

# Experiences in Using Immersive Virtual Characters to Educate Medical Communication Skills

Kyle Johnsen, Robert Dickerson,  
Andrew Raji, Benjamin Lok<sup>1</sup>

Computer and Information Science  
and Engineering  
University of Florida

Jonathan Jackson, Min Shin<sup>2</sup>

Computer Science Department  
University of North Carolina at Charlotte

Jonathan Hernandez, Amy  
Stevens, D. Scott Lind<sup>3</sup>

VA Hospitals  
University of Florida

## ABSTRACT

This paper presents a system which allows medical students to experience the interaction between a patient and a medical doctor using natural methods of interaction with a high level of immersion. We also present our experiences with a pilot group of medical and physician assistant students at various levels of training. They interacted with projector-based life-sized virtual characters using gestures and speech. We believe that natural interaction and a high level of immersion facilitates the education of communication skills. We present the system details as well as the participants' performance and opinions. The study confirmed that the level of immersion contributed significantly to the experience, and participants reported that the system is a powerful tool for teaching and training. Applying the system to formal communication skills evaluation and further scenario development will be the focus of future research and refinement.

**CR Categories and Subject:** J.3. Life and Medical Sciences, K.3 Computers and Education, H.5 Information Interfaces and Presentation

**Additional Keywords:** Virtual Characters, Multimodal Interaction, Medical Education, Immersive Virtual Environments

## 1 OVERVIEW



Figure 1. A female standardized patient (trained actor) complains of abdominal pain.

*"Doctor, I have a pain in my side! Please make it go away!"*  
Training medical students on how to diagnose and handle this common patient complaint is a challenge for medical educators.

Initially, education is often done with role play and standardized patients (actors trained to represent specific conditions). Later, an apprenticeship model is employed where the student observes the expert, and then gradually takes over. Current methods for learning communication skills have *repetition*, *experience diversity*, *quality control* (keeping the experience consistent across students), and *effectiveness* concerns.

Researchers have been exploring the benefits of simulating social situations. USC's CARTE group has applied intelligent agents to pedagogical and training applications [1], and the ICT group has created experiences to train military personnel in interpersonal leadership [2]. Similar in spirit to this work, Research Triangle Institute's (RTI) Virtual Standardized Patient is a commercial virtual character system for training medical students [3]. Students interact with 3D virtual patients displayed on a monitor while using natural speech (with natural language processing), mouse, and keyboard. Pertaub, *et. al.*, observed participants with a fear of public speaking, speaking to an audience of virtual characters. They responded similarly to when they spoke to an audience of real people. Further, experiencing a virtual social situation, may reduce anxiety in reality [4].

We aim to add to this research by providing our experiences in studying the following:

- How does a higher level of immersion (such as life-size displays, head-tracked rendering, and gesture recognition) impact cognition and education of communication skills?
- How do people react in a highly immersive social experience with a virtual character?

By developing and evaluating interaction methods to increase the level of immersion, we hope to more fully exploit the capabilities of virtual characters

**Hypothesis:** Experiencing a virtual character interpersonal scenario at a high level of immersion and with natural interaction, will facilitate effective training, teaching, and testing of communication skills.

We created a system where a student may interact naturally with a virtual patient (a life-sized virtual character) and installed it in a facility currently used for standardized patient training. An observing expert (also a life-sized virtual character) provides scenario information and gives immediate feedback on the student's performance. The student interacts with the characters through speech recognition, gestures, and head-tracked rendering.

This paper details the experience of creating this immersive virtual patient system and discusses the feedback from a pilot group of students who experienced it. They reported that the system would be a powerful tool for teaching and training, though formal communication skills evaluation is still a long term goal.

<sup>1</sup> [kjohnsen, rfd, raji, lok] @cise.ufl.edu

<sup>2</sup> [jojackso, mcshin] @uncc.edu

<sup>3</sup> jhern@med.ufl.edu, Amy.Stevens@med.va.gov, lindds@surgery.ufl.edu

## 2 PREVIOUS WORK

### 2.1 Medical Training

Medical interview simulators commonly use either text or prerecorded-video based scenarios. One example system uses a web-based teaching approach that enabled distributed students to cooperatively solve a problem [5]. Most approaches can be classified as either narrative (students are presented possible actions given the current state) or problem-solving (students have a list with all the possible actions). In a comparative study, participants remarked the narrative approach was more effective because they felt they got to know the patient better (despite her saying the same statements) [6]. Both approaches were labeled artificial and restrictive.

Virtual characters have been applied as interactive agents for training and procedure planning. The Human Patient Simulator is a full sized mannequin with motors and actuators to produce physical input and output for emergency room training [7]. The Just VR system uses an immersive approach to have students experience and react to health emergency scenarios [8].

### 2.2 Virtual Characters

Creating useful or believable agents has been a goal for artificial intelligence and simulation research for such applications as simulating military forces [9] and crowds [10]. MIT's Synthetic Characters group is investigating using artificial intelligence to drive interactive autonomous agents [11]. Virtual characters have also been used as interactive guides [12]. There are software libraries, such as Microsoft Agents that add 'interactive personalities' to application interfaces.

