–73 (M = 36.4, SD = 11.11) completed the study. We carried out factorial ANOVAs to discern the effect of animation, voice, and rendering style on our outcome measures. Post-hoc analyses were performed using Tukey HSD.

There were significant main effects of animation, $F(2, 293) = 20.62$, $p < 0.01$; and voice, $F(1, 293) = 26.73$, $p < 0.01$ on the animation quality measure (Table 5). The hand-animated agent (M = 34.18, SD = 1.42) was rated significantly higher than both the gestuary (M = 33.52, SD = 1.39), $p < 0.05$, and static agents (M = 2.91, SD = 1.38), $p < 0.01$. Additionally, the gestuary agent was rated significantly higher than the static, $p < 0.01$. The animation quality of the agent with the human voice (M = 33.96, SD = 1.53) was rated significantly higher than the synthesized voice agent (M = 33.17, SD = 1.33), $p < 0.01$.

There was a significant interaction effect of voice and animation on the voice quality measure, $F(2, 293) = 5.63$, $p < 0.01$ (Table 5). Participants rated the agent with the human voice as having a significantly higher voice quality than the synthesized voice

Table 5 ANOVA results for the Realism study across animation, voice, and rendering conditions

Factor	Statistic	Effect size (η_p^2)	Levels comparisons (mean, SD)
Animation quality ratings			
Animation	F(2, 293) = 20.62**	0.123	Manual (4.18, 1.42) > Gestuary (3.52, 1.39) Manual (4.18, 1.42) > Static (2.91, 1.38) Gestuary (3.52, 1.39) > Static (2.91, 1.38)
Voice	F(1, 293) = 26.73**	0.084	Human (3.96, 1.53) > Synth (3.17, 1.33)
Voice quality ratings			
Voice x Animation	F(2, 293) = 5.63**	0.037	Human (5.85, 1.33) > Synth (3.38, 1.44) Synth+Manual (3.92, 1.41) > Synth+Gestuary (3.01, 1.26) Synth+Manual (3.92, 1.41) > Synth+Static (3.08, 1.45)
Appearing realistic ratings			
Animation	F(2, 293) = 4.91**	0.032	Manual (3.81, 1.54) > Static (3.19, 1.5) Gestuary (3.68, 1.65) > Static (3.19, 1.5)
Appearing familiar ratings			
Voice	F(1, 293) = 6.36*	0.012	Synth (4.36, 1.97) > Human (3.83, 1.85)
Appropriate for the job ratings			
Voice	F(1, 293) = 7.70**	0.022	Human (5.5, 1.58) > Synth (4.98, 1.93)
Animation	F(2, 293) = 11.84**	0.056	Manual (5.43, 1.85) > Static (4.61, 1.91) Gestuary (5.59, 1.58) > Static (4.61, 1.91)
Rendering	F(1, 293) = 5.16*	0.017	Toon (5.44, 1.78) > Realistic (5.04, 1.79)
Agent satisfaction ratings			
Animation	F(2, 293) = 7.59**	0.049	Manual (5.08, 1.26) > Static (4.44, 1.27) Gestuary (4.96, 1.2) > Static (4.44, 1.27)
Voice	F(1, 293) = 13.93**	0.045	Human (5.13, 1.22) > Synth (4.57, 1.24)
Trust ratings			
Voice	F(2, 293) = 6.93**	0.023	Human (6.44, 1.41) > Synth (5.98, 1.55)
Information credibility ratings			
Voice	F(1, 293) = 4.37**	0.015	Human (5.84, 1.18) > Synth (5.48, 1.40)
Animation	F(2, 293) = 3.40*	0.023	Gestuary (5.86, 1.13) > Static (5.38, 1.39)
Persuasion ratings (includes Computer Literacy as covariate)			
Animation	F(2, 220) = 11.53**	0.095	Static (1.35, 2.33) > Manual (0.41, 1.82) Static (1.35, 2.33) > Gestuary (−0.04, 1.72) Manual (0.41, 1.82) > Gestuary (−0.04, 1.72)

**Indicates $p < 0.01$ and * $p < 0.05$. The last column shows which level of the independent variable was significantly higher than another

agent, across all animation and rendering levels, M = 35.85 (1.33) versus M = 33.38 (1.44), $p < 0.01$. Additionally, our analysis revealed that in the conditions where the agent had a synthesized voice, the manually-animated agent had significantly higher ratings of voice quality than the gestuary agent, M = 33.92 (1.41) versus M = 33.01 (1.26), $p < 0.05$, and the static agent, M = 33.08 (1.45), $p < 0.05$.

