Navigating the Combinatorics of Virtual Agent Design Space to Maximize Persuasion

Dhaval Parmar
Northeastern University
Boston, MA, USA
d.parmar@northeastern.edu

Stefán Ólafsson
Northeastern University
Boston, MA, USA
olafsson.s@northeastern.edu

Dina Utami
Northeastern University
Boston, MA, USA
utami.d@northeastern.edu

Prasanth Murali
Northeastern University
Boston, MA, USA
murali.pr@northeastern.edu

Timothy Bickmore
Northeastern University
Boston, MA, USA
bickmore@ccs.neu.edu

ABSTRACT

Designers of virtual agents have a combinatorically large space of choices for different media that comprise the look and behavior of their characters. We explore the systematic manipulation of animation quality, speech quality, and rendering style, and its impact on the perceptions of virtual agents in terms of naturalness, engagement, trust, credibility, and persuasion within a health counseling domain. The agent's counseling behavior was based on live video footage of a human counselor. We conducted a between-subjects study that had 12 conditions. Character animation was varied between a static image, procedural animation using a gestuary, and manually rotoscoped animation. Voice quality was varied between recorded audio of the human counselor and synthetic speech. Character rendering style was varied between 3D-shaded realistic and toon-shaded. Prior studies indicate that people prefer and attribute more sociality to other people and agents when modalities are consistent in their level of quality. Thus, we hypothesize that people will be most affected by agents whose animation, voice, and rendering style are consistent, rather than the effects of channel quality being purely additive. Results indicate that natural animations and recorded voice are more appropriate for general acceptance, trust, and credibility of the agent, and how appropriate she seems for the task. However, our results indicate that 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.

KEYWORDS

Virtual agents; animation fidelity; voice quality; rendering style; agent perception

ACM Reference Format:
Dhaval Parmar, Stefán Ólafsson, Dina Utami, Prasanth Murali, and Timothy Bickmore. 2020. Navigating the Combinatorics of Virtual Agent Design Space to Maximize Persuasion. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 9 pages.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand. © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

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: their study participants preferred toon-shaded characters for entertainment, but more photorealistic characters for serious applications such as medical counseling [20]. In another study comparing the effects of maximizing the realism of each channel (e.g., photorealism, voice, etc.) vs. using channels that were matched in realism, Nass and Gong found that matching channel realism had more impact on positive user perceptions [18]. 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" [16], following predictions of the uncanny valley effect [17]. 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 [15, 20, 32], few have explored the impact of the animation quality of nonverbal behavior on user perceptions of and attitudes towards virtual agents. The quality of conversational behaviors may be particularly important for "Embodied Conversational Agents" (ECAs [6]) 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 (an animated vs. static character) led to significantly greater perceptions of co-presence and greater emotional response [31]. 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 user 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. 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 [32].

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 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.

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 [25] to avoid the uncanny valley effect [17]. 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 [13]. 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 [31]. 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 [7] [32]. Thus, although virtual characters have been shown to be effective in the context of health counseling [4, 12], 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 [16]. Tinwell, et al. demonstrated that a visual–auditory mismatch correlates with uncanniness [23]. 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 [22]. 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 vs. 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 [11]. However, the TTS-only voice was preferred due to its consistency and ability to facilitate the users' interaction with the interface.

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 [28]. McDonnell, et al. investigated how different rendering styles affect user perceptions of a 3D character and identified that rendering style affects the appeal and trustworthiness of the characters [15]. However, 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 to be more appropriate for serious tasks, such as medical counseling [20]. Similarly, Zibrek, et al. also found toon-style characters as having a more agreeable personality [34]. Thus, a toon-shaded visualization may better suit an agent created for building a relationship with the user, whereas a more realistic look may be more appropriate for task-oriented agents. 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 [33].

These inconsistent findings prompted us to investigate this space further, particularly in a serious task-oriented domain, such as health counseling.

3 VIRTUAL AGENT DESIGN

In our current effort, we evaluate the effect of varying the animation fidelity, speech quality, and rendering style 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).

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 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 [24] 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 (Figure 1). The system utilized programmatic triggers within the dialogue script to drive agent animations.

Figure 1: (A) The counseling agent screen as seen by the participants. The agent has her arms ready in gesture space. (B) The agent with the Realistic rendering style. (C) The agent with the toon shaded rendering style.