Highly detailed human models, complete with physical constraints, can be integrated with virtual parts, vehicles, and tools to evaluate ergonomics and usability. The Human Modeling and Simulation Group at the University of Pennsylvania use virtual humans for task analysis and assembly validation [13]. Safework, EDS's Jack system, and Boston Dynamics are commercial packages that incorporate virtual humans for ergonomics and human factors analysis.

Finally, VQ Interactive's BOTizen and Haptik Inc.'s HapPlayer software provide virtual characters to help commercial website visitors to enjoy a more interactive and memorable experience.

### 2.3 Interaction

Researchers in human-computer interaction, virtual environments, natural language processing, and computer vision have developed and studied the effects of natural interaction.

There is substantial gesture recognition literature in computer vision. Aggarwal, *et. al.*, classified approaches to human motion analysis, tasks involved, and major areas related to human motion interpretation [14]. A review by Pavlovic, *et. al.*, addresses the main components and directions in gesture recognition research [15]. Hidden Markov models have also been employed to address the highly stochastic nature of human gestures [16].

Most virtual environment systems use tracked devices (*e.g.* data gloves, joysticks, mice) and translating hardware actions (*e.g.* button pushes or gestures) to perform actions such as grasping. Pfeiffer described a framework to incorporate speech and gestures for manipulating objects in an immersive (large projection) virtual environment [17]. Another approach is to engineer a device for a specific type of interaction to allow the user to naturally perceive the object's role [18]. However, specialized engineering can be time-consuming and the results are often application specific.

### 2.4 Immersion

Researchers have investigated how people respond to virtual characters as an important component to interaction. In studies, the importance of rendering fidelity was examined for interpersonal communication [19]. Results suggest that emotions, such as embarrassment, fear, irritation, and self-awareness, could be elicited in real people by virtual characters [4].

Nass has pioneered research into the affective power of computers and intelligent agents. His research group has shown that people can ascribe very human characteristics to computers, such as the helpfulness, usability, and friendliness [20]. Lisetti has investigated the role of emotion and personality of intelligent agents has on human computer interaction [21].

Previous work by virtual reality researchers has investigated the link between interaction, immersion and cognition. Insko's studies showed improved memory recall through natural locomotion and haptic feedback [22]. Others studies suggest that natural locomotion improves higher order cognitive performance [23], and naturally interacting with real objects improves problem solving task performance [24].

## 3 PROJECT DESCRIPTION

### 3.1 Project Goals



Figure 2. A female virtual patient complains of abdominal pain. The instructor on the right coordinates the diagnosis.

Figure 2 shows interactive virtual patients and instructors which are used to aid in the instruction of communication skills. These characters are a combination of high-quality rendering, animation, and display, with speech and gesture recognition. The virtual characters provide scripted responses to the student's speech and gestures. With its high level of immersion and interactivity, this system allows participants to go through compelling experiences. Repeated exposure would provide valuable communication skills education. It can occur *more often (experience frequency)*, on *more scenarios (scenario variety)*, and with *increased standardization (quality control)* over traditional real-person training. The repetitive practice and potential availability of this technology has real merit in training and documenting physician competency before real patient interactions. Ultimately, the development of more varied and less common patient encounters will make this a powerful educational tool. In the future this tool could be part of the standardized testing process.

### 3.2 Driving Application: Patient-Doctor Interaction

The interaction between patients and doctors during initial condition diagnosis is an interpersonal scenario 1) *which can be simulated despite the limits of current technology*, 2) *where practice is costly*, 3) *where immersion and fidelity is important*, and 4) *which benefits with immediate feedback*.

Currently, high-level components of patient physician interaction, such as bed-side manner and grief counseling, would be severely hampered by even slight compromises in fidelity (inevitable with current technology). However, the system described in this paper has the potential to educate students on the more basic skills of 1) gathering the information by asking core questions and 2) developing a list of possible diagnoses.

Patient-doctor interaction requires substantial first-hand experience to become proficient, but there are insufficient opportunities to go through the wide variety of scenarios. The result is that many medical students do not have sufficient practice in communication skills when they reach the real patient's bedside.

Communication skills are an important part of medical school curriculum and research shows they can be taught and do not just improve over time with clinical experience. Practicing patient-doctor interaction improves communication skills [25]. To provide maximum benefit, the experience should simulate, as closely as possible, the actual scenario.

An informal survey of a sample of third year Shands Hospitals medical students highlighted that although students feel they get adequate instruction in communication skills, they receive little feedback on their real time performances. The design of the virtual scenario includes a virtual instructor who can address these issues of feedback.

### 3.3 Interactive Virtual Characters

Virtual characters can provide benefits for simulating patient-doctor interaction not easily realized with real people.

- *Repetition*. The number of opportunities students receive for experiencing scenarios is increased.
- *Experience diversity*. Scenarios can be experienced with patients of varying age, language, ethnicity, and gender (which affects questions and diagnosis)
- *Quality control*. Virtual characters respond the same way to each student, enabling evaluation with objective measures. This is difficult to ensure with standardized patients and subjective instructor evaluation.

It is estimated that 65% of information in interpersonal interaction comes from nonverbal communication, such as gestures, pose, and gaze [14]. Virtual characters may provide this information to the student while natural interaction techniques enable the student to use nonverbal communication with the character. This pilot explores whether current technology can be capable of providing an adequate fidelity level to achieve basic communication education objectives.