Our composite measure regarding the appearance of the agent had a low internal consistency, $\alpha = 0.44$. The low Cronbach's alpha implied that we could not consider the appearance composite scale to be measuring the same single factor of agent appearance. An aligned rank transform (ART) [43] would allow us to perform multi-factor non-parametric analysis on the single items that comprised the scale. Since ART transforms the data using ranks, it allows for conducting analysis using ANOVA and Tukey HSD methods. Therefore, we carried out a non-parametric aligned rank transform procedure to perform multi-factor analysis on the single items that comprised the scale.

For appearing realistic, we found a significant main effect of animation level, $F(2, 293) = 4.91$, $p < 0.01$, in which both the manually animated and gestuary agents were rated significantly more realistic than the static agent, respectively M = 3.81 (1.54) versus M = 3.19 (1.5) $p < 0.01$ and M = 33.68 (1.65) versus M = 33.19 (1.5) $p < 0.05$. There was a significant main effect of voice level on the agent appearing familiar, $F(1, 293) = 6.36$, $p < 0.05$, with the synthetic voice agent rated significantly more familiar than the human voice agent, M = 34.36 (1.97) versus M = 33.83 (1.85), $p < 0.01$. We did not find any effect of rendering level—realistic shader versus toon shader—on any of the agent appearance rating items.

Participants were asked how appropriate they felt the agent was for the job and we found significant effects of all three independent variables: voice, $F(1, 293) = 7.70$, $p < 0.01$; animation, $F(2, 293) = 11.84$, $p < 0.01$; and rendering style, $F(1, 293) = 5.16$, $p < 0.05$ (Fig. 5). For voice, the agent with the human voice was rated significantly more appropriate than the synthesized one, M = 5.5 (1.58) versus M = 4.98 (1.93). For animation levels, the hand-animated agent was significantly more appropriate than the static agent, M = 35.43 (1.85) versus M = 34.61 (1.91), as was the gestuary agent significantly more appropriate than the static agent, M = 5.59 (1.58) versus M = 4.61 (1.91), $p < 0.01$. For rendering style, the toon shaded agent was rated more appropriate for the job than the realistic agent, M = 5.44 (1.78) versus M = 5.04 (1.79), $p < 0.05$.
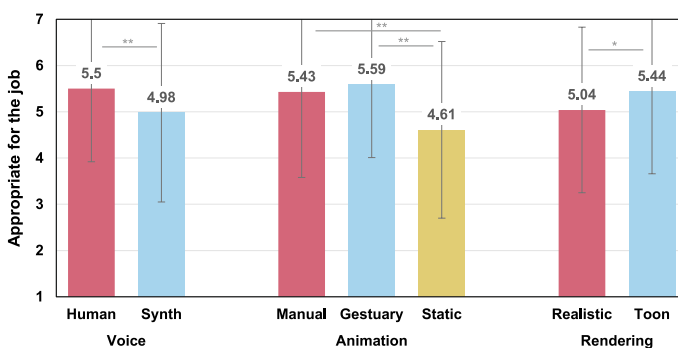

Fig. 5 Realism study: means and standard deviations for the effect of voice, animation, and rendering on the appropriateness of the agent for her job. **Indicates p values < 0.01 and *$p < 0.05$