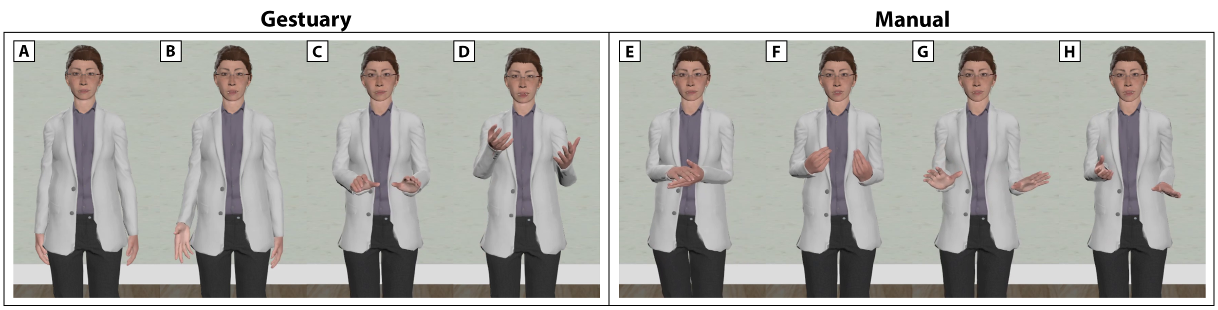

Figure 2: Animation frames from the Gestuary (A–D) and Manual (E-H) conditions, with examples of nonverbal behaviors such as beat (B), contrast (D, H), and palms-down (C,G).

3.2 Animation

Following work on the effects of animation fidelity of virtual humans in medical settings [26, 27, 31], 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 Figure 2. Mapping library features to annotated movements, the animated behaviors were generated by an automatic nonverbal behavior generator [8] 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 still performed algorithmically. Compared to the Gestuary agent, the Manual agent was naturally nuanced (see E and F in Figure 2) and varied (see G vs. C and H vs. D). Looking at Figure 2-H, a gesture contrasting two items, the Manual agent encodes additional information of one being lower than the other.

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

3.3 Voice

Similar to the work by Mitchell, et al. [16] and Nass & Gong [18], our voice quality manipulation had two levels: human recording (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 [5]. 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 [14] and HTK acoustic models trained from 16000 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 [9] to generate the speech audio, the phonemes, and timing markers used by the Unity client to animate the speech.

3.4 Rendering Styles

Following Ring et al. [20], we created two versions of a virtual agent model to evaluate the effect of rendering style, as shown in Figure 1. 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.

3.5 Study Stimuli

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

Table 2: The 12 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

The counseling dialogue script was the same in each condition, i.e., about 15-20 turns long and lasting approximately 10 minutes (Figure 3). At the end of the conversation, the system automatically redirected the user to a post-interaction questionnaire website.

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

<u>Topic introduction (having a healthcare proxy):</u>
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.*

<u>Addressing barriers and decision-making:</u>
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.*

<u>Emphasizing importance in the face of resistance:</u>
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.*

<u>Wrap-up and farewell:</u>
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.*

Figure 3: Samples from the counseling dialogue between the user (U) and the agent (A) during various segments of the conversation.

4 VIRTUAL AGENT EVALUATION

4.1 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

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.

study (Table 2). Following enrollment, participants interacted with the agent over the web and then filled out self-report questionnaires.

4.1.1 Participants. The study was conducted on the Amazon Mechanical Turk (AMT) platform [2]. 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 (α = 0.93) and the voice quality measure (α = 0.96). These measures were administered after the interaction with the agent.

Trust in the agent: The 15-item Wheeless trust inventory [29] 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 [10] to measure participant perception of the credibility of the information provided by the agent, administered after the agent interaction, as shown in Table 4.

Table 4: The items and anchors for the general agent ratings and information credibility.

General Agent Ratings *(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?

Agent satisfaction: An 18-item, 7-point scale measure assessing different perceptions of the agent, including satisfaction, likability, friendliness, and caring, 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.

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).

Table 5: ANOVA results across animation, voice, and rendering conditions. ** indicates p < .01 and * p < .05. The last column shows which level of the independent variable was significantly higher than another.

Measure	Factor	Statistic	Effect Size (η_p^2)	Levels Comparisons (Mean, SD)
Animation quality	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	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	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	Voice	F(1, 293) = 6.36*	0.012	Synth (4.36, 1.97) >Human (3.83, 1.85)
Appropriate for the job	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	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	Voice	F(2, 293) = 6.93**	0.023	Human (6.44, 1.41) >Synth (5.98, 1.55)
Information credibility	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 (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)

4.2 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.

