The Right Agent for the Job?

The effects of agent visual appearance on task domain

Lazlo Ring, Dina Utami, Timothy Bickmore

College of Computer and Information Science, Northeastern University, Boston, MA
{lring,dinau,bickmore}@ccs.neu.edu

Abstract. The visual design of virtual agents presents developers with a very large number of choices. We conducted a series of studies using Amazon's Mechanical Turk that demonstrate that there are no design universals for characters, optimal design of character proportion and rendering style depends on the task domain and user characteristics. Specifically, we found these adjustments to an agent's appearance directly effected how users rated it based on whether it was discussing social or medical content. The results of this research aim to help create visual guidelines for the development of domain specific virtual agents.

Keywords: Virtual Agents, Rendering Styles, Character Proportions

1 Introduction

There are many design decisions that must be made when creating a virtual agent for a new application and user demographic. These decisions range from the species of the agent (humanoid, animal, robot, etc.), genre of the character (anime, cartoon proportioned, realistic), the apparent role of the character, demographic parameters (gender, race, age), selection of clothing and accessories, hairstyle, and rendering style. Although several studies have attempted to systematically explore parts of this design space [17, 4] they have all constrained their investigations to single application domains, such as entertainment. However, the most appropriate character design for one domain is not necessarily the most appropriate for another. For example, a toon-shaded anime style character may be ideal in a social networking application but possibly inappropriate as the interface to a retirement planning system.

In this paper, we further investigate the visual design space for intelligent virtual agents, but include in our research the systematic manipulation of task type. This work is motivated by our experience building health counseling agents [21] in which it is usually not obvious whether more playful, cartoony agents would be preferred, because they may increase engagement, or more realistic agents are better, because of the seriousness of the health topics being discussed (e.g., chemotherapy protocols for cancer patients). We have made many design decisions based on small sample user studies or anecdotal feedback, and wanted a more systematic answer to the question of which character design is the most appropriate in any given situation.

Given our application domains of interest, we have limited our exploration to humanoid characters (no monkeys or parrots), hold behavioral realism (animation)

and character environment constant, and systematically explore rendering style and character proportions in correlation with genre of the character's design across task domains. Based on our experience and previous studies (Section 2) our primary hypotheses relate character realism and the levity/seriousness of a task domain as follows.

H1: Realism will be judged more appropriate for task domains high in seriousness.

H2: Frivolity (non-realism) will be judged more appropriate for domains low in seriousness.

In the remainder of this paper we briefly review related studies on character design and describe our experimental methodology, before presenting two design studies and conclusions.

2 Related Work

Several researchers have investigated the effects of visual design choices on user perceptions of a virtual agent. In this section we review the concept of the how realism, rendering style and character proportion have been shown to change people's attitudes towards virtual agents, and why these may be influenced by the task domain.

2.1 Effects of Rendering Styles:

Using the concept of the Uncanny Valley [18], McDonnell et al. investigated how different rendering styles affect how users perceive a 3D character [17]. Using ten points along the realism spectrum, ten shaders were created and applied to a 3D model of a human. Using these variants, two studies were conducted to investigate how users evaluated the different rendered models on social aspects such as friendliness, trustworthiness and appeal of the character. The results of this study showed that toon shaded and highly realistic models were best received by participants across the various social aspects, with the toon shaded version slightly outperforming the high quality version in the majority of comparisons.

2.2 Effects of Character Proportions:

The proportions of animated characters have also been explored as ways to understand and work around the uncanny valley. Kenn McDonald, a Sony Pictures Imageworks animator said that "A good way to avoid the uncanny valley is to move a character's proportions and structure outside the range of human.'" and attributed the success of Gollum from the Lord of The Rings and Grendel in Beowulf to their disproportion. His reasoning is that when viewers see the characters, they will think that they are not human and will not judge them by the same rule as if they were [11].

2.3 Effects of Realism:

Several researchers have studied the effects of realism on user perceptions. There are two kinds of realism studied: appearance realism and behavior realism. For appearance realism, Kang et al. [15] found that social co-presence is higher when dynamic high iconic avatars are used in mobile video telephones. For behavioral

realism, Garau et al. [10] and Bailenson et al. [3] found that a large mismatch between behavioral realism and appearance realism of avatars lowered social realism. Bailenson and Yee [4] also found that the more realistic the behavior of the agent, the more persuasive the agent is. Finally, Guadagno et al. [12] and Bailenson [2] found that social influence within immersive virtual environments is higher with virtual humans with high behavioral realism. Furthermore, in their study, the researchers found that this effect was moderated by the gender similarity between human and the avatar. In 2007, Yee et al. conducted a meta-analysis of 25 experimental studies of anthropomorphism, embodied agents, or agent realism and found that human-like representations with higher realism generated more positive subjective user ratings than representations with lower realism [24].

