Can Immersive Virtual Humans Teach Social Conversational Protocols?

Sabarish Babu, Evan Suma, Tiffany Barnes, Larry F. Hodges

Department of Computer Science

University of North Carolina at Charlotte

ABSTRACT

We investigated the effects of using immersive virtual humans to teach users social conversational verbal and non-verbal protocols in south Indian culture. The study was conducted using a between-subjects experimental design, and compared instruction and interactive feedback from immersive virtual humans against instruction based on a written study guide with illustrations of the social protocols. Participants were then tested on how well they learned the social conversational protocols by exercising the social conventions in front of videos of real people. The results of our study suggest that participants who trained with the virtual humans performed significantly better than the participants who studied from literature.

CR Categories and Subject Descriptors: K.3 Computers and Education, I.6 Simulation and Modeling, H.5 Information Interfaces and Presentation, I.3 Computer Graphics.

Additional Keywords: Virtual Characters, Embodied Agents, Multimodal Interaction, Human-Computer Interaction, Immersive Virtual Environments

1 INTRODUCTION

1.1 Motivation

Virtual humans have the potential to engage and train human users in tasks that involve interpersonal verbal and non-verbal conversational behaviors in face-to-face social contact. To quantify the potentials of using immersive virtual humans to train users in social verbal and non-verbal behaviors, we investigated the following two questions:

1. Is it possible to train users in the verbal and non-verbal behaviors associated with real life social conversational protocols through natural multimodal interaction with immersive virtual humans?
2. How effective is the use of immersive virtual humans as a tool for training users in social conversational protocols as compared to a text-based approach using a written study guide with illustrations?

In particular, this study explores the use of immersive virtual humans to train users in social conversational behaviors pertaining to conversation initiation and disengagement in south Indian culture. Social conversational protocols in south Indian culture are highly structured and specific to the gender, age, and status of the interlocutor. The temporality, intensity, and synchronicity of the verbal greetings, non-verbal gestures, and eye gaze are well-defined by the rules of etiquette for conversation initiation and disengagement [1]. Learning of these social protocols for most people comes through social grounding, interactive feedback, and reinforcement from an immersive experience in the culture.

Our hypothesis: Natural multi-modal interaction with immersive virtual humans can successfully train naïve users in south Indian social protocols.

In order to evaluate our hypothesis we performed a study where we compared natural multi-modal interaction with immersive virtual humans to reading a written study guide with illustrations of social protocols. Participants in both conditions were given equal amounts of training time, and were then asked to demonstrate their ability to greet and say goodbye in response to video presentations of real south Indians.

1.2 Previous Work

1.2.1 Virtual Humans in Training and Pedagogy

Anecdotal evidence suggests that human communication consists of a high bandwidth of modalities such as gestures, facial expressions, speech, and body language [2]. In addition, researchers have found that users can learn a task from demonstrations far more effectively than learning to perform a task from text-based instructions alone, especially when that task involves spatial motor skills [3]. Many virtual human interfaces have been developed for training, pedagogy, and education that provide feedback to human users using multiple verbal and nonverbal channels such as speech, gestures, and facial expressions [2, 3, 4, 7].

1.2.2 Social effects of Virtual Humans

Researchers have investigated how people respond to computers and virtual humans. Nass and Moon have shown that people react to and attribute very human characteristics to computers - such as the computer's helpfulness, expertise, and friendliness [6]. Using a virtual human interface minimizes the need for training users since they already know how to interact with other people [7]. Mel Slater's group at UCL has conducted studies on the effects of social ramifications of having avatars in virtual environments. They found that the presence of avatars was significant for social interaction and task performance [8]. Raij et. al. examined perceived similarities and differences in experiencing an interpersonal scenario with a real and virtual patient [9]. They found lower ratings on participants' rapport and conversational flow with the virtual patient which was attributed to the limited expressiveness of the virtual patient. Level of immersion and natural interaction also facilitated the participants' ability to perform a training task with a virtual patient as effectively as with a real patient.

1.2.3 Virtual Humans in Social Conversational Protocol Training in a Foreign Culture

We have found little work that directly focuses on using virtual humans in training users in performing verbal and non-

{sbabu, easuma, tbarnes2, lfhodges}@uncc.edu

IEEE Virtual Reality Conference 2007
March 10 - 14, Charlotte, North Carolina, USA
1-4244-0906-3/07/$20.00 ©2007 IEEE

verbal behaviors in a foreign culture. The existing training systems described in the literature are focused on training users in social communication skills that are primarily verbal. The research that is closest to ours is the Virtual Environment for Operational Readiness (VECTOR) [10]. This system trains solders in the critical communication skills for survival and mission success. The effects of the trainee's positive and negative conversational input with indigenous virtual Iraqi civilians result in appropriate behavioral outcomes (such as hostile, helpful, worrisome etc.,) based on cultural expectations. Our research focuses on training users in performing the standard verbal and non-verbal (gestures and gaze) social conversational conventions using virtual reality technology with life size virtual humans and multi-modal interaction.