There were significant main effects of animation and voice on the animation quality measure (Table 5). The hand-animated agent (M=4.18, SD=1.42) was rated significantly higher than both the gestuary (M=3.52, SD=1.39), p<.05, and static agents (M = 2.91, SD=1.38), p<.01. Additionally, the gestuary agent was rated significantly higher than the static, p<.01. The animation quality of the agent with the human voice (M=3.96, SD=1.53) was rated significantly higher than the synthesized voice agent (M=3.17, SD=1.33), p<.01.

There was a significant interaction effect of voice and animation on the voice quality measure (Table 5). Participants rated the agent with the human voice as having a significantly higher voice quality than the synthesized voice agent, across all animation and rendering levels, M=5.85 (1.33) vs. M=3.38 (1.44), p<.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=3.92 (1.41) vs. M=3.01 (1.26), p<.05, and the static agent, M=3.08 (1.45), p<.05.

Our composite measure regarding the appearance of the agent had a low internal consistency, α = 0.44. Therefore, we carried out a non-parametric aligned rank transform procedure on the single items that comprised the scale [30].

For appearing realistic, we found a significant main effect of animation level, in which both the manually animated and gestuary agents were rated significantly more realistic than the static agent, respectively M=3.81 (1.54) vs. M=3.19 (1.5) p<.01 and M=3.68 (1.65) vs. M=3.19 (1.5) p<.05. There was a significant main effect of voice level on the agent appearing familiar, with the synthetic voice agent rated significantly more familiar than the human voice agent, M=4.36 (1.97) vs. M=3.83 (1.85), p<.01. We did not find any effect of rendering level—realistic shader vs. 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, animation, and rendering style (Figure 4). For voice, the agent with the human voice was rated significantly more appropriate than the synthesized one, M=5.5 (1.58) vs. M=4.98 (1.93). For animation levels, the hand-animated agent was significantly more appropriate than the static agent, M=5.43 (1.85) vs. M=4.61 (1.91), as was the gestuary agent significantly more appropriate than the static agent, M=5.59 (1.58) vs. M=4.61 (1.91), p<.01. For rendering style, the toon shaded agent was rated more appropriate for the job than the realistic agent, M=5.44 (1.78) vs. M=5.04 (1.79), p<.05.

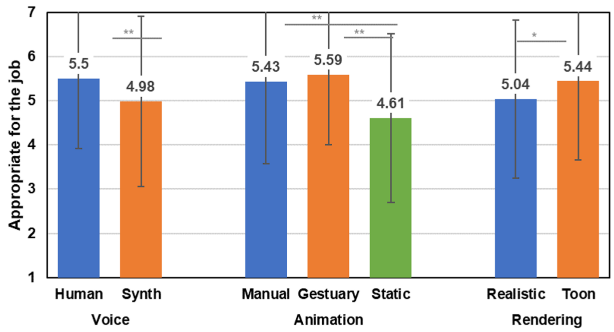

Figure 4: Means and standard deviations for the effect of voice, animation, and rendering on the appropriateness of the agent for her job.

There were significant effects of animation and voice on ratings of satisfaction with the agent. The manually-animated agent was rated significantly higher than the static one, M=5.08 (1.26) vs. M=4.44 (1.27), p<.01, as was the gestuary agent rated significantly higher than the static agent, M=4.96 (1.2) vs. M=4.44 (1.27), p<.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) vs. 4.57 (1.24), p<.01.

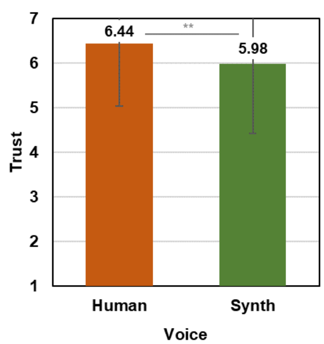

Figure 5: Means and standard deviations for the effect of voice on the participants' level of trust in the agent.