Researchers have also compared the effects of watching cartoons vs. videos of humans. Han et al. [14] showed using fMRI studies that different parts of the human brain are used when presented with cartoons compared to videos of real people. Chen, et al. demonstrated that exposure to cartoon video clip shifts preferences of human faces towards larger eyes [9].

2.4 Effects of Agent Appearance on Tasks:

Many studies have also investigated task-specific effects of different character designs. Several researchers have shown that the gender [5, 6, 12, 16] and race [5, 13, 19] of pedagogical agents have significant effects on a student's self efficacy and motivation. However these generalizations have been shown to be context dependent. For example, female agents are more effective in trying to convince students of the merits of engineering as a career regardless of user gender [1].

An agent's attractiveness, coolness and age have also been shown to be influential in pedagogical agents. While undergraduate female students are more likely to identify themselves with young, attractive, and cool female agents, they tend to choose to learn more about engineering from male agents that were attractive but uncool [7, 20, 1].

3 The Renderlab System

In order to conduct systematic investigations into the effects of agent appearance on user perceptions, we developed an online system integrated into Amazon's Mechanical Turk with real-time support to dynamically alter a 3D agent's graphical appearance. The platform uses a Unity-based 3D environment to render animated virtual agents over the web. The agents interact with users in brief dialogue sessions using a hierarchical state-machine-based dialogue engine, template-based text generation, conversational nonverbal behavior generation using BEAT [8], and synthesized speech. User contributions to the dialogue are made via a multiple-choice menu. Since task domain is a focus of our studies, we felt it was important that users engage in interactive dialogue rather than just passively listen to the agents. The Unity-based animation engine run in users' web browsers included support for the dynamic loading of Cg/HLSL shaders for the agent (3D rigged models) at runtime. A single set of animation files were also incorporated into the agent via Unity's Mecanim system to ensure that

there were no variations between the animations the agents performed between study conditions. All virtual agents used the same range of nonverbal behavior including: visemes and eyebrow raises synchronized to speech, head nods, facial displays of emotion, posture shifts, gazing at and away from the user, and idle behavior (blinking, etc.).

In each of the following studies, we created and used dialogue scripts for social interaction and medical counseling. Each dialogue was 6 to 10 turns long. The social interaction scripts discussed the user's favorite books and movies, while the medical scripts discussed two about cancer related topics (Table 1).

Table 1: Sample Dialogue Excerpts

	Medical Dialogue
Agent:	Hi, today I would like to talk to you about the importance of having regular colorectal screening for cancer.
User:	Go on.
Agent:	Screening is the process of looking for cancer in people who have no symptoms of the disease.
User:	Sure.
Agent:	Colorectal cancer is the third most common cancer diagnosed and is the third leading cause of cancer-related deaths in the United States... It also allows more colorectal cancers to be found earlier, when the disease is easier to cure.
	Social Dialogue
Agent:	Hi, do you like watching movies?
User:	Yes.
Agent:	Great! Me too!
User:	Sure.
Agent:	So, what kind of movie do you like?

Common Measures: In both of the following studies, we assessed user attitudes towards the agent using ten 7-point Likert-scale self-report questions following each interaction with an agent. The items assessed were: *realism, appeal, familiarity, eeriness, friendliness, trustworthiness, easiness to interact with, desire to continue working with, likability, caring, appropriateness, and the quality of motion*. Two open-ended questions were also given, one asking the user how they would describe the character appearance and one for general comments about the agent.

4 Experiment 1: Shading Styles

In our first experiment, we sought to replicate part of McDonnell's Render Me Real? study by investigating the impact of rendering style on user perceptions, but in two different task domains.

4.1 Methods:

We selected commercial shaders to match two of the conditions used in the Render Me Real? study (Human High Quality 1 we refer to as Realistic, and (Toon Shaded) as closely as possible for a human-proportioned character (Figure 1). To create the Toon Shaded version of the model, the MatCaps shader library

from the Unity assest store was integrated into the client for real time render support.