### 3.4 Acute Abdominal Pain (AAP) Diagnosis

Acute abdominal pain (AAP) is defined as follows:

“The term *the acute abdomen* refers to the presence of an acute attack of abdominal pain that may occur suddenly or gradually over a period of several hours. The patient with this symptom complex may confront the surgeon, internist, pediatrician, and obstetrician, creating a problem in clinical diagnosis requiring an immediate or urgent decision regarding the etiology (cause) and method of treatment [26].”

AAP is one of the most common ailments encountered by doctors. It is also a basic scenario in patient-doctor interaction and communication skills education. The doctor begins diagnosis by asking the patient a series of questions about the pain (*history of present illness or HPI*). At this stage, the doctor is trying to ascertain more information, such as the pain's location and character, symptoms exhibited, family history, current medication, and aggravation if certain motions are performed. Sample questions include “What brought you into the clinic today?”, “How long have you had the pain?”, and “On a scale from 1 to 10, please rate the pain.”

The patient's responses will guide the doctor down different routes of questioning. The doctor evaluates the patient's response, gestures, and physical and auditory cues, such as wincing of pain, weight, posture, difficulty in making instructed motions, or pointing to specific areas. Based on asking the appropriate questions and evaluating the answers, treatment options can vary from immediate surgery to observation.

In AAP diagnosis, doctors should *ask the proper questions* and *determine the correct diagnosis*. The traversal of questions is a very well-defined process. The students are graded on *how many of the eleven key questions they asked*. These questions must be asked to ascertain the correct diagnosis. The common mistakes are omitting questions (*e.g. sexual history*), incorrectly asking questions (*e.g. pointing to the wrong quadrant of the patient's abdomen*), or not evaluating and noting the non-verbal cues from the patient's gestures and verbal responses (*e.g. recurring cough*).

## 4 SYSTEM DESCRIPTION

### 4.1 Overview

**Scenario:** The scenario involved a 19 year-old Caucasian female with a complaint of acute abdominal pain. The student needed to ask appropriate questions to come to a diagnosis of appendicitis.

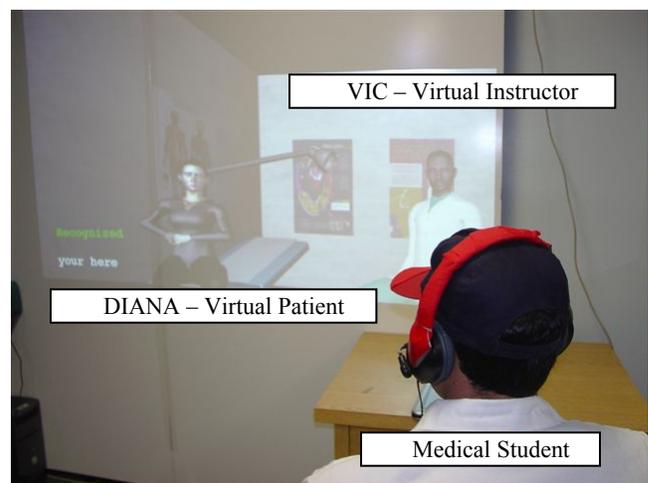


Figure 3. The (right) student is diagnosing (left) DIANA, a patient with acute abdominal pain, while (middle) VIC observes. The colored headset is for head tracking.

DIANA (DIGital ANimated Avatar), a female virtual character, plays the role of the patient with appendicitis, while VIC (Virtual Interactive Character), a male virtual character, plays the role of an observing expert (Figure 3.). DIANA and VIC's scripts, which included gestures and audio responses, were created in consultation with several teaching medical faculty, with substantial standardized patient experience.

The system is composed of the following (Figure 4):

- Two networked PCs
- A data projector to display the virtual characters at life-size
- Two web cameras to track and recognize the user's head and hand movement
- TabletPC
- Microphone

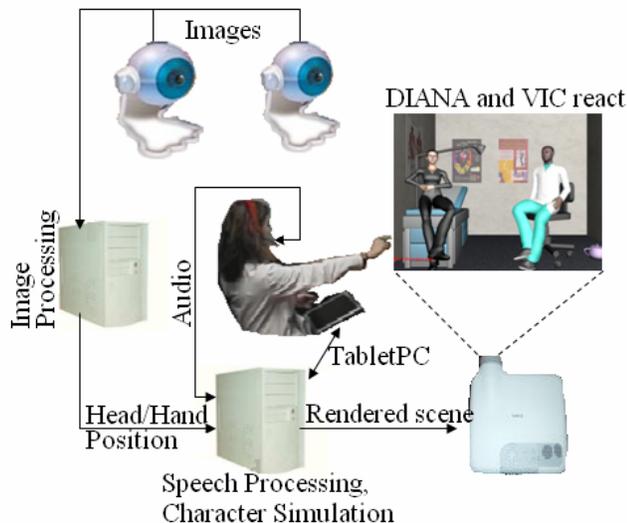


Figure 4. System layout.

We installed the virtual patient system at the Harrell Adult Development and Testing Center at Shands Hospitals at the University of Florida. This testing site is where students interact with standardized patient. Each exam room is set up with closed circuit monitoring for direct observation, as well as to record each session. The prototype system cost less than \$7,000.

In our system the medical student asks a single question or makes a gesture, which the virtual character *perceives*, *thinks*, and then visually and audibly *responds*.