On trusting the agent, we found a significant main effect of voice (Figure 5). The agent with the human voice was rated significantly higher than the synthesized voice agent, M=6.44 (1.41) vs. M=5.98 (1.55), p<.01. Additionally, there was a significant main effect of voice and animation levels on how credible participants found the information (Figure 6). The information given by the agent with a human voice was rated more credible than the one with the synthetic voice, M=5.84 (1.18) vs. M=5.48 (1.40), p<.01. Similarly, the information delivered by the gestuary agent was regarded more credible than the information coming from the static agent, M=5.86 (1.13) vs. M=5.38 (1.39), p<.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

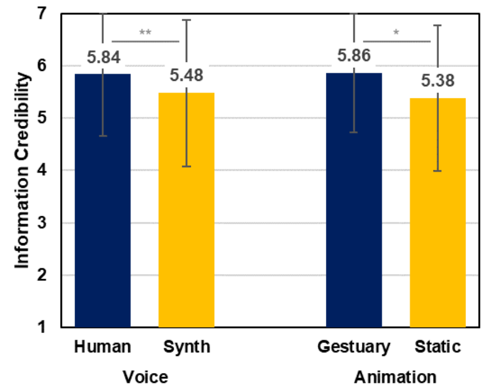

Figure 6: Means and standard deviations for the effect of voice and animation on the credibility of the information provided by the agent.

the agent-as-authoritative-counselor supposition as much as those less familiar with 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.

Table 6: Bivariate correlations from analysis of ordinal and ratio measures. ** indicates p-values < .01.

	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

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 (Figure 7). Participants in the static condition showed a significantly greater change, M=1.35 (2.33), than those in the gestuary, M=-0.04 (1.72) p<.01, and manually animated conditions, M=0.41 (1.82) p<.01, as well as a significantly greater change for participants in the manually-animated condition compared to those in the gestuary, p<.01. There were no significant differences between the levels of voice and rendering styles.

5 DISCUSSION

The manipulation checks for animation and voice quality showed that they were correctly perceived by our participants. When asked

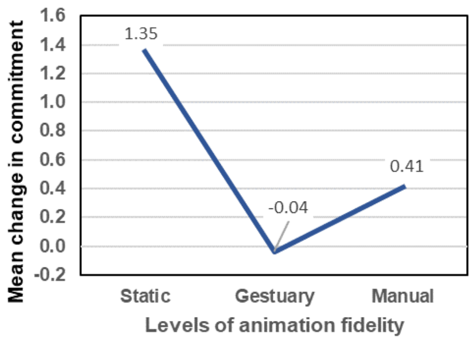

Figure 7: Persuasion: Means and standard deviations for the effect of animation on the change in participants' level of commitment to getting a health care proxy.

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 [21], 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. [20], 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 [19]) 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.

5.1 Limitations

Our study has several 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 [3], and so may not be representative of what would happen in longer interactions, or after users have established working relationships with the agent. The task of obtaining a health care proxy was likely not 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 CONCLUSION

In this work, we studied the impacts of animation, voice and rendering styles 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.

6.1 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. Our finding that animation can act as a distraction from the comprehension of key information warrants further investigation. Finally, we plan to investigate how these effects change over time in longitudinal tasks.

ACKNOWLEDGMENTS

This work was supported by the US National Institutes of Health grant R01NR016131.

REFERENCES

[1] Adobe. 2020. Adobe: Creative, marketing and document management solutions. https://www.adobe.com/. (2020). Accessed: 2020-02-20.

[2] Amazon. 2020. Amazon Mechanical Turk. https://www.mturk.com/. (2020). Accessed: 2020-02-20.

[3] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.

[4] Timothy W Bickmore, Rebecca A Silliman, Kerrie Nelson, Debbie M Cheng, Michael Winter, Lori Henault, and Michael K Paasche-Orlow. 2013. A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society* 61, 10 (2013), 1676–1683.

[5] Brigitte Bigi and Daniel Hirst. 2012. SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. In *Speech Prosody*. Tongji University Press, Shanghai, China, 19–22. https://hal.archives-ouvertes.fr/hal-00983699

[6] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. 2000. *Embodied conversational agents*. MIT press, Cambridge, MA, USA.

[7] Justine Cassell and Kristinn R Thorisson. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 4-5 (1999), 519–538.

[8] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2004. *BEAT: the Behavior Expression Animation Toolkit*. Springer Berlin Heidelberg, Berlin, Heidelberg, 163–185. https://doi.org/10.1007/978-3-662-08373-4_8

[9] Cereproc. 2020. CereProc text-to-speech. https://www.cereproc.com/. (2020). Accessed: 2020-02-20.