2 The Virtual Reality Setup

2.1 System Overview

The immersive virtual reality social conversational protocol training system implemented using VHIF [11] was housed in an office where participants could experience training with the virtual characters with no one else present (Figure 1).

Our system used two networked PCs. One handled speech and gesture recognition. The second PC handled the visual rendering component of the system. A data projector was used to display the virtual humans at life-size. The rendering PC was an Alienware Aurora with a duel nVidia 7900 GT SLI graphics card. The virtual humans were rendered at 40-45 FPS. Figure 2 shows a schematic of the hardware infrastructure of our system.

Figure 1: Shows a participant (right) greeting Anita, a virtual south Indian of the same gender (left), while Radha, the virtual south Indian instructor (middle) observes and shows the trainee how the greeting is performed.

The virtual humans were projected on a large screen in front of the participant. The dimensions of the projected image measured 1.8 meters in width by 1.35 meters in height. The participant stood two meters from the projection screen. The trainee's head, hands, and waist were tracked in 6DOF using a Polhemus Fastrack electro-magnetic tracker. This tracking data served as input for non-verbal gesture and gaze recognition (Figure 2). Additionally, the head tracker data was also used in rendering the image according to the correct perspective warping effects. Speech input was taken through a microphone attached to a head band. Audio output of the system was provided by speakers positioned on either side of the screen. Participants were requested to stand on a marked spot on the floor facing the screen when training with the virtual humans. A student-apprentice pedagogical model was chosen as the basis of the instruction, similar to Johnson et al. [12]. The virtual south Indian rendered on the right (figure 1), who is also the same gender as the participant, acts as the virtual instructor. The virtual south Indian rendered on the left (figure 1), acts as a conversational partner for the user, with whom the user practices the conversational protocols while observed by the instructor.

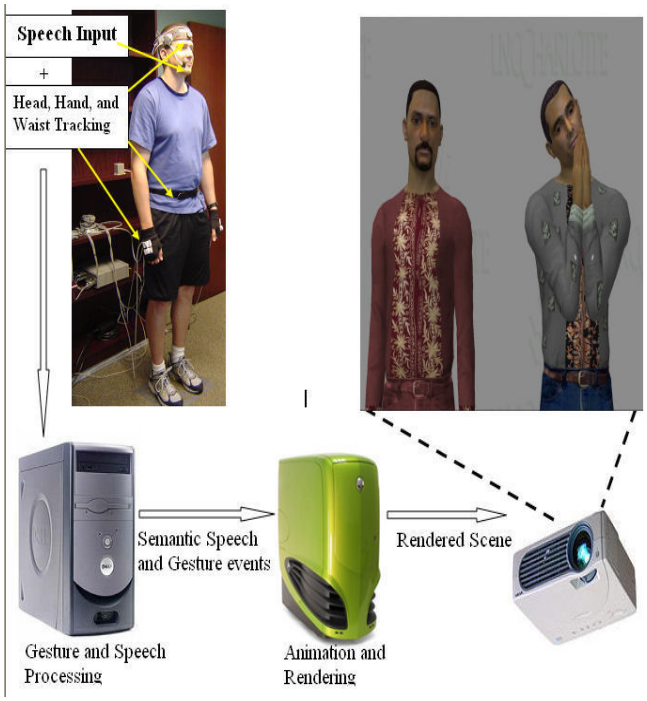

Figure 2: System Hardware Infrastructure

2.2 Virtual Human Instruction and Interactive Feedback

2.2.1 Pedagogical Instruction

The immersive virtual human social conversational protocol training system presented two life-size virtual humans to the user. Female participants were presented with a female virtual instructor (Radha), depicted on the right of the screen. Male participants were presented with a male virtual instructor (Sameer). The content of the instructions came from literary sources describing south Indian social customs, norms, and etiquette [1]. Four conversational protocol tasks were chosen:

1. Greeting someone of the same gender
2. Saying goodbye to someone of the same gender
3. Greeting someone of the opposite gender
4. Saying goodbye to someone of the opposite gender

The sequence and presentation of each protocol type instruction were carefully designed to ensure consistent descriptions of each step. Each task description consisted of a sequence of five steps. The instructions provided by the virtual south Indian instructor in the virtual reality condition (VR) were also consistent with the instructions in the handout provided in the literature condition (L). In condition VR, the virtual instructor also demonstrated each step of the protocol to the participant using animated gestures, speech, and facial expressions. Images of the virtual instructor's non-verbal behaviors were also included along with the instructions in the literature provided to participants in condition L (Table 1). Each participant learned a

total of 20 steps (4 tasks x 5 steps) consisting of verbal and non-verbal behaviors.

