

Virtual Patients for Future Leaders

**Patrick Kenny, Thomas D. Parsons,
Albert 'Skip' Rizzo**

**Institute for Creative Technologies /
USC**

Marina Del Rey, CA

{kenny,tparsons,rizzo}@ict.usc.edu

Greg M. Reger

CTR USA

USAMEDCOM

greg.reger@us.army.mil

**Caroly Pataki, Michele Pato, Jeff Sugar,
Cheryl StGeorge**

USC Keck School of Medicine

Los Angeles, CA

{pataki,mpato,jsugar,stgeorge}@usc.edu

ABSTRACT

War is one of the most challenging environments that persons may experience. The cognitive, emotional and physical demands of combat environments place enormous stress on even the best-prepared military personnel. The OIF/OEF combat theatre, with its ubiquitous battlefronts, ambiguous enemy identification, and repeated extended deployments have resulted in a significant number of returning American SMs with PTSD and other mental disorders. As a result, military leaders and clinicians in training need to develop clinical skills for identifying potential stress related disorders. Although traditional approaches make use of standard clinic patients to teach, there is limited ability to evaluate skills in a systematic fashion. There is the concern related to the time and money needed to train those involved in the role play for standardized patients. Perhaps most difficult is the “standardization” of standardized patients—will they in fact consistently proffer psychometrically reliable and valid interactions with the training clinicians.

Virtual Human technology has evolved to a point where researchers are developing mental health applications that make use of virtual standardized patients. These virtual patients are embodied characters that have the ability to recognize speech, respond to questions and generate verbal and non-verbal behavior. We have conducted several pilot studies with clinical residents at USC’s Keck School of Medicine and will describe the ongoing study and methodology of our virtual patient approach that allows novice mental health clinicians to conduct an interview with a character emulating PTSD. The paper will summarize the data from the studies and discuss the preliminary standardization of the interactions with the virtual patients. The underlying virtual patient technology will be described. Finally future work will be discussed and recommendations related to the ways in which these characters may enable future leaders to learn, train and win.

ABOUT THE AUTHORS

Patrick G. Kenny has over fifteen years of industry experience working in the field of software development and artificial intelligence. He is currently a researcher and system integration scientist leading the effort to design and integrate the next generation virtual human architecture with cognitive and simulation architectures for the Virtual Human immersive and interactive training environment. Mr. Kenny previously worked at the University of Michigan Artificial Intelligence Lab, researching and developing robotics, mission planning and cognitive models for unmanned ground robotic vehicles. Mr. Kenny is also a founder of Soar Technology Inc. a company specializing in AI and cognitive models. Mr. Kenny’s research interests are in creating highly realistic high fidelity interactive virtual humans, personality models for virtual humans, robotics and gaming. Mr. Kenny has a BS from The University of Minnesota and MS from University of Michigan. He is a member of SIGGRAPH, AAAI and IGDA.

Thomas D. Parsons is a Clinical Neuropsychologist, Assistant Professor, and Research Scientist at the University of Southern California's Institute for Creative Technologies. Dr. Parsons currently co-directs the VRPSYCH Laboratory with Dr. Skip Rizzo, helping to facilitate research integrating Virtual & Augmented Reality,

Psychology, and Social Neuroscience. In addition to his patents, he has over 100 publications. His research makes use of artificial neural networks and virtual/augmented reality to investigate frontal subcortical circuits that underlie neurocognitive functioning and emotion regulation. He has also designed and implemented studies using virtual human agents to assess differential diagnostic abilities of clinicians in training.

Albert ‘Skip’ Rizzo Research Scientist and Research Assistant Professor, Institute for Creative Technologies and School of Gerontology, University of Southern California. Rizzo received his Ph.D. in Clinical Psychology from the State University of New York at Binghamton. He is a Research Scientist at the ISC IT and has a faculty appointment with the USC School of Gerontology. Dr. Rizzo conducts research on the design, development and evaluation of Virtual Reality systems targeting the assessment and training/rehabilitation of spatial abilities, attention, memory, executive function and motor abilities. Additionally, he is conducting research on VR applications that use 360 Degree Panoramic video for exposure therapy (social phobia), role-playing applications (anger management, etc.). He is also investigating the use of VR for pain distraction at CHLA and is currently designing game-based physical rehabilitation VR scenarios for the elderly and persons with impairments due to do Central Nervous System dysfunction. His latest project has focused on the translation of the graphic assets from the Xbox game, Full Spectrum Warrior, into an exposure therapy application for combat-related PTSD with Iraq War veterans. In the area of Gerontology, Dr. Rizzo has served as the program director of the USC Alzheimer’s Disease Research Center and is the creator of the Memory Enhancement Seminars for Seniors (MESS) program at the USC School of Gerontology. He is the associate editor of 4 journals, and is on a number of editorial boards for journals in the areas of cognition and computer technology.

Greg Rigor Greg Reger, PhD: Dr. Greg Reger is a clinical psychologist at the Telehealth and Technology Center of the Defense Center of Excellence. Dr. Reger is a former Army captain who served in Iraq for a year with the 62nd Medical Brigade. Dr. Reger develops and implements virtual reality exposure therapy applications for combat-