Evaluating Teacher Avatar Appearances in Educational VR

Jason W. Woodworth   Nicholas G. Lipari   Christoph W. Borst

University of Louisiana at Lafayette

Figure 1: The four avatars considered in the pilot study. From left to right: Kinect-Video, Human-Model, Robot-Model, and Hybrid.

Abstract

We present a pilot study of four teacher avatars for an educational virtual field trip. The avatars consist of a depth-video-based mesh of a real person, a game-style human model, a robot model, and the robot with its head replaced by a video feed of the teacher's face. Multiple avatars were developed to consider alternatives to the mesh representation that required high-bandwidth networks and a non-immersive teacher interface. The pilot study presents a random avatar to the participant at each of 4 educational stations, and follows up with a subjective questionnaire. Most notably, we find positive affinity for the plain robot model to be similar to that of the video mesh, which was previously shown to provide high co-presence and good results for education. Results are guiding a larger study that will measure the educational efficacy of revised avatars.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

The design of a pedagogical avatar can have a substantial impact on the efficacy of an agent-driven educational application [5]. Important aspects include avatar realism [1], color [3], and enthusiasm [6]. We address the design of a teacher avatar for a collaborative VR system that allows a non-immersed teacher to guide immersed students through an environment, akin to a virtual field trip.

The virtual field trip system [2] was initially developed with a single avatar: a 3D RGBD-based mesh rendered from videos captured by a Kinect V2, intended to present a realistic view of the teacher with high co-presence for students [4]. Instruction could be presented in a live format or in prerecorded form with students triggering clips. Student assessments of the avatar were positive, showing very high co-presence and affective attraction. However, the avatar required high network and rendering performance and a non-immersive teacher interface, and it is unknown how critical the avatar appearance was to the experimental results. The need to send depth data losslessly and to maintain a fast audio response rate resulted in up to 20 Mbit bandwidth. The desire to have an unobstructed view of the teacher's face was one factor in the design of a non-immersive teacher interface requiring a Kinect and a large display. This motivates us to understand which aspects of the teacher avatar are most important, and to seek other avatar styles that impose fewer requirements while being similarly effective.

Two factors considered for the design of additional avatars were likeness to a human and showing the original teacher. We designed 3 additional avatars considering different values of these factors: a human model, a robot model, and the robot with its head replaced by a video screen showing the original teacher's head. We present initial findings from a pilot study intended to determine what aspects of the teacher's avatar are subjectively important to students.

2 Avatars

We included four avatars in the pilot study (seen in Figure 1). Each avatar's movement is based on a recording from a Microsoft Kinect; recordings for each avatar were taken simultaneously, meaning every recording can be played back with any avatar.

Kinect-Video: A 3D mesh built per frame from color and depth videos captured by the Kinect as described by Ekong et al. [4]. This avatar is the most realistic, shows imagery of the real teacher, and is our baseline, because its educational efficacy and subjective impression were previously studied [2]. We included techniques for the teacher to maintain eye gaze toward a desired point [4].

Human-Model: A standard game-type human model made in MakeHuman to have similar dimensions to the real-life teacher. It is less realistic and less like the original teacher than the Kinect-Video,

This is an author-formatted version. Original publication: J. W. Woodworth, N. G. Lipari, and C .W. Borst, "Evaluating Teacher Avatar Appearances in Educational VR", IEEE Virtual Reality 2019; Osaka, Japan, 2019

but there is still a resemblance. To simulate talking motion, the model's jaw moves according to the volume of the teacher's audio.

Robot-Model: A robot model that matches the industrial atmosphere of the environment. It does not appear human and bears minimal resemblance to a real teacher. Because the model lacks a mouth, its eyes pulse based on the volume of the teacher's audio.

Hybrid: The same robotic model with the head replaced by a video screen of the actual teacher's face. The video is taken from the Kinect's color video and is synchronized so that the face always appears in the center of the screen. The Hybrid is not overall human-like, but the video head bears resemblance to the real teacher and therefore may provide some helpful presence or facial cues.

3 User Study

We conducted a within-subjects pilot study to gather opinions about the four avatars, with the results intended to guide a larger study to judge each avatar's effect on educational results. Thirteen subjects participated, with 8 being from a local computer graphics class. Subjects were immersed with an Oculus Rift CV1 and were placed in a virtual solar energy plant shown in Figure 1.

In this study, subjects traveled between four educational stations with a different avatar presented at each station. Stations were always visited in the same order, but the order of the avatars was randomized to avoid order effects. Each avatar was shown playing two presentation clips, triggered by the subject, totalling approximately two minutes of educational material per avatar.