Fig. 6 Realism study: means and standard deviations for the effect of voice on the participants' level of trust in the agent. ** indicates p values < 0.01
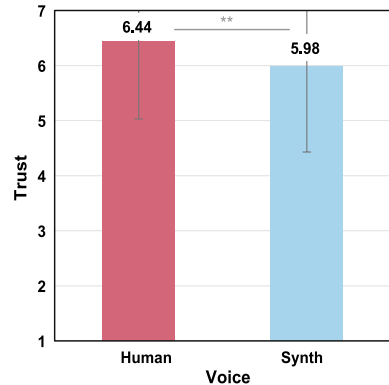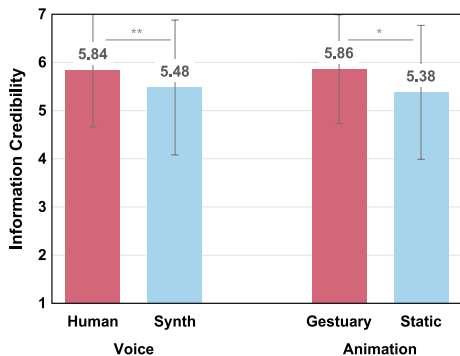

Fig. 7 Realism study: means and standard deviations for the effect of voice and animation on the credibility of the information provided by the agent. ** indicates p values < 0.01 and *$p < 0.05$

There were significant effects of animation, $F(2, 293) = 7.59$, $p < 0.01$; and voice, $F(1, 293) = 13.93$, $p < 0.01$ on ratings of satisfaction with the agent. The manually-animated agent was rated significantly higher than the static one, M = 35.08 (1.26) versus M = 34.44 (1.27), $p < 0.01$, as was the gestuary agent rated significantly higher than the static agent, M = 34.96 (1.2) versus M = 34.44 (1.27), $p < 0.01$. As for voice, the general agent ratings for the agent with the human voice were significantly higher than the ones for the synthesized voice agent, M = 5.13 (1.22) versus 4.57 (1.24), $p < 0.01$.

On trusting the agent, we found a significant main effect of voice, $F(2, 293) = 6.93$, $p < 0.01$ (Fig. 6). The agent with the human voice was rated significantly higher than the synthesized voice agent, M = 36.44 (1.41) versus M = 35.98 (1.55), $p < 0.01$. Additionally, there was a significant main effect of voice, $F(1, 293) = 4.37$, $p < 0.01$; and animation levels, $F(2, 293) = 3.40$, $p < 0.05$ on how credible participants found the information (Fig. 7). The information given by the agent with a human voice was rated more credible than the one with the synthetic voice, M = 35.84 (1.18) versus M = 35.48 (1.40), $p < 0.01$. Similarly, the information delivered by the gestuary agent was regarded more credible than the information coming from the static agent, M = 35.86 (1.13) versus M = 35.38 (1.39), $p < 0.05$.

Correlational analysis of select ordinal and ratio measures yielded several interesting insights (Table 6). The ability of the agent to persuade participants decreases with self-reported computer literacy, indicating that more tech-savvy participants may not buy into the agent-as-authoritative-counselor supposition as much as those less familiar with

Fig. 8 Realism study: means and standard deviations for the effect of animation on the change in participants' level of commitment to getting a health care proxy as a measure of persuasion
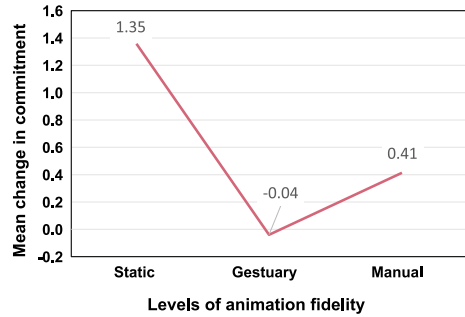

Table 6 Bivariate correlations from analysis of ordinal and ratio measures for the realism study

	Age	Trust	Information credibility	Persuasion
Computer literacy	− 0.116*	− 0.047	0.008	− 0.178**
Age		0.163**	0.153**	− 0.068
Trust			0.670**	0.007
Information credibility				− 0.005

**p* values < 0.01 and **p* < 0.05

computers. Participant trust in the agent, and their ratings of credibility of information delivered by the agent, both increased with participant age, indicating that older participants were more willing to give the agent the benefit of the doubt, regardless of study manipulation.