A sample protocol excerpt of verbal and non-verbal instructions taken from the instruction manual provided to female participants in condition L is shown in Table 1 for a female participant greeting a person of the opposite gender.

Table 1: Sample conversational protocol for the task of greeting a person of the opposite gender, taken from the instruction manual provided for female participants. Each step is described on the left, and an image of the virtual instructor performing the step is shown on the right.

Description	Illustration
1. Stand facing the woman at two arms length. During the greeting the posture should be erect; both feet should be placed together.	
2. Gaze at the other person's mouth.	
3. Place both hands together in front of your neck with the tip of your fingers touching your chin. The elbows should be kept close to your body, and your arms should be tucked close to your chest.	
4. Smile and say "namaste" (nam-as-TAY), while bowing your head slightly forward, and shifting your gaze to the other person's feet. And then bring your head back to the normal position.	
5. Wait with your hands together, until the other person has completed greeting you similarly. Then bring your hands down, and relax your posture.	

2.2.2 Virtual Human Response and Feedback

For every social conversation task, such as greeting a person of the same gender, the virtual instructor first describes the steps involved in the task using a combination of verbal speech instructions as well as non-verbal gestures and facial expressions in a step-by-step manner. A male or female virtual partner either enters the scene or leaves depending on the nature of the social conversational protocol task being currently described. Every participant in the study learned to greet and say goodbye to someone of the same gender and opposite gender. The order of instruction of the four tasks was randomized. For each social conversational protocol, the virtual instructor provides instruction and interactive feedback after the user practiced the protocol with the virtual conversation partner, three times.

2.2.3 Gesture and Gaze Processing

The participant's head, hands, and waist were tracked in 6DOF using a Polhemus Fastrack electro-magnetic tracker. The head tracked data were used to control the gaze direction of the virtual humans. Participant head gaze was detected by finding the intersection of a ray from the tracked head to a virtual sphere representing the head of the virtual conversation partner. Hand clasping gestures were detected by the surface intersections of tracked virtual objects that corresponded to the position and orientation of the participant's real hands.

3 Study Design

A study was conducted between two conditions to determine the effectiveness of using immersive virtual humans in teaching users verbal and non-verbal social conversational protocols, as compared to learning the protocols from a study guide with illustrations (a non-technological, traditional learning approach). Participants were randomly assigned to one of two conditions:

- **Condition L**: Participants were provided an 8-page study guide with illustrations.
- **Condition VR**: Participants received instructions from a virtual south Indian instructor and practiced with a virtual conversation partner.

A total of 40 participants completed the study, with 20 participants in each condition. Participants were required to be of non-Asian culture and able to communicate comfortably in English.

3.1 Experiment Procedure

The pre-experiment session, training session, testing session, and post-experiment session took each participant approximately one hour to complete.

3.1.1 Pre-Experiment Session

The participant in each condition first read the Participant Information Sheet and was asked if he/she had any questions. The participant then read and signed the Informed Consent Form.

3.1.2 Training Session

The participants were then trained in social conversational protocols based on their assigned condition.

Condition VR:

Participants were fitted with the electro-magnetic tracking equipment for head, hands, and waist. The trackers were affixed to a pair of gloves, a head band, and a belt. Each participant then trained the speech recognition program by reading from a short passage. The participant was then told that he/she would now be trained in south Indian social protocols by virtual humans, and that the training session would last for approximately 20 minutes. During the training session, the participant was left alone in the experiment room with the virtual reality training system.

Condition L:

Participants were given a written study guide with illustrations of the social conversational protocols, and were told that they had a maximum of 20 minutes to study the material provided. During this time, the participant was left alone in the experiment room.

3.1.3 Testing Session

Participants were tested immediately after completing either the VR or text-based instruction. During testing, each participant

was told that they will be presented with videos of real south Indians on the screen, and were instructed to carry out the appropriate protocol (either greeting or saying goodbye). In every testing scenario the participant always initiated the greeting or the goodbye. They were also told that, after a certain delay, the person presented on the screen would respond. Their greetings or goodbyes were recorded by video camera and the video recordings were used to score how well the participants performed. Three south Indian evaluators who were blind to each participant's condition viewed the digital videos and scored the participant's performance. The evaluators were provided with evaluation criteria which consisted of the instructions used in training for each of the conversational protocol tasks. The evaluators were asked to evaluate if the actions were performed correctly and in the right order by assigning a score between 1 and 7 for each step described in the protocol (1 = not at all, 7 = perfectly).

3.1.4 Post-Experiment Session

Finally, participants in both conditions were orally debriefed.

4 Results and Analysis

4.1 How well did participants in condition VR learn the social conversational protocols as compared to participants in condition L?