After a subject finished with one avatar, they answered brief in-game prompts about their perceptions of the avatar regarding clarity of pointing, presence, and overall likability. After a subject finished with the entire experience, they answered a written questionnaire to explain their reasons for liking or disliking any particular avatars and to describe any distracting elements of the avatars.

4 Results

We review subjective results to determine if any aspects of the avatars should be changed before a full experiment is conducted. When asked which avatar subjects preferred the most, 5 chose the Kinect-Video, 5 chose the Robot-Model, 2 chose the Hybrid, and 1 chose the Human-Model. Four proponents of the Kinect-Video remarked that it was the most realistic, and thus took them out of the experience the least. Four proponents of the Robot-Model remarked that it fit the environment (a solar energy plant) the best, and 2 stated it had the most understandable pointing.

When asked which avatar was least liked, 7 subjects chose the Hybrid, 4 chose the Human-Model, 2 chose the Kinect-Video, and none chose the Robot-Model. All who disliked the Hybrid remarked that it felt strange and unnatural, and 2 noted that the head monitor floating above the body was distracting. Two who disliked the Human-Model specifically mentioned the uncanny valley, and 2 stated it had unclear pointing and mouth movements. The 2 who disliked the Kinect-Video remarked that the image was unclear and that white edges of the mesh were distracting. When asked if anything about the avatars distracted from the educational experience, 3 reported minor strange avatar-environment interactions (avatars interpenetrating objects or floating), 3 reported unnatural Robot-Model movement, and 5 reported the white edges of the Kinect-Video.

Subjects answered in-game questions after exposure to each avatar to rate that avatar on 7-point Likert-style questions. The questions asked how well pointing could be understood, how much it felt like the avatar was standing in front of users, and how much subjects liked the avatar in general (with 7 being most likable). Average and median scores are presented in Table 1. Notably, the Kinect-Video and Robot-Model avatars share similar results for all questions. The Human-Model and Hybrid share similar means, slightly lower than Kinect-Video and Robot-Model means. The Hybrid has lowest median likability of 2, suggesting a skewed distribution with mostly low numbers. In combination with the earlier most/least-liked ratings, the hybrid and human models do not seem subjectively promising.

Table 1: Mean and median scores from 7-point Likert-style questions asked after exposure to each avatar.

Question	Kinect-Video		Human		Robot		Hybrid	
	Mn	Md	Mn	Md	Mn	Md	Mn	Md
Pointing	6.25	7	5.8	6	6.6	7	5.8	6
Presence	5.6	6	5.5	6	5.6	6	5	4
Likability	4.6	4	3.6	4	4.6	4	3.6	2

5 Conclusion and Future Works

The popularity of the Kinect-Video was expected, due to the positive responses in a prior study [2]. But, it was not expected that the Robot-Model would receive a similarly positive response. Judging by written responses, we infer that an avatar's "fit" with the environment may be more important in the pilot study's ratings than other factors we had considered.

Future studies will determine the educational impact of these avatar styles, and see if favorability or other factors translate into educational results. It is possible that the added realism of the Kinect-Video is not necessary once the best alternative is found, which would allow the use of a simpler interface and network.

Considering feedback given in this pilot, the Kinect-Video will be improved by recording with a solid dark-colored background to remove the white edges, which may give it a subjective advantage over the Robot-Model. The Human-Model and Hybrid avatars were judged to have a strange appearance, which we will attempt to mitigate before further study, potentially by improving mouth movement on the human avatar and remodeling the head for the Hybrid.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 1815976 and 1451833, and additionally by CNS-1531176 through the Mozilla Gigabit grant program.

References

[1] A. L. Baylor and Y. Kim. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, and F. Paraguaçu, editors, *Intelligent Tutoring Systems*, pages 592–603, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[2] C. W. Borst, N. G. Lipari, and J. W. Woodworth. Teacher-guided educational vr: Assessment of live and prerecorded teachers guiding virtual field trips. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 467–474, March 2018.

[3] I. X. Dominguez and D. L. Roberts. Asymmetric virtual environments: Exploring the effects of avatar colors on performance. In *2015 Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2015.

[4] S. Ekong, C. W. Borst, J. Woodworth, and T. L. Chambers. Teacher-student vr telepresence with networked depth camera mesh and heterogeneous displays. In *International Symposium on Visual Computing*, pages 246–258. Springer, 2016.

[5] S. Greenwald, A. Kulik, A. Kunert, S. Beck, B. Frohlich, S. Cobb, S. Parsons, N. Newbutt, C. Gouveia, C. Cook, et al. Technology and applications for collaborative learning in virtual reality. In *2017 International Conference on Computer Supported Collaborative Learning*. International Society of the Learning Sciences, 2017.

[6] T. W. Liew, N. A. Mat Zin, and N. Sahari. Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences*, 7(1):9, May 2017.