Title: Teaching Communication Skills with Virtual Humans

Benjamin Lok, University of Florida

At the University of Florida (UF), our research team has worked on applying virtual humans as partners in interpersonal communication scenarios. The goal of the project: *teach communication skills using virtual humans (VHs)*. The simulation of an interaction between people to facilitate teaching, training, and testing of communication skills would be a powerful new application of VR. With VHs, is the human representation engaging users on such a powerful level that the interface significantly impacts effectiveness?

Our approach is to employ *natural methods of interaction* with the VH. We are striving to create an experience similar to two (initially) people talking. During the past two years, we have worked on a system where the user's speech, gestures, and body language are inputs to life-sized rendered and animated 3D conversation VHs. This is as opposed to using traditional interfaces, such as monitors, mice, and keyboards.

Technology limitations, such as speech recognition accuracy, emotion perception, and eye-gaze tracking, would not allow for a generalized interpersonal experience. Thus, to study virtual humans, we have created and refined a ***constrained*** scenario involving a virtual patient (VP), DIANA (DIgital ANimated Avatar), being interview by a medical student. Currently, students learn, practice, and are tested on communication skills with standardized patients (SPs), actors trained to mimic condition symptoms.

The research team includes experts in computer science, medicine, communication, and education, from the Computer and Information Sciences and Engineering Department, College of Education, and the College of Medicine at UF, and the College of Medicine at the Medical College of Georgia (MCG).

DIANA is a female college student with acute abdominal pain (AP). Students interview DIANA in a standard exam room, and the experience attempts to mirror a patient encounter. DIANA is a life-sized VH and communicates with speech, gestures, and body language. By using a natural interface, we aim to have the student interact with DIANA naturally, and thus allow the training of interviewing and communication skills.

In the scenario, DIANA has appendicitis, while VIC (Virtual Interactive Character), a male virtual character, is an observing expert (Figure 1). DIANA and VIC's gestures and responses were created by medical faculty and archival videos of AP SP interviews.

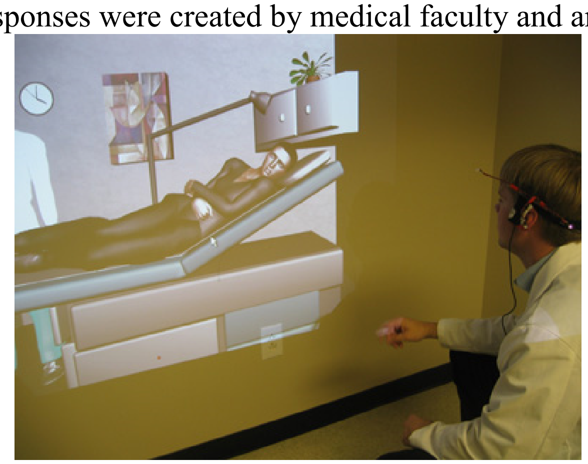

Figure 1 - A medical student interviews DIANA in a standard examination room.

The system is composed of the following (Figure 2): two PCs, two cameras to track LEDs on the user's head, hand, and in the environment, a data projector to display the VHs at life-size, a wireless microphone, Dragon Naturally Speaking 8 Professional, and Haptek Inc. virtual characters. The system costs less than $8,000 and the use of commodity components makes adoption and distribution a realistic long-term goal.

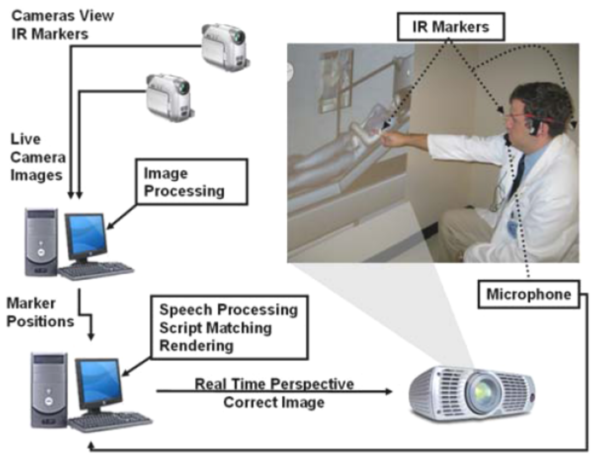

Figure 2 - The system is composed of commodity-off-the-shelf components.