[10] BJ Fogg, Jonathan Marshall, Tami Kameda, Joshua Solomon, Akshay Rangnekar, John Boyd, and Bonny Brown. 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)*. Association for Computing Machinery, New York, NY, USA, 295–296. https://doi.org/10.1145/634067.634242

[11] Li Gong and Jennifer Lai. 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)*. Association for Computing Machinery, New York, NY, USA, 158–165. https://doi.org/10.1145/365024.365090

[12] Everlyne Kimani, Timothy Bickmore, Ha Trinh, Lazlo Ring, Michael K. Paasche-Orlow, and Jared W. Magnani. 2016. A Smartphone-Based Virtual Agent for Atrial Fibrillation Education and Counseling. In *Intelligent Virtual Agents*. Springer International Publishing, Cham, 120–127.

[13] H Chad Lane, Matthew Jensen Hays, Mark G Core, and Daniel Auerbach. 2013. Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology* 105, 4 (2013), 1026.

[14] Akinobu Lee and Tatsuya Kawahara. 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*. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, Sapporo, Hokkaido, Japan, 131–137. http://hdl.handle.net/2115/39653

[15] Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. 2012. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 91.

[16] Wade J Mitchell, Kevin A Szerszen Sr, Amy Shirong Lu, Paul W Schermerhorn, Matthias Scheutz, and Karl F MacDorman. 2011. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 1 (2011), 10–12.

[17] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[18] Clifford Nass and Li Gong. 1999. Maximized Modality or constrained consistency?. In *Auditory-Visual Speech Processing, AVSP '99, Santa Cruz, CA, USA, August 7-10, 1999*, Dominic W. Massaro (Ed.). ISCA, Santa Cruz, CA, USA, 1. http://www.isca-speech.org/archive_open/avsp99/av99_001.html

[19] Richard E Petty, Pablo Briñol, Lee Fabrigar, and Duane Wegener. 2010. Attitude Structure and Change. In *Advanced Social Psychology: The State of the Science*. Oxford University Press, USA, New York, NY, USA, Chapter 6, 217–259.

[20] Lazlo Ring, Dina Utami, editor="Bickmore Timothy Bickmore, Timothy", Stacy Marsella, and Candace Sidner. 2014. The Right Agent for the Job?. In *Intelligent Virtual Agents*. Springer International Publishing, Cham, 374–384.

[21] Ladan Shams and Robyn Kim. 2010. Crossmodal influences on visual perception. *Physics of life reviews* 7, 3 (2010), 269–284.

[22] Steven E Stern, John W Mullennix, Corrie-lynn Dyson, and Stephen J Wilson. 1999. The persuasiveness of synthetic speech versus human speech. *Human Factors* 41, 4 (1999), 588–595.

[23] Angela Tinwell, Mark Grimshaw, and Andrew Williams. 2010. Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds* 2, 1 (2010), 3–25.

[24] Unity. 2020. Unity Real-Time Development Platform. https://unity.com/. (2020). Accessed: 2020-02-20.

[25] Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2005. Building characters: Lessons drawn from virtual environments. In *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*. Cognitive Science Society, Stresa, Italy, 119–126.

[26] Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily, and Tracy Fasolino. 2016. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1326–1335.

[27] Matias Volonte, Andrew Robb, Andrew T Duchowski, and Sabarish V Babu. 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)*. IEEE, IEEE, Reutlingen, Germany, 25–32. https://doi.org/10.1109/VR.2018.8446364

[28] Robert B Welch, Theodore T Blackmon, Andrew Liu, Barbara A Mellers, and Lawrence W Stark. 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 (1996), 263–273.

[29] Lawrence R Wheeless and Janis Grotz. 1977. The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3, 3 (1977), 250–257.

[30] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 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)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[31] Yanxiang Wu, Sabarish V Babu, Rowan Armstrong, Jeffrey W Bertrand, Jun Luo, Tania Roy, Shaundra B Daily, Lauren Cairco Dukes, Larry F Hodges, and Tracy Fasolino. 2014. Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics* 20, 4 (2014), 626–635.

[32] 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)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/1240624.1240626

[33] Catherine Zanbaka, Paula Goolkasian, and Larry Hodges. 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)*. Association for Computing Machinery, New York, NY, USA, 1153–1162. https://doi.org/10.1145/1124772.1124945

[34] Katja Zibrek and Rachel McDonnell. 2014. Does Render Style Affect Perception of Personality in Virtual Humans?. In *Proceedings of the ACM Symposium on Applied Perception (SAP '14)*. Association for Computing Machinery, New York, NY, USA, 111–115. https://doi.org/10.1145/2628257.2628270