Fig. 1: Screenshot of Realistic (left) and Toon Shaded (right) agent

4.2 Participants:

Participants were recruited on Amazon's Mechanical Turk for a counterbalanced, within-subjects experiment in which they interacted with four variants of the agent, Shaded-Social, Realistic-Social, Shaded-Medical and Realistic-Medical.

4.3 Results:

Sixty-seven participants (36 Males, 31 Females) successfully completed the study, resulting in a total of 268 agent interactions (4 interactions per user). A 2x2 (rendering style vs. task) repeated measures ANOVA was performed using the ex package in R. Table 2 and 3 show the main results and interaction effects of the study.

Table 2: Main Effects in Study 1

	Medical		Social		p-value	
	Shaded	Realistic	Shaded	Realistic	Rendering	Dialogue
Realistic	4.24(1.59)	4.76(1.44)	4.03(1.68)	4.39(1.47)	<0.01	<0.01
Friendly	4.96(1.34)	4.88(1.31)	5.48(1.19)	5.07(1.48)	0.04	0.02
Familiar	4.6(1.82)	4.99(1.68)	4.2(1.92)	4.31(1.84)	0.18	<0.01
Trustworthy	5(1.4)	4.93(1.51)	4.58(1.35)	4.46(1.47)	0.43	<0.01
Appropriate	5.37(1.48)	5.49(1.39)	4.87(1.57)	5.15(1.53)	0.15	<0.01
Desire to Cont.	4.45(1.73)	4.36(1.77)	4.48(1.8)	4.24(1.87)	0.24	0.76
Likeable	4.42(1.6)	4.49(1.73)	4.72(1.6)	4.25(1.78)	0.19	0.82
Caring	4.13(1.83)	4.25(1.84)	4(1.83)	3.76(1.73)	0.66	0.06

4.4 Manipulation Check:

The *realism* question was used as a manipulation check. This test was found to be significant in the expected direction, with users rating the Realistic agent

significantly higher on the *realism* question compared to the Shaded version of the agent $F(1,66) = 11.10$, $p < .01$.

Table 3: Interaction effect beween dialogue condition vs rendering style in Study 1

	Interaction p-value
Friendly	0.16
Familiar	0.44
Appropriate	0.56
Desire to Continue	0.46
Likeable	0.03
Caring	0.08

4.5 Outcome Analysis:

We found main effects of task on appropriateness, $F(1,66) = 7.83$, trustworthiness, $F(1,66) = 13.77$, and familiarity, $F(1,66) = 14.97$, $p < .01$, with these factors being rated higher for medical task compared to the social task. Two main effects were also found on friendliness, with the agent being rated as more friendly in the social task $F(1,66) = 6.97$, $p < .05$, and the shaded version, $F(1,66) = 4.33$, $p < .05$.

A significant interaction of task and rendering style was found on likeability, $F(1,66) = 5.22$, $p < .05$, with the shaded agent being rated as more likeable than the the realistic agent, but only for the social task. The interaction of task and rendering style on caring was found to be trending towards significance, $F(1,66) = 3.18$, $p = 0.08$, with the realistic agent being rated as more caring for the medical task, and the shaded agent being rated as more caring for the social task.

4.6 Discussion:

This experiment demonstrated that there are significant effects of agent rendering on a user's impressions of it, particularly in social tasks. We found that for social dialogue the Toon Shaded agent was rated as being more likeable and caring. This finding replicates the results of McDonnell's study, in which the more cartoon like character was rated higher than the realistic ones on these measures. However, in the medical task we only found this result for friendliness. One possible explanation for this is that the medical scripts were more task oriented than the social dialogue, in which there was a clear primary purpose to the dialogue (education in this case).

5 Experiment 2: Character Proportions

In our second study we explored the effects of character proportions (cartoony vs. realistic, Figure 2) on user attitudes across tasks.

5.1 Methods:

We used a modified version of the Toon-Shaded agent from Experiment 1, and compared it to a cartoon-proportioned character that was otherwise equivalent in dress, hairstyle and skin tone (Figure 2).

Fig. 2: Screenshot of Human (left) and Cartoon (right) proportioned agents

5.2 Results:

Forty-seven participants (31 Males, 16 Females) participated in this study, resulting in a total of 188 agent interactions. A 2x2 (rendering style vs. task) repeated measures ANOVA was performed using the ex package in R. Table 4 shows the main results of the study.