#### 4.2 Virtual Character Perception

The student used speech and gestures to interact with DIANA and VIC. The system received audio and video input from the microphone and web cameras.

The audio was processed into phrases by a commercial speech recognition engine (Dragon Naturally Speaking Professional). To improve accuracy, each participant created a voice profile. During the experience, the system displayed the recognized phrase on screen, allowing the student to identify if the system misrecognized a phrase.

The HPI (history of present illness) portion of the exam consists of a set of questions which the students are taught to ask. The script contained the most likely forms of each question, and several questions could map to the same response. Ex. "Are you nauseous?" and "Have you been vomiting?" both cause DIANA to tell the student that she has felt sick to her stomach. The system used a heuristic to establish a cost to morph the recognized phrase into each question, and then chose the lowest cost (below a threshold) as the understood question.

The system tracked the 3D trajectory of the students' hand with a marker-based tracking algorithm [27]. Two gestures were recognized, *handshaking* and *pointing*. Handshaking was signaled if the student held their hand in front of their body for more than two seconds. Pointing was detected by finding the

intersection of a ray (from the tracked head to the hand) and objects in the scene. A "laser pointer" red dot appeared where the system determined the student was pointing.

While these were simple speech and gesture recognition techniques, they appeared adequate for the scenario.

Tracking the student's head position enabled DIANA and VIC's eyes to focus on the student. Correct perspective warping [28] of the rendered image emphasized the characters' gaze directions, and maintained the illusion of the virtual examination room as an extension of the real room.

#### 4.3 Virtual Character Cognition

The system used a simple state-based machine that transitioned between actions depending on input from the perception stage. Transition rules were based on accepted medical doctrine for the scenario. Actions included the virtual character speaking statements, changes in emotion, or animation. Our medical collaborators verified that acute abdominal pain diagnosis training lent itself well to this architecture.

#### 4.4 Virtual Character Response

DIANA and VIC were displayed at life-size using data projectors. This research proposes that seeing a human face and form at the appropriate size (as opposed to on a monitor) increases immersion and triggers psychological responses. The system used Haptex Inc.'s character animation library, which can generate high-quality, dynamic facial expressions and gestures. The characters verbal responses were driven by the doctor-designed scenario and audio generated by AT&T Wizard's Natural Text-To-Speech software.

Secondary devices provided the student with information and more realistically simulated the encounter. The student used a TabletPC as a notepad and to receive scenario information (e.g. patient file information).

#### 4.5 Immersion



Figure 5. Student asks, "Does it hurt here?"

Medical students learn communication skills by interacting with (real) standardized patients in the Harrell Adult Development and Assessment Testing Center. Standardized patients await students in the center's examination rooms. DIANA and VIC were in Examination Room #3.

This provided the students with the highest level of immersion, closely matching the current standardized patient experience. We were interested in how much of an education 'value-add' would be realized if the system provided natural interaction and a high

level of immersion. For example, one question in AAP diagnosis is “Does it hurt here?” where the doctor points either to a place on herself or on the patient (Figure 5). Does it help to more closely simulate the real experience, *i.e.* instead of moving a cursor and clicking, the student actually points at DIANA?

## 5 CASE STUDY

### 5.1 Design

In an AAP standardized patient scenario, the medical student is given basic information about the patient before entering the room. They then enter the room, greet the patient, and obtain the HPI. To determine this, the student needs to ask a prescribed set of questions to find out why the patient has come into the clinic. Asking a question can consist of *both* a verbal and gesture component. A physical exam usually follows. This study focused on *only* the greeting and HPI derivation.

We replaced the standardized patient with virtual characters. Medical students conducted AAP diagnosis with DIANA and VIC to evaluate if we had achieved our goal of developing a training, teaching, and evaluation tool. Further, we looked to evaluate the impact of the various immersive system components, such as gestures, speech, and immersion, as well as the much asked-for feedback and instruction provided by VIC.

### 5.2 Procedure

#### 5.2.1 Pre-Experience

After arriving at the Harrell Center, each participant signed a consent form. The participant then created a voice profile which took ten minutes. We do not consider this to be a major inconvenience because each student need do this only once, and the saved profile could be used for a variety of scenarios.

The participant is then led to another room to fill out a background survey on prior exposure to standardized patients, abdominal pain scenarios, and medical examinations. Meanwhile, the virtual character system was started. The participant was then brought to the examination room where DIANA’s patient information chart was displayed on a TabletPC in the door basket. They picked up the TabletPC, reviewed the information, and then entered the room to meet DIANA and VIC.

#### 5.2.2 Experience

The participant was seated in a chair facing DIANA. They put on a headset with a colored marker. Another marker is wrapped around their left finger for tracking the hand. The participant is then instructed how to take handwritten notes on the tablet. Before the experimenter left the room, the participant was shown how to use gestures.

VIC began the experience by explaining some information regarding the scenario. He also reminded the participant how to perform the gestures, how to interrupt DIANA if she did not understand correctly, and how to end the session early. VIC then asked them to begin their examination of DIANA. After eight minutes, VIC interrupted the session and announced that two minutes remained. After ten minutes, VIC ended the session and asked the participant for their differential diagnosis.

Finally, VIC reported which, if any, of the eleven key questions were not asked by the participant. VIC then explained the correct differential diagnoses.