Prior to the VH interaction, students spend ten minutes creating a voice recognition profile. Students enter the exam room and immediately see VIC and DIANA projected life-size on the exam room wall. VIC welcomes the student and provides instructions on communicating with the system. VIC then leaves the room and students converse with DIANA naturally to develop a differential diagnosis (no physical exam).

The microphone and video cameras provide audio and video input. The only additional hardware attached to the user is a ballcap (for tracking) and the wireless microphone and transmitter. We focused on reducing infrastructure to reduce breaks in presence. Speech recognition software and a straightforward algorithm for parsing utterances allow the participant to talk to VIC and DIANA naturally within the scope of the AP scenario. The student's hand is tracked, allowing the participant to localize DIANA's pain with simple pointing gestures. The participant's head is tracked to render the scene from their viewpoint (prospectively-correct warping of the 3D world), and to allow DIANA and VIC to maintain eye contact with the participant.

Life-sized projection with speech and gestures more closely mimics a doctor-patient interaction (interaction videos are at http://www.cise.ufl.edu/research/vp/videos). Care was exercised to keep the level of immersion as high as possible, such as always having DIANA and VIC present in the exam room as the student enters and leaves the room. No icons, desktops, mice, or keyboard are ever visible to the student.

This system was reported in [1, 2], and here we report on lessons learned (what we think we know), unresolved issues (what we want to know), and ongoing developments (how we are going to find out).

Lessons Learned

Eight studies (*n*=101) have been conducted that has investigated integration into two universities, curriculum integration, comparing real and virtual interactions, and exploring VH diversity. This has resulted in the following preliminary conclusions (for complete details and statistics, please consult the publications):

Using VHs as participants in a constrained interpersonal scenario barely works.

Interactions were evaluated using three metrics:

Patient rated: Maria (an SP) or DIANA scored the interaction using current SP evaluation forms. Currently, a passing score is based on this component alone.

Student rated: the medical student scored the DIANA's performance using a common SP evaluation questionnaire.

Expert rated: expert observers (physicians and medical faculty) reviewed videos of the interviews and rated the student's performance. Grading criteria includes verbal and nonverbal cues such as eye contact, posture, and vocal inflection

Overall the interactions and task performance was similar between participants interviewing Maria or DIANA; both groups scored similarly on patient ratings (eliciting the same information from the VP/SP). This supports the virtual scenario as having a strong correlation to its real world counterpart. It also shows participants put forth a similar level of effort into achieving the goals in a VP interaction as in SP interactions.

Students reported that the tool appeared authentic and stimulated them to ask questions. Students reported a moderate sense of presence (SUS raw count - 5.12). Students rated highly the VH gestures, life-size VH representation, VH using speech, head-tracking, speech recognition, and the experience occurring in the SP center.

Participants' overall rating of DIANA (μ=7.2) was in the same ballpark as SP national ratings (7.44). Largely, students were enthusiastic about the VH interaction and its value as a teaching tool. Most students felt the system would aid in preparation for interaction with SPs and real patients. Students were willing to interact with VHs, and believed that they had a place in learning how to practice medicine.

However, a battery of lower level questions better highlights the important similarities and differences. Participant responses showed differences in whether the patient appeared authentic, and whether the encounter was similar to other prior patient encounters. Students reported that the SP communicated how she felt better and appeared to be a better listener than the VP.

The VP differed primarily in her expressiveness and limited ability to handle more than basic conversation flow. The expressiveness of good SPs sets the bar very high for the VP and could have contributed to differences. Challenges in speech recognition still remain, and students at times forced speech in the VP interaction. One significant VP advantage was the clear uniform feedback delivered.

Limitations to current technologies aid novice users with constrained domain-specific conversations, but hamper experts with complex conversation styles.

The system has been evaluated with participants that span a range of experience levels, from beginning medical students to practicing clinicians. The results showed that the system is still rudimentary for experienced communicators as they used a more conversational style. However, novices' errors usually involved phrasing that most laypeople would find confusing. DIANA's inability to understand these phrasings forced users to reword questions using more basic language. Observing faculty mentioned this 'bug' was of significant major teaching benefit. Our focus is on 1st and 2nd-year medical students. These students have had minimal interactions with patients and use a more 'by the book' style.

Teachers and students of interpersonal scenarios want simulators.

The patient-doctor interview represents an interaction where it is difficult to provide adequate practice, teaching, and evaluation on a critical skill. With few effective tools other than role-playing and first-hand experience, medical students have overwhelmingly commented on the expanding role of simulation and VHs.