Table 4: Main Effects for Study 2

	Medical		Social		p-value	
	Shaded	Realistic	Shaded	Realistic	Rendering	Dialogue
Realistic	4.82(1.48)	4.23(1.73)	4.71(1.69)	4.15(1.98)	<0.01	0.49
Friendly	4.58(1.23)	5.06(1.2)	4.86(1.77)	5.36(1.31)	0.02	0.06
Familiar	4.71(1.56)	4.85(1.62)	4.54(1.73)	4.4(1.73)	0.91	0.05
Trustoworthy	5.04(1.28)	5.21(1.17)	4.67(1.53)	4.91(1.27)	0.22	0.02
Appropriate	5.36(1.37)	5.08(1.57)	4.67(1.64)	5.21(1.57)	0.59	0.08
Desire to Cont.	4.42(1.73)	4.96(1.65)	4.54(1.75)	4.70(1.59)	0.29	0.55
Likeable	4.47(1.63)	4.77(1.61)	4.54(1.75)	4.70(1.59)	0.40	0.94
Caring	4.40(1.71)	4.67(1.74)	4.06(1.74)	4.23(1.80)	0.31	<0.01

5.3 Manipulation Check:

Significance was found for how realistic the agent was in the predicted direction, with the human proportioned character being rated as more realistic than the cartoon proportion character $F(1,47) = 7.23$, $p < .01$.

5.4 Outcome Analysis:

Similar main effect results were found for familiarity, trustworthiness, appropriateness and caring for the social and medical tasks as compared to Experiment 1. The minor discrepancy in significance between the two experiments may be due to having fewer subjects in Experiment 2 compared to Experiment 1. For friendliness however, an additional main effects of cartoon proportioned was found, $F(1,46) = 6.94$, $p < .05$, in which the cartoon proportioned character were rated as being significantly friendlier than the human proportioned character.

A significant interaction of task and character proportion was also found on approriateness $F(1,46) = 7.12$, $p < .05$, with participants rating the realistic agent as being significantly more appropriate for the medical task, while rating the cartoon proportioned character as being significantly more appropriate for the social task (Figure 3).

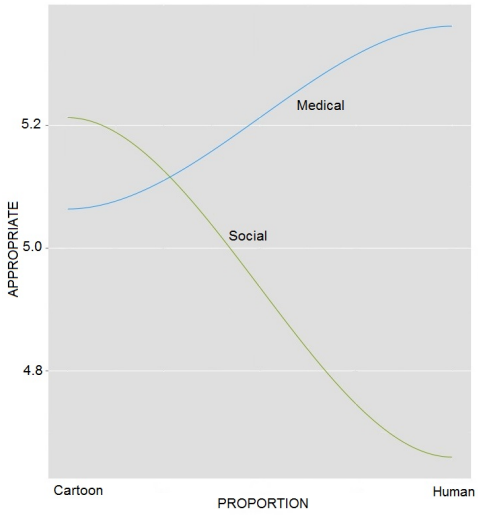

Fig. 3: Interaction effect of agent proportion and appropriateness

5.5 Discussion:

This experiment suggests that the design rules for the visual appearance of an agent may not be universal, but depend on the application domain. Although participants found the cartoon-proportioned character to be friendlier regardless of dialogue content, they found it more appropriate for the human proportioned character to talk to them about medical content. The effect of character proportions on friendliness is supported by character design heuristics that specify that larger-eyed characters can more easily express emotion [22]. The interaction of proportion and task on appropriateness, however, may be caused by participants' mental model of what they think of as a medical professional, which is most likely not a cartoon character.

6 Conclusion

In this paper we explored the effects of an agent's appearance and application domain on user perceptions of the agent. We found partial support for our hypotheses relating agent realism and task seriousness on user perceptions of the agent. Specifically, we found that changes in an agent's appearance effected how users rated its friendliness, likability, caring, and appeal depending on the content of its dialogue.

For our first experiment we investigated the effects of manipulating an agent's rendering styles, comparing toon shaded and realistic looking agents. This experiment found that toon shaded characters were rated as being more likable, and caring when compared to realistic characters in social task contexts.

In our second experiment we looked at changing an agent's proportions, comparing a human and cartoon proportioned character. Through this investigation we found that cartoon proportioned characters were rated as being more friendly regardless of task domain, but found that more realistic characters were rated as being more appropriate for medical tasks.

The findings from these studies suggest designing an agent may not be as simple as make the most realistic or cartoony agent possible. Our results suggest that a purely medical system a highly realistic agent may be a better design, whereas for a social system a cartoon like agent may work better.