This completed the experience, and the student left the examination room. From start to finish, DIANA and VIC were always present.

#### 5.2.3 Post-Experience

After the examination, the participant filled out a presence [29] and co-presence [30] questionnaire. Two other measures were also used. One was a survey used to measure the performance of real standardized patients [31] which was adapted to evaluate DIANA. The other gauged the importance of the system components and evaluated the system as a whole. Finally, in a oral debriefing, the participant provided qualitative feedback.

## 5.3 Results

### 5.3.1 Population

A total of seven participants were involved in the study. There were three male and three female medical students (three 3<sup>rd</sup> years and three 4<sup>th</sup> years) and one 3<sup>rd</sup> year female physician assistant student. All had substantial prior experience with standardized patients (at least 5, average: 10-20). Five had experiences with standardized patients with AAP, and six had experiences with real patients with AAP. This indicated that the participants had significant prior experience in similar scenarios.

### 5.3.2 Task Performance

*Scoring:* Each participant was evaluated on the following criteria: 1) Correct greeting etiquette (*Introduce self, shake hands, query for chief complaint*), 2) Eleven core AAP diagnosis questions that need to be asked to obtain the correct diagnosis, 3) Differential diagnosis (what is the final evaluation).

Five out of seven students introduced themselves and shook hands. All queried for the chief complaint. Out of the eleven core questions, the average number asked was 6.28 (7.0 is a passing grade). Four out of seven received passing marks. Our medical collaborators verified that this was a typical result.

All students forgot to ask DIANA to “tell them more about the pain,” a common mistake, but important because patients often provide only simple answers until asked to elaborate.

After the HPI, there were six possible correct differential diagnoses. The most critical were *appendicitis* and *ectopic pregnancy* because these would require immediate surgery. One student included ectopic pregnancy, while four included appendicitis. Four students had one correct diagnosis and three had two. Only one included an incorrect diagnosis. We did not specifically ask them for all possible diagnoses so we consider this acceptable.

### 5.3.3 Sense of Presence and Co-Presence

The mean SUS Sense of Presence Questionnaire score was 1.57 (number of answers of 6 or 7 for all 6 questions). The average raw SUS score was 4.1. Students had a high sense of actually being in the room (mean: 5.86) with all students saying at least 5.

Co-presence was measured using the 14 question Slater Co-Presence Questionnaire. Responses were on a scale of 1 to 7 (1. *not at all* to 7. *a great deal*). The mean (number of answers of 6 or 7 for all 14) was 3.29. The raw mean was 4.21. Of note is that students generally felt that they contributed positively to the task (5.29), and that DIANA and VIC also contributed positively to the task (5.71). Also, they indicated that they would like to meet DIANA and VIC (5.0).

### 5.3.4 Technology Survey

This self-created survey was used to gauge the impact of technical components on the experience. Responses are on a 7-point Likert scale (1. strongly disagree, 4. neutral, 7. strongly agree).

We had many positive responses towards the immersive technology. Participants indicated that interacting with DIANA

and VIC with speech (mean: 6.71), seeing DIANA and VIC at life-size (6.33), and having the system at the Harrell Center (5.86) were very important to the experience. Also training (6.29) and testing (6.29) were seen as potential uses for the system. To use the system for evaluation (4.86), however, had mixed results. “If this system were installed in a room at Shands (available 24/7) I would use it (never, daily, weekly, monthly, once a semester, never)”, three responded they would use it weekly and four monthly.

We had some unexpected negative results. Interacting with DIANA using handshaking and pointing gestures was not viewed as important to the experience (3.0). The scene ‘moving when they moved their head’ (perspective-correct rendering) was generally not thought critical to the experience (3.29) although opinions were mixed. Interestingly, students responded to the possibility of typing their questions, if it improved accuracy (3.86) all being either high or low, depending on their frustration with the accuracy of DIANA’s responses.

### 5.3.5 Standardized Patient Satisfaction Survey

The questionnaire is adapted from one used to evaluate the performance of real standardized patients [31]. Answers are on a 5-point Likert scale (1. strongly disagree, 3. neutral, 5. strongly agree).

Students indicated DIANA appeared authentic (mean: 4.14), communicated how she felt during the session (4.29) and stimulated them to ask questions (4.14). A mixed feeling about DIANA is that students did not seem to know whether she was listening to them or not (3.0). The only generally negative response was that students felt that DIANA did not answer questions in a natural manner (2.43).

Scores reported that VIC interrupting them was helpful (4.29), his criticism was constructive (4.29); however, they liked only getting feedback at the end of the session (4.14).

Overall, on a scale from one to ten, DIANA was given an overall score of 6.36 for the interaction (authenticity, accuracy, and symptom display).

## 5.4 Discussion

### 5.4.1 Presence and Co-Presence

While Slater suggests the SUS presence questionnaire should be used to compare two conditions, the SUS scores suggest that students report a moderate sense of presence. Students at least thought that they were partly “in” the virtual examination room indicating a high level of immersion.

The raw co-presence mean suggests that the user, at some moderate level, treated DIANA and VIC as other humans as opposed to software components.