When asked how often they would use such a system to practice if it were available at all times***, 50% of participants indicated weekly and 40% responded monthly*** (10% - no answer). Considering the over 64,000 medical students nationwide, there is significant potential demand and applicability of such a system.

Integration of a VH scenario into the medical school curriculum is a realistic goal

With the eventual goal of integration into the medical school curriculum, a pilot study was conducted that examined the feasibility of running a VP station alongside SP stations during a UF Essentials of Patient Care (EPC) lab section. Exam rooms in the clinical skills center were used simultaneously in a timed fashion with SPs in all the rooms except for one room equipped with a VP. 34 of the 128 enrolled in the EPC course (2nd-year medical students) were randomized to the VP interaction and conducted the interaction concurrently with other students interacting with SPs.

Empathy can be elicited; however empathetic actions towards the VP were more superficial.

Both the VP and SP were instructed to provide an empathetic moment, "I'm scared, can you help me?" about two minutes into the conversation. The percentage of participants that expressed empathy (a critical communication skill) was equal between the students who experienced the VP and who experienced the SP. However, the number and depth of the empathetic actions were still stronger with the SP (such as touching the SP's leg or exam bed as opposed to simple verbal reassurances.

The VH's speech affects the types of communication skills that can be taught. [3]

The VHs speech can be created either using recorded speech from a voice talent (high fidelity, low flexibility, and resource intensive) or using a speech synthesis engine (lower fidelity, high flexibility and low resource requirements). We studied if the VH could get away with using synthesized speech.

There exist subtle – yet important – differences between VPs and SPs, primarily relating to conversation flow and the significant difference in level of expressiveness. Part of the lowered expressiveness is auditory, and thus synthesized speech DIANA's lower level of emotive expression impacts the overall experience.

For lower level learning of communication skills, (knowledge on Bloom's Taxonomy of Learning), there appears to be little difference between recorded and synthesized speech. If the goal is to teach the student *which questions to ask*, synthesized speech provides a compelling dynamic approach with minimal loss of educational objectives. However, if the goal is to teach the student *how to ask the correct questions*, a high level of expressiveness in the VH is needed. Essential information of the patient's condition could be lost from using synthesized speech. This in turn necessitates the higher cost – even with the lower flexibility – of recorded speech.

Ultimately, the patient-doctor interaction is an appropriate scenario through which to study virtual humans.

Unresolved Issues

This initial work is scratching only the surface of the potential of virtual humans in interpersonal scenarios. Major questions we are still unclear about include:

- How close is experiencing a virtual interaction to a real interaction?
- What are the key components to research and develop as to have virtual interactions overcome technical limitations and make performance strides?
- Does experiencing a virtual interaction impact learning outcomes? (teaching and testing)
- How do we evolve the interface to capture non-verbal cues to augment the system? Can we quantify skills?
- What other scenarios would benefit from an 'interpersonal simulator'?

Ongoing Developments

The system has undergone testing, development, and evolution to its current state of presenting a surprisingly effective simulation of a basic patient-doctor interaction. Yet, it still falls short of an SP interview in several aspects. However, the studies also suggest that training and educational objectives are still obtainable, and ongoing improvements will only improve VH performance.

Measure communication cues (verbal & non-verbal)

We are currently working to track several verbal and non-verbal cues as inputs to the VP simulation. Currently, the student's head-gaze direction and body-lean are tracked. We are developing methods to track body symmetry, body posture, gesture frequency analysis, and audio frequency analysis.

Virtual Human Diversity

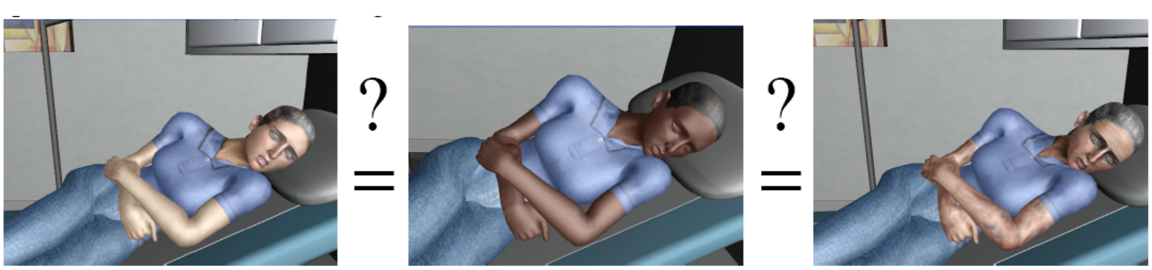

Figure 3 - Are there differences in the interactions with virtual humans of different backgrounds?