The ratings described in section 3.1.3 were summed and translated to provide a score on a scale from 0 to 96. Participants in condition VR (M = 91.97, SD = 2.41) scored higher than participants in condition L (M = 84.90, SD = 4.79). A t-test was used to compare the differences in video evaluation scores between conditions L and VR. Levene's test for equality of variances was significant, F = 5.04, p = 0.031, and so results were generated without assuming homogenous variances. There was a significant difference between the two groups, t(28.02) = 5.90, p < 0.001, indicating that the participants who were instructed in condition VR were able to learn the relevant social protocols better than those in condition L. Additionally, there was less variation in scores for those in condition VR. This experimental design provided an estimated power of 0.46 to detect medium-size effects.

Table 2: Shows the descriptive statistics for participants in conditions L and VR on performance in the testing session.

CONDITIONS	N	MEAN	STD. DEV.	STD. ERROR
L	20	84.90	4.79	1.070
VR	20	91.97	2.41	0.538

t(28.02) = 5.90, p < 0.001

4.2 Discussion

Overall, average scores for both conditions were high. Participants learned the social conversational protocols with either type of training. However, participants in condition VR performed significantly better than participants in condition L. An important difference between the two conditions was that the variance in test performance scores for the VR condition was four times lower than the variance for the L condition. This suggests that training with virtual humans provides a more consistent and reliable result.

5 Summary and Future Work

In our study, we explored if immersive virtual humans can train and teach users in social conversational protocols, specifically in the verbal and non-verbal behaviors associated with conversation initiation and disengagement in south Indian culture. Results of our study suggest that participants who trained with and gained interactive feedback from the immersive virtual humans performed significantly better than participants who learned from the written study guide with illustrations.

Comments and suggestions from our participants indicate that a great number of participants in condition L found the written study guide not as useful as learning from example, practice, and feedback. Based on our participants' comments, we believe that our system can also be useful in training users in social conversational protocols in other cultures.

In future work, we would like to compare learning with immersive virtual humans to learning from video-based instructions. We would also like to evaluate the effect of learning social conversation protocols from immersive virtual humans as part of a narrative scenario as opposed to the current constrained system with precise instructions.

Reference

[1] M. S. Thirumala, "Communication via Gesture", in Language in India, Strength for Today and Bright Hope for Tomorrow, vol 3, 2003.

[2] J. Cassell, "Embodied Conversational Interface Agents," in Communications of ACM, vol. 43, pp. 70-78, 2000.

[3] W. L. Johnson, and J. Rickel, "Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environmnents," in SIGART Bulletin, vol. 8, pp. 16-21, 1998.

[4] K. R. Thorisson, "Gandalf: An Embodied Humanoid Capable of Real-Time Multimodal Dialogue with People," in Proceedings of the First ACM International Conference on Autonomous Agents, Marina Del Rey, California, 1997.

[5] D. Pertaub, M. Slater, and C. Baker, "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience," in Presence: Teleoperators and Virtual Environments, vol. 11, pp. 68-78, 2001.

[6] C. Nass, and Y. Moon, "Machines and Mindlessness: Social Responses to Computers," in Journal of Social Issues, vol. 56, pp. 81-103, 2000.

[7] K. R. Thorisson, and J. Cassell, "Why Put an Agent in a Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue," in Proc. of Lifelike Computer Characters '96, Snowbird, Utah, 1996.

[8] M. Slater, M. Sadagic, M. Usoh, and R. Schroeder, "Small-Group Behavior in a Virtual and Real Environment: A Comparative Study," in Presence: Teleoperators and Virtual Environments, vol. 9, pp. 37-51, 2000.

[9] A. Raij, K. Johnsen, R. Dickerson, B. Lok, M. Cohen, A. Stevens, T. Bernard, C. Oxendine, P. Wagner, D. S. Lind, "Interpersonal Scenarios: Virtual § Real?" in Proc. of IEEE Virtual Reality 2006, Alexandria, VA, March 2006.

[10] J. Deaton, C. McCollum, "Applying a cognitive architecture to control of virtual non-player characters," in Proc. of the 2004 Winter Simulation Conference, pp. 883-889.

[11] S. Babu, S. Schmugge, R. Inugala, S. Rao, T. Barnes, L. F. Hodges, "Marve: a prototype virtual human interface framework for studying human-virtual human interaction," in Proc. of the 5th International Working Conference on Intelligent Virtual Agents (IVA 2005), Kos, Greece, pp. 120-133, 2005.

[12] K. Johnson, R. Dickerson, A. Raij, B. Lok, J. Jackson, M. Shin, J. Hernandez, A. Stevens, and D. S. Lind, "Experiences in Using Immersive Virtual Characters to Educate Medical Communication Skills," in Proc. of IEEE Virtual Reality 2005 (VR 2005), Bonn, Germany, 2005.