### 5.4.2 Technology Survey

With an average score of 6.71 for using speech, continuing to use speech recognition as the primary input method seems appropriate. We believe that providing visual feedback of what the system heard made the student more confident in the system, and enabled them to adapt their speaking to enhance recognition accuracy. While this might not be exactly the same as communication with a standardized patient, the goal of educating a student on the *key questions* was not affected. The desire of some students to type their questions was primarily due to the script being insufficient to handle the many possible questions, and not to the *accuracy* of the recognized speech. While the speech recognition was at or above 90% for all participants, we estimate that only 60% of the asked questions were included in

the script. Refining the script to include more possible questions should help reduce frustration.

Students did not think gestures were important to the experience. This suggests that the scenario did not properly utilize gestures, or the quality of gesture recognition needs improvement.

Perspective-correct rendering was viewed negatively by the students. We originally thought this added heavily to the immersive aspect of the system making the virtual space seem to be simply an extension of the room. Unfortunately the scenario did not require much movement, and thus did not illustrate the value of such a technique. Also, jitter in the optical tracking component occasionally caused noticeable jitter in the display which students commented on in the debriefing.

Participants agreed life-size characters were very important. We believe this suggests that using data projectors to show full-body life-sized virtual characters (as opposed to the limited display space of monitors) was critical to the experience.

All students felt that the system would be an *invaluable* tool in training and testing, especially to those without much patient experience (students in their first two years). Feelings were more neutral on the system as a skills evaluation tool. This is understandable, as system errors and fidelity compromises are much more critical if used to evaluate performance.

All of the participants said they would use the system with a high frequency. Even in its current developmental state, the system and technology was viewed as acceptable for training and teaching.

### 5.4.3 Virtual Patient Satisfaction

High scores for DIANA’s authenticity and communication skills validate our virtual character approach, as well as the content of the scenario.

Work needs to be done to improve how DIANA responds. The lack of higher level information in text to speech, such as tone may have made her responses seem simulated. VIC was highly praised, which correlated with the students’ desire for feedback on their performance. This suggests finding more ways to incorporate VIC, such as context sensitive help and feedback on interruptions or non-verbal communication.

### 5.4.4 TabletPC

Many students write information down when performing an examination. We provided a TabletPC for note taking. This was used extensively by most participants (Figure 6).

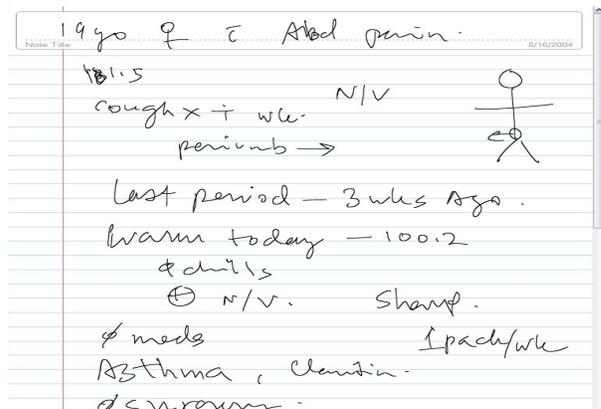


Figure 6. A student’s notes about the session with DIANA.

This added positively to the experience. We see many potential uses of the TabletPC such as providing evaluation results, viewing

patient information (e.g. X-rays and MRIs) or interacting with the system (e.g. restart scenario).

#### 5.4.5 Debriefing

The debriefings yielded many interesting comments, constructive criticism and positive feedback.

**VIC's presence enhanced the experience** by providing helpful information during the session and feedback immediately after.

"I liked that [VIC] gave me feedback at the end and told me exactly which questions I forgot to ask."

One participant said VIC should elaborate on his responses.

"[VIC] should say 'You forgot to ask about fever, which is important because fever is a sign of infection.'"

Multiple participants suggested **VIC could provide hints** when asked, particularly for first and second year medical students who are not experienced enough to get through an AAP scenario.

**Speaking with DIANA and VIC enhanced the experience.**

"I don't think of it as much as watching a computer screen as actually interacting with a person. You're actually talking to a patient, you're not typing in something and waiting for a response."

**Detracting from the experience was that DIANA occasionally answered questions incorrectly or repeated previous answers.** This made it harder to get diagnosis-critical information.

"I wanted to be able to get an answer to those questions."

"I felt like she didn't always answer the question I asked her."

For example, one person asked DIANA several times whether she had a history of gall bladder problems, but the script for DIANA had no information about her gall bladder and thus the system responded incorrectly. This visibly frustrated the student.

**Participants reported learning how to ask DIANA questions** to avoid improper responses. Often, medical professionals need to ask the same question of a patient multiple ways to learn the information they need to make a diagnosis.

"I think it's good for us, too, because if the patient doesn't understand the question, which is inevitably going to happen in real life, too, it forces you to think about other ways to ask questions."

Others noted that it was quite distracting to them and thus implies that students would benefit from an improved system.

"[I got] caught up with trying to think of a way to phrase [the question] rather than taking her history."

From the debriefing we learned that DIANA had a tendency to offer too much information, which is not typical of most patients.

"Often times I actually thought she gave more detailed answers than real people."

Another participant joked that he did not need to interview her after her initial complaint. Others made the point that some patients give up information more readily than others. One participant suggested simulating this variability by providing varying difficulty levels. DIANA, in her current state, might be considered an easy patient because she offers much information. Harder difficulty level patients would provide less information. The script is easily modifiable to enable this.