Can VHs elicit real world biases (Figure 3)? To begin exploring the social components of the VHs, sixteen volunteer 2nd and 3rd year MCG medical students participated in a diversity study. Participants experienced either a DIANA of a similar ethnic group as the student or DIANA of a different ethnic group. Two ethnic groups were tested (four conditions), Caucasian and African American. The VP differences were only in her skin color, hair color, and recorded voice (different SP voice talent was used, but the actual words spoken were identical).

Behavioral measures were observed (e.g. eye gaze, interaction time, body lean, questions asked) and a post-experience cultural bias survey was administered. Data analysis is ongoing, and preliminary results show promise that VPs do elicit participant biases. An elderly DIANA (Edna) has already been created, and a large repository of VPs of varying gender, age, ethnicity, and weight is being constructed.

Improve expressiveness / conversation flow

We are applying the formal emotion coding systems, FACS (The Facial Action Coding System) to creating DIANA's expressions. This will enable the system to leveraging understood emotive triggers to more powerfully drive the student into understanding DIANA's emotive state.

The conversation engine is primarily a user-initiated dialogue system (where the user begins conversation units with a statement or question and the system interprets and responds). However, this is not reflective of effective patient-doctor interviews. The conversation system is being expanded to handle medical interview conversation-specific idioms, such as summarization, acknowledgement, incomplete phrases, and empathy.

Visualize

We are developing a toolset of effective visualizations that will allow an analysis of a completed VH interaction. The tool would allow the replay of the interaction including visualizing the 3D real-time concurrently with other data, such as vocal inflection, head gaze (where a student was looking), gestures, and body posture. For example, a powerful ability would be to allow a medical student to review the interaction from the other participant's viewpoint (in the patient-doctors cenario, DIANA's).

This visualization tool will allow a medical educator to create a set of objective evaluation measures to codify good communication skills. Through novel visualization methods, we can more fully real realize the potential of virtual humans and provide effective training for interpersonal scenarios.

Component Evaluation

We are taking a component by component investigation into their impact to interpersonal communication simulation. Specific components, such as speech fidelity, speech interfaces, and level of immersion, are examined in formal studies to identify the role each plays in the interaction.

Integration into the cirriculum

DIANA will be incorporated after the first semester in communication courses at UF (n=135) and MCG (n=200). ***All*** 1st-year medical students at UF and MCG will conduct a VP interview and students will then complete remainder of the courses. We will then try to correlate course performance to VH interaction performance.

Our goal is to lead the world-wide incorporation of virtual humans as partners in interpersonal scenario simulations.

The Virtual Patients Research Team:

University of Florida

CISE – Andrew Raij, Kyle Johnsen, Rebecca Wells, Robert Dickerson, Benjamin Lok

College of Medicine – Juan Cendan, Amy Stevens, Marc Cohen, Rebecca Pauly, Margaret Duerson

College of Education – Rick Ferdig

Medical College of Georgia – Scott Lind, Peggy Wagner

Please visit http://www.cise.ufl.edu/research/vegroup for more information.

For Further Reading:

1. Johnsen, Kyle, Robert Dickerson, Andrew Raij, Benjamin Lok, Jonathan Jackson, Min Shin, Jonathan Hernandez, Amy Stevens, Scott Lind (2005). "Experiences in Using Immersive Virtual Characters to Educate Medical Communication Skills," In *Proceedings of IEEE Virtual Reality 2005*.

2. Johnsen, Kyle, Robert Dickerson, Andrew Raij, Benjamin Lok, Jonathan Jackson, Min Shin, Jonathan Hernandez, Amy Stevens, Scott Lind (2005); Evolving an Immersive Medical Communication Skills Trainer; To appear in *Journal on Presence: Teleoperators and Virtual Environments*.

3. Dickerson, Robert, Kyle Johnsen, Andrew Raij, Benjamin Lok, Thomas Bernard, Amy Stevens, D. Scott Lind (2005). "Virtual Patients: Assessment of Synthesized Versus Recorded Speech," In the *Proceedings of Medicine Meets Virtual Reality 14*.