Our studies have many limitations, including the relatively small convenience samples recruited on Mechanical Turk that may not generalize to any particular user demographic for a target application. We have also only explored a tiny corner of the very large space of design parameters for virtual agents. In future studies we aim to further explore this space by looking at various other graphic manipulations such as lightning and color, and also investigate how these effects change over time in longitudinal tasks.

References

1. Ashby Plant, E., Baylor, A.L., Doerr, C.E., Rosenberg-Kima, R.B.: Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. Computers & Education 53(2), 209–215 (2009)
2. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Interpersonal distance in immersive virtual environments. Personality and Social Psychology Bulletin 29(7), 819–833 (2003)
3. Bailenson, J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J.: The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. Presence: Teleoperators and Virtual Environments 14(4), 379–393 (2005)
4. Bailenson, J.N., Yee, N.: Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. Psychological science 16(10), 814–819 (2005)
5. Baylor, A.L., Kim, Y.: Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In: Intelligent Tutoring Systems. pp. 592–603. Springer (2004)
6. Baylor, A.L., Kim, Y.: Simulating instructional roles through pedagogical agents. International Journal of Artificial Intelligence in Education 15(2), 95–115 (2005)
7. Baylor, A.L., Rosenberg-Kima, R.B., Plant, E.A.: Interface agents as social models: the impact of appearance on females' attitude toward engineering. In: CHI'06 Extended Abstracts on Human Factors in Computing Systems. pp. 526–531. ACM (2006)
8. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: Life-Like Characters, pp. 163–185. Springer (2004)

9. Chen, H., Russell, R., Nakayama, K., Livingstone, M.: Crossing the "uncanny valley": adaptation to cartoon faces can influence perception of human faces. Perception 39(3), 378 (2010)
10. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., Sasse, M.A.: The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 529–536. ACM (2003)
11. Geller, T.: Overcoming the uncanny valley. IEEE Computer Graphics and Applications 28(4), 11–17 (2008)
12. Guadagno, R.E., Blascovich, J., Bailenson, J.N., Mccall, C.: Virtual humans and persuasion: The effects of agency and behavioral realism. Media Psychology 10(1), 1–22 (2007)
13. Gulz, A., Haake, M., Tärning, B.: Visual gender and its motivational and cognitive effects–a user study. Lund University Cognitive Studies 137 (2007)
14. Han, S., Jiang, Y., Humphreys, G.W., Zhou, T., Cai, P.: Distinct neural substrates for the perception of real and virtual visual worlds. NeuroImage 24(3), 928–935 (2005)
15. Kang, S.H., Watt, J.H., Ala, S.K.: Social copresence in anonymous social interactions using a mobile video telephone. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1535–1544. CHI '08, ACM, New York, NY, USA (2008), http://doi.acm.org/10.1145/1357054.1357295
16. Kim, Y., Baylor, A.L., Shen, E.: Pedagogical agents as learning companions: The impact of agent emotion and gender. Journal of Computer Assisted Learning 23(3), 220–234 (2007)
17. McDonnell, R., Breidt, M., Bülthoff, H.H.: Render me real?: investigating the effect of render style on the perception of animated virtual humans. ACM Transactions on Graphics (TOG) 31(4), 91 (2012)
18. Mori, M.: The uncanny valley. Energy 7(4), 33–35 (1970)
19. Pratt, J.A., Hauser, K., Ugray, Z., Patterson, O.: Looking at human–computer interface design: Effects of ethnicity in computer agents. Interacting with Computers 19(4), 512–523 (2007)
20. Rosenberg-Kima, R.B., Baylor, A.L., Plant, E.A., Doerr, C.E.: The importance of interface agent visual presence: Voice alone is less effective in impacting young women's attitudes toward engineering. In: Persuasive Technology, pp. 214–222. Springer (2007)
21. Schulman, D., Bickmore, T.W., Sidner, C.L.: An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In: AAAI Spring Symposium: AI and Health Communication (2011)
22. Thomas, F., Johnston, O., Thomas, F.: The illusion of life: Disney animation. Hyperion New York (1995)
23. van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Veldhuis, J.: Why fat interface characters are better e-health advisors. In: Intelligent Virtual Agents. pp. 1–13. Springer (2006)
24. Yee, N., Bailenson, J.N., Rickertsen, K.: 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. pp. 1–10. ACM (2007)