**Most felt the gestures were not very useful**, and many did not even remember to use them.

"I think the whole shaking hand thing and pointing is not really that important."

Some said DIANA pointing to the right place on her abdomen indicated where her pain was, so they did not see a need for pointing at her. Some saw handshaking as a novelty while others saw no value in it because DIANA is not real.

"[People] would not accept an image as someone they can shake their hands with."

Despite this, a few participants felt the handshake gesture enhanced the training aspect of the system.

"I think [handshaking] is important because that's one of the things that they try to make sure we do automatically every time we walk in the room."

The ability to give a physical exam via hand gestures may be an important feature that could be added. When we asked participants to indicate what additional gestures they would like to see in the system, the almost unanimous response was the ability to give a physical exam. In fact, one participant asked DIANA twice for permission to do the physical exam asking her to "lie down" to continue the examination.

Students emphasized that **having the system available 24 hours a day, 7 days a week, would be very valuable.**

"A big benefit would be to be able to go in and do this at your leisure and practice with it. I would definitely use it."

The low cost and commodity hardware makes it viable to install similar systems in training facilities, medical schools, and hospitals. Work would need to be done to make the system more robust for continuous use.

Finally, almost all the study participants felt that **the current system was most appropriate for 1<sup>st</sup> and 2<sup>nd</sup> year medical students.** They explained that:

1. Students currently do not get any exposure to standardized patients before early in their 2<sup>nd</sup> year, but the virtual patient system could open the door for students to practice interviewing patients earlier and more often.
2. 1<sup>st</sup> and 2<sup>nd</sup> year students are typically very nervous when asking questions of patients. Such a system would provide good practice at the early stages when "you're so nervous and you just need the idea of how it should go."
3. The AAP scenario used is a textbook case that is not challenging for 3<sup>rd</sup> and 4<sup>th</sup> year students but would be for 1<sup>st</sup> and 2<sup>nd</sup> year students.
4. Ultimately, 3<sup>rd</sup> and 4<sup>th</sup> year students need to practice with real people. "It's not as good for a 2<sup>nd</sup> and say a 3<sup>rd</sup> year, because you are now really supposed to be working with people that have different personalities."

## 6 CONCLUSION AND FUTURE WORK

We believe that in creating an immersive virtual character system with natural interaction, we have achieved our goal of more fully exploiting the capabilities of virtual characters. The performance of our virtual patient, DIANA, was given a high mark (mean 6.36 out of 10 by standardized patient guidelines), and results indicate that having a virtual instructor give *real-time* performance feedback added heavily to the experience. A moderate level of presence and co-presence was reported during the study. Participants indicated that life-sized virtual characters, speech recognition, and having the system at the Harrell Center were *crucial* while gesture recognition and immersive head-tracked displays need work for the majority to accept them.

We are extending the system to incorporate additional devices to increase immersion. For example, triggering the student's beeper to "go off" during an interview enables practice on maintaining patient rapport while handling unexpected urgent situations. Also, we are detecting other actions, such as if the examination room sink is used to remind the student to wash their hands. We also aim to provide high-level evaluation techniques, such as the student's use of eye contact (using tracked head information) and the number of times they interrupt DIANA. Finally, we plan on implementing the participant suggestions of difficulty levels for DIANA and the ability to give physical exams via hand-gestures and commands.

Given the overall positive feedback on the system, a large controlled study is planned. Students would interview either DIANA or a standardized patient with the same symptoms (both work from the same script). It is our goal to show no significant difference between the performances. Finally, a future study is planned to evaluating if experiencing virtual patients would lower the anxiety of inexperienced medical students before their first standardized patient encounter.

We hope the lessons learned and experiences gained will provide insight to developers of similar projects. Understanding how to interact with virtual characters is the *critical first step* to better realize their potential for educating interpersonal skills.