Regarding our main outcome measure of persuasion (i.e., commitment to getting a health proxy) we found significant pre-post differences across all study conditions, $W =$ 2176.5, $p < 0.01$ (Pre: M = 7.35 SD = 2.3 vs. Post: M = 7.88 SD = 2.35). Given the significant influence of self-reported computer literacy on persuasion (Table 6), we included computer literacy as a covariate in our MANOVA analysis. We saw a main effect of animation on the change in persuasion (Fig. 8). Participants in the static condition showed a significantly greater change, M = 31.35 (2.33), than those in the gestuary, M = −0.04(1.72)$p < 0.01$, and manually animated conditions, M = 30.41 (1.82) $p < 0.01$, as well as a significantly greater change for participants in the manually-animated condition compared to those in the gestuary, $p < 0.01$. There were no significant differences between the levels of voice and rendering styles.

4.3 Realism study: discussion

The manipulation checks for animation and voice quality showed that they were correctly perceived by our participants. When asked about animation quality, the hand-animated motion was rated higher than procedural ("gestuary" based) animation, which in turn was rated higher than a (mostly) static agent. When asked about voice quality, recorded human voice was rated higher than synthetic voice.

However, our manipulation check on rendering styles failed to demonstrate that participants consistently rated the 3D shaded character higher than the toon-shaded

character as having a more realistic appearance. We did find that animation quality significantly impacted ratings of appearance, with hand-animated and gestuary rated higher than the static agent, so it may be that the impact of animation overwhelms any influence of rendering style, or that participants had an overall positive reaction to the toon-shaded character that biased their judgment of appearance quality.

Manipulations of one media channel often influence perceptions of other channels [32], and we found this in four cases. Manipulations of voice quality significantly impacted ratings of animation quality and character appearance ("familiar" and "appropriate"), and manipulations of animation quality significantly impacted ratings of voice quality and character appearance ("realistic" and "appropriate").

Contrary to Ring et al. [31], our participants rated the toon-shaded character as being significantly more appropriate for the health counseling task than the 3D shaded character. However, our character design, setting, and task were all different from theirs, indicating there may be more complex moderators that govern the most appropriate rendering style for a character.

Overall satisfaction with the character, as well as trust in the character, and ratings of information credibility were significantly greater with a human recorded voice compared to a synthetic one. Satisfaction and credibility were also significantly greater when hand ("rotoscoped") animation was used.

Our most surprising result was that our primary persuasion outcome—change in intent to obtain a health care proxy—was significantly greater when character animation was minimized and was not influenced by any other manipulations. It could be that in brief, information-rich, counseling sessions (i.e., for "central route" persuasion [29]) animation acts as a distraction from the comprehension of the information required to make a decision. This is further supported by the finding that the highest quality animations led to higher persuasion than gestuary-based animations, under the assumption that rotoscoped animations were the most natural, and thus least distracting, of those two conditions.

5 Empathy study

5.1 Empathy study: method

To evaluate the difference in empathic response, we conducted a between-subjects study evaluating the three levels of agent empathic response (described in Sect. 3.5): NO-EMPATHY (no acknowledgment of the user's affective state), VERBAL-EMPATHY (verbal acknowledgment only), and FULL-EMPATHY (verbal and nonverbal acknowledgment).

5.1.1 Virtual agent design

For the agent design, we chose the levels of animation, voice, and rendering fidelity that resulted in the highest change in commitment from the previous evaluation study: static animation, human voice, and realistic rendering. These design factors were kept the same across all three empathy conditions.

5.1.2 Participants, measures, and procedure

The study was conducted on AMT, with the same qualifications as the Virtual Agent Evaluation study (Sect. 4). We used the same measures as in the previous study, except those measuring voice and animation quality. The procedure was also the same as in the previous study, besides being randomized to 3 conditions instead of 12.