## 7 ACKNOWLEDGEMENTS

## REFERENCES

- [1] Rickel, J. and Johnson, W. (1999) Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence* 13:343-382.
- [2] Hill, R, Gratch, J., Marsella, S., Rickel, J., Swartout, W., and Traum, D. (2003) Virtual Humans in the Mission Rehearsal Exercise System. *KI special issue on Embodied Conversational Agents*.
- [3] Hubal, R., Kizakevich, P., Merino, D. and West, S. (2000) The Virtual Standardized Patient: Simulated Patient-Practitioner Dialogue for Patient Interview Training. In *Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue*, J.D. Westwood, H.M. Hoffman, G.T. Mogel, R.A. Robb, & D. Stredney (Eds.), IOS Press: Amsterdam, 2000.
- [4] Pertaub, D., Slater, M., and Barker, C. (2001) An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience, *Presence: Teleoperators and Virtual Environments*, 11(1), 68-78.
- [5] Beer, M., Green, S., Nefti, S., Sixsmith, A. and Whatley, J. (2001). "Education for Health Care Professionals Through the Medium of the Internet", in *Proceedings of IEEE International Conference on Advanced Learning Technologies*, 31-32.
- [6] Bearman, M. (2003). Is Virtual the Same as Real? *Medical Students' Experiences of a Virtual Patient?* *Academic Medicine*, 78(5).
- [7] Van Meurs, W., Good, M., and Lampotang, S. (1997). Functional Anatomy of Full-Scale Patient Simulators, *Journal of Clinical Monitoring* 13:317-324, 1997.
- [8] Ponder, M., Herbelin, B., Molet, T., Scherteneib, S., Ulicny, B., Papagiannakis, G., Magnenat-Thalmann, N. and Thalmann, D. (2002) Interactive Scenario Immersion: Health Emergency Decision Training in JUST Project, *VRMHR2002 Conference Proceedings*.
- [9] Stytz, M. and Banks, S. (2003) An Architecture to Address Uncertain Requirements and Composability for Intelligent Agents in Distributed Simulations. *Design and Application of Hybrid Intelligent Systems*, 749-758.
- [10] Hamagami, T. and Hirata, H. (2003) Method of Crowd Simulation by Using Multiagent on Cellular Automata. *Intelligent Agent Technology, 2003 IEEE/WIC International Conference*. 13-16 Oct. 2003, 46- 52.
- [11] Burke, R., Isla, D., Downie, M., Ivanov, Y., and Blumber, B. (2001) "CreatureSmarts: The Art and Architecture of a Virtual Brain" in the *Proceedings of the Game Developers Conference*, San Jose, CA, 147-166.
- [12] Thórisson, K. (1997). Gandalf: An Embodied Humanoid Capable of Real-Time Multimodal Dialogue with People. In the *Proceedings of the ACM 1st International Conference on Autonomous Agents*, 536-537.
- [13] Badler, N., Erignac, C. and Liu, Y. (2002). Virtual Humans for Validating Maintenance Procedures, In *Communications of the ACM*, 45(7), 56-63.
- [14] Aggarwal, J. and Cai, Q (1997). Human Motion Analysis: A Review. In *IEEE Non-Rigid and Articulated Motion Workshop*, San Juan, Puerto Rico, 90-102.
- [15] Pavlovic, V., Sharma, R., and Huang, T. (1997) Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677-695.
- [16] Vogler, C., Sun, H., and Metaxas, D. (2001) A Framework for Motion Recognition with Applications to American Sign Language and Gait Recognition. In *Proceedings of Workshop on Human Motion*, Los Alamitos, CA, December 2001, 33 - 38.
- [17] Pfeiffer, T. and Latoschik, M. (2004) Resolving Object References in Multimodal Dialogues for Immersive Virtual Environments, In *Proceedings of IEEE Virtual Reality 2004*, Chicago, IL.
- [18] Pierce, J. and Pausch, R. (2002) Comparing Voodoo Dolls and HOMER: Exploring the Importance of Feedback in Virtual Environments, In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, 105-112.
- [19] Garau, M., Slater, M., Bee, S., and Sasse, M. (2001). The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems*, March 31- April 5, 2001, Seattle, WA USA, 309-316.
- [20] Nass, C. and Moon Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.
- [21] Lisetti, C. (2002) Personality, Affect and Emotion Taxonomy for Socially Intelligent Agents. In *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference (FLAIRS'02)*, Menlo Park, CA: AAAI Press, (Pensacola, FL).
- [22] Insko, B. (2001). *Passive Haptics Significantly Enhances Virtual Environments*, Department of Computer Science, UNC-Chapel Hill, Unpublished dissertation.
- [23] Zambaka, C., Lok, B., Babu, S., Xiao, D., Ulinksi, A., and Hodges, L. (2004). Effects of Travel Technique on Cognition in Virtual Environments, In *Proceedings of IEEE Virtual Reality 2004*, Chicago, IL, 149-156, 286.
- [24] Lok, B., Naik, S., Whitton, M., and Brooks, F. (2003). Effects of Interaction Modality and Avatar Fidelity on Task Performance and Sense of Presence in Virtual Environments, *Journal on Presence: Teleoperators and Virtual Environments*, 12(6), 615-628.
- [25] Lang, F., McCord, R., Harvill, L., and Anderson, D. (2004) Communication Assessment Using the Common Ground Instrument: Psychometric Properties. *Family Medicine*, March 2004, 36(3).
- [26] Diethelm, A. (1997) "The Acute Abdomen". *Textbook of Surgery: The Biological Basis of Modern Surgical Practice*, 15th Edition. Ed. D. Sabiston. Saunders W B Co, 1997, Ch. 29.
- [27] Jackson, J., Lok, B., Kim, J. Xiao, D., Hodges, L., and Shin, M. Straps: A Simple Method for Augmenting Primary Tracking Systems in Immersive Virtual Environments. *Future Computing Lab Technical Report*. University of North Carolina at Charlotte.
- [28] Raskar, R. (2000). Immersive Planar Displays Using Roughly Aligned Projectors. In *Proceedings of IEEE Virtual Reality 2000*.
- [29] Usoh, M., Catena, E., Arman, S., and Slater, M. "Using Presence Questionnaires in Reality" *Presence: Teleoperators and Virtual Environments*, Vol 9, No 5, pp 497-503
- [30] Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., and Whitton, M. (2002) Collaboration in Tele-Immersive Environments. In *Proceedings of the Eighth Eurographics Workshop on Virtual Environments*, May 30-31, 2002.
- [31] Lidewij A Wind, Jan van Dalen, Arno M M Muijtjens, and Jan-Joost Rethans. "Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients)." *Blackwell Publishing Ltd, Medical Education* 2004; 38: 39-44.