5.2 Empathy study: results

A total of 95 participants (42% Female) aged 21–67 (M = 336.34, SD = 11.73) completed the study. We conducted one-way aligned ranks transformation (ART) ANOVAs to discern the differences between the three levels of empathic feedback on our outcome measures. Table 7 shows a summary of these results.

Table 7 ANOVA results across no-empathy, verbal-empathy, and full-empathy conditions for the Empathy study

Statistic	Effect size (η_p^2)	Levels comparisons (mean, SD)
Friendly ratings		
$F(2, 92) = 3.42*$	0.069	Full-Empathy (6.25, 0.94) > No-Empathy (5.44, 1.19)
Charismatic ratings		
$F(2, 92) = 4.23*$	0.084	Full-Empathy (5.42, 1.44) > No-Empathy (4.26, 1.56)
Warm ratings		
$F(2, 92) = 2.93*$	0.059	Full-Empathy (5.75, 1.54) > Verbal-Empathy (4.86, 1.71)
Paid close attention to the agent ratings		
$F(2, 92) = 7.03*$	0.133	Full-Empathy (6.71, 0.69) > Verbal-Empathy (5.75, 1.45)
Appropriate non-verbal behavior ratings		
$F(2, 92) = 4.54*$	0.089	Full-Empathy (5.38, 1.53) > Verbal-Empathy (4.2, 1.76)
Appropriate for the job ratings		
$F(2, 92) = 4.21*$	0.083	Full-Empathy (6, 1.44) > Verbal-Empathy (5.14, 1.73)
Trust ratings		
$F(2, 92) = 3.5*$	0.071	Verbal-Empathy (5.95, 1.57) > Full-Empathy (4.9, 1.85)
Appearance ratings		
$F(2, 92) = 8.79***$	0.16	Full-Empathy (4.95, 1.02) > Verbal-Empathy (3.99, 0.79)
Agent satisfaction ratings		
$F(2, 92) = 4.19*$	0.083	Full-Empathy (5.57, 1.03) > Verbal-Empathy (4.86, 1.1)

$***p < 0.001$ and $*p < 0.05$. The last column shows which level of the independent variable was significantly higher than another

Participants reported the Full-Empathy agent to be significantly more friendly (M = 6.25, SD = 0.94 vs. M = 5.44, SD = 1.19; F(2, 92) = 3.42, $p < 0.05$), and charismatic (M = 5.42, SD = 1.44 vs. M = 4.26, SD = 1.56; F(2, 92) = 4.23, $p < 0.05$) than the No-Empathy agent. The Full-Empathy agent was also rated significantly more warm (M = 5.75, SD = 1.54) than the Verbal-Empathy agent (M = 4.86, SD = 1.71), F(2, 92) = 2.93, $p < 0.05$.

Likewise, participants paid significantly more attention to the Full-Empathy agent (M = 6.71, SD = 0.69) than the Verbal-Empathy agent (M = 5.75, SD = 1.45), F(2, 92) = 7.03, $p < 0.05$. They also reported that the Full-Empathy agent showed more appropriate nonverbal behavior (M = 5.38, SD = 1.53) than the Verbal-Empathy agent (M = 4.2, SD = 1.76), F(2, 92) = 4.54, $p < 0.05$. Further, the Full-Empathy agent was rated more appropriate for the role of a health counselor (M = 6, SD = 1.44) than the Verbal-Empathy agent (M = 5.14, SD = 1.73), F(2, 92) = 4.21, $p < 0.05$.

Participants indicated that they trusted the agent in the Verbal-Empathy condition (M = 5.95, SD = 1.57) more than the Full-Empathy agent (M = 4.9, SD = 1.85), F(2, 92) = 3.5, $p < 0.05$. Finally, the Full-Empathy agent was rated higher on general satisfaction (M = 5.57, SD = 1.03 vs. M = 4.86, SD = 1.1; F(2, 92) = 4.19, $p < 0.05$) and appearance (M = 4.95, SD = 1.02 vs. M = 3.99, SD = 0.79; F(2, 92) = 8.79, $p < 0.001$) than the Verbal-Empathy agent.

5.3 Empathy study: discussion

Our results showed that participants correctly perceived the full-empathy agent as being more empathic. The agent exhibiting concurrent verbal and nonverbal empathic behaviors was rated highest on friendliness, warmth, and charisma, indicating that an agent that maximizes the quality available for both the verbal and nonverbal channels of empathic expression is preferable to one that does not.

However, inconsistency in the empathic channels affected participants' perception of the agent. The agent exhibiting only the verbal channel of empathy was rated lowest on warmth, showing appropriate nonverbal behavior, garnering attention, and being appropriate for her role. Further, the ratings of overall satisfaction and appearance were also lowest for the verbal empathy agent.

The exception to this was trust. Participants rated the verbal-empathy agent as more trustworthy than the full-empathy agent. Perhaps the richer channels of empathy in the form of added immediacy [30] and facial display acted as a distraction, similar to our prior findings (Sect. 4.3). The brief duration of the interaction could also have played a role, possibly preventing the participants from establishing a trusting relationship with the agent.

6 Conclusion

In this work, we studied the impacts of animation, voice, rendering styles, and simulated empathy of a virtual human on people's intention to get a health care proxy. Our results have important implications for the design of interactive virtual characters. We found that natural animations and a human-sounding voice affected how users rated the virtual human's overall acceptance, trust, and appropriateness in delivering health counseling information. For critical moments when we want to maximize persuasion, our results suggest that it might be more appropriate for the agent to be less animated, to shift the focus momentarily to the speech channel. We found few interaction effects in our results,

indicating that media channels (animation, rendering, voice) act independently, in support of the "maximization" hypothesis: the best quality available for each channel should be used, independent of the other channels, as opposed to the "consistency" hypothesis, where channels should always be matched in fidelity. Finally, we found that having consistent and high levels of empathic feedback improves the user perceptions of the agent, but may distract from forming a trusting alliance with the virtual counselor.

6.1 Limitations

Our study has some important limitations, including the relatively small convenience samples recruited on Mechanical Turk that may not generalize to any particular user demographic for a target application. Our results are from a very brief counseling session with an agent that involved essentially no rapport or relationship-building interaction [5], and so may not be representative of what would happen in longer interactions or multiple interactions, or after users have established working relationships with the agent. The research was conducted on a single agent, and therefore does not provide insights on stereotypical preconceptions of agent gender, race, age, clothing, etc. The task of obtaining a health care proxy may not have been personally-relevant to most of our participants, so the results largely reflect those from a hypothetical decision scenario. Finally, our self-report task outcome lacks the validity of an objective, behavioral outcome, such as following up to determine whether participants actually obtained health care proxies or not.

6.2 Future work

In future studies, we aim to further explore the design space of virtual characters in serious task applications, investigating manipulations of lighting and color for rendering the character. The effects of gender, age, race, and general appearance in different task scenarios with different user populations is also a large but important design space to explore. An expanded set of rendering styles beyond the two extremes evaluated in this paper, such as those studied by Zibrek et al. [48] and beyond can be explored. Our finding that animation can act as a distraction from the comprehension of key information also warrants further investigation. Finally, we plan to investigate how these effects change over time in longitudinal tasks.

References

1. Adobe: Adobe: Creative, marketing and document management solutions. https://www.adobe.com/ (2020). Retrieved 2020 July 20.
2. Amazon: Amazon mechanical turk. https://www.mturk.com/ (2020). Retrieved 2020 July 20
3. Bickmore, T., & Schulman, D. (2007). Practical approaches to comforting users with relational agents. In: CHI'07 extended abstracts on human factors in computing systems, CHI EA'07, pp. 2291–2296. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1240866.124096.
4. Bickmore, T. W., Fernando, R., Ring, L., & Schulman, D. (2010). Empathic touch by relational agents. *IEEE Transactions on Affective Computing, 1*(1), 60–71.
5. Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human–computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293–327.

6. Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L., & Paasche-Orlow, M. K. (2013). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society, 61*(10), 1676–1683.

7. Bigi, B., & Hirst, D. (2012). SPeech phonetization alignment and syllabification (SPPAS): A tool for the automatic analysis of speech prosody. In *Speech prosody* (pp. 19–22). Tongji University Press, Shanghai, China. https://hal.archives-ouvertes.fr/hal-00983699

8. Cassell, J., Sullivan, J., Churchill, E., & Prevost, S. (2000). *Embodied Conversational Agents*. MIT Press.

9. Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence, 13*(4–5), 519–538.

10. Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2004). BEAT: The Behavior Expression Animation Toolkit, pp. 163–185. Springer, Berlin. https://doi.org/10.1007/978-3-662-08373-4_8

11. Cereproc: Cereproc text-to-speech. https://www.cereproc.com/ (2020). Retrieved 2020 July 20.

12. Dai, Z., & MacDorman, K. F. (2018). The doctor's digital double: How warmth, competence, and animation promote adherence intention. *PeerJ Computer Science, 4*, e168.

13. Fogg, B., Marshall, J., Kameda, T., Solomon, J., Rangnekar, A., Boyd, J., & Brown, B. (2001). Web credibility research: A method for online experiments and early study results. In *CHI'01 extended abstracts on human factors in computing systems, CHI EA'01* (pp. 295–296). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/634067.634242

14. Gong, L., & Lai, J. (2001). Shall we mix synthetic speech and human speech? Impact on users' performance, perception, and attitude. In *Proceedings of the Sigchi conference on human factors in computing systems, CHI'01* (pp. 158–165). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/365024.365090

15. Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, nonverbal cues, and user characteristics. *International Journal of Human Computer Studies, 53*(2), 251–267. https://doi.org/10.1006/ijhc.2000.0368

16. Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology, 6*, 390. https://doi.org/10.3389/fpsyg.2015.00390

17. Kim, J., Kim, W., Nam, J., & Song, H. (2020) "i can feel your empathic voice": Effects of nonverbal vocal cues in voice user interface. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems, CHI EA'20* (pp. 1–8). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3334480.3383075

18. Kimani, E., Bickmore, T., Trinh, H., Ring, L., Paasche-Orlow, M.K., & Magnani, J.W. (2016). A smartphone-based virtual agent for atrial fibrillation education and counseling. In *Intelligent virtual agents* (pp 120–127). Springer, Cham.

19. Klein, J., Moon, Y., & Picard, R. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers, 14*(2), 119–140. https://doi.org/10.1016/S0953-5438(01)00053-4

20. Lane, H. C., Hays, M. J., Core, M. G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology, 105*(4), 1026.

21. Lee, A., & Kawahara, T. (2009). Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific signal and information processing association, 2009 annual summit and conference* (pp. 131–137). Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, Sapporo, Hokkaido, Japan. http://hdl.handle.net/2115/39653

22. MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition, 146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

23. McDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real?: Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG), 31*(4), 91.

24. Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. i-Perception **2**(1), 10–12 (2011)

25. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100.

26. Nass, C., & Gong, L. (1999). Maximized modality or constrained consistency? In D.W. Massaro (ed.), *Auditory-visual speech processing, AVSP '99, Santa Cruz, CA, USA*, August 7–10, 1999, p. 1. ISCA, Santa Cruz, CA, USA. http://www.isca-speech.org/archive_open/avsp99/av99_001.html

27. Nguyen, H., & Masthoff, J. (2009). Designing empathic computers: The effect of multimodal empathic feedback using animated agent. In *Proceedings of the 4th international conference on persuasive technology, persuasive '09*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1541948.1541958

28. Parmar, D., Ólafsson, S., Utami, D., Murali, P., & Bickmore, T. (2020). Navigating the combinatorics of virtual agent design space to maximize persuasion. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems, AAMAS'20* (pp. 1010–1018). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2020)

29. Petty, R. E., Briñol, P., Fabrigar, L., & Wegener, D. (2010). Attitude structure and change. In *Advanced social psychology: The State of the Science* (pp. 217–259). Oxford University Press.

30. Richmond, V. P., & McCroskey, J. C. (1995). *Immediacy*. Boston: Allyn & Bacon.

31. Ring, L., Utami, D., & Bickmore, T. (2014). The right agent for the job? In *Intelligent virtual agents* (pp. 374–384). Springer, Cham.

32. Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews, 7*(3), 269–284.

33. Slote, M. (2003). Empathy, immediacy, and morality. In: H. Dyke (ed.) Time and ethics: Essays at the intersection, pp. 179–188. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-3530-8_12.

34. Stern, S. E., Mullennix, J. W., Dyson, C. I., & Wilson, S. J. (1999). The persuasiveness of synthetic speech versus human speech. *Human Factors, 41*(4), 588–595.

35. Taipale, J. (2015). Beyond cartesianism: Body-perception and the immediacy of empathy. *Continental Philosophy Review, 48*(2), 161–178. https://doi.org/10.1007/s11007-015-9327-3

36. Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. *Journal of Gaming& Virtual Worlds, 2*(1), 3–25.

37. Unity: Unity real-time development platform. https://unity.com/ (2020). Retrieved 2020 July 20

38. Vinayagamoorthy, V., Steed, A., & Slater, M. (2005). Building characters: Lessons drawn from virtual environments. In *Proceedings of toward social mechanisms of android science: A CogSci 2005 workshop* (pp. 119–126). Cognitive Science Society, Stresa, Italy

39. Volonte, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S. B., & Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Transactions on Visualization and Computer Graphics, 22*(4), 1326–1335.

40. Volonte, M., Robb, A., Duchowski, A. T., & Babu, S. V. (2018). Empirical evaluation of virtual human conversational and affective animations on visual attention in inter-personal simulations. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR), pp. 25–32. IEEE, IEEE, Reutlingen, Germany. https://doi.org/10.1109/VR.2018.8446364

41. Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., & Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. Presence: Teleoperators & Virtual Environments **5**(3), 263–273

42. Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*(3), 250–257.

43. Wobbrock, J.O., Findlater, L., Gergle, D., & Higgins, J.J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, p. 143-146. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1978942.1978963. https://doi.org/10.1145/1978942.1978963

44. Wu, Y., Babu, S. V., Armstrong, R., Bertrand, J. W., Luo, J., Roy, T., Daily, S. B., Dukes, L. C., Hodges, L. F., & Fasolino, T. (2014). Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE Transactions on Visualization and Computer Graphics, 20*(4), 626–635.

45. Yee, N., Bailenson, J. N., & Rickertsen, K. (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* (pp. 1–10). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1240624.1240626

46. Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '06* (pp. 1153–1162). Association for Computing Machinery, New York, NY, USA (2006). https://doi.org/10.1145/1124772.1124945

47. Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., McDonnell, R., & Botsch, M. (2015). To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics* **34**(6) (2015). https://doi.org/10.1145/2816795.2818126

48. Zibrek, K., Kokkinara, E., & Mcdonnell, R. (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), 1681–1690. https://doi.org/10.1109/TVCG.2018.2794638

49. Zibrek, K., & McDonnell, R. (2014). Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM symposium on applied perception, SAP'14*, pp. 111–115. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2628257.2628270

Authors and Affiliations

Dhaval Parmar[1] [iD] **· Stefan Olafsson[1,2] · Dina Utami[1] · Prasanth Murali[1] · Timothy Bickmore[1]**

Stefan Olafsson
olafsson.s@northeastern.edu; stefanola@ru.is

Dina Utami
utami.d@northeastern.edu

Prasanth Murali
murali.pr@northeastern.edu

Timothy Bickmore
t.bickmore@northeastern.edu

[1]   Northeastern University, Boston, MA, USA

[2]   Present Address: Reykjavik University, Reykjavík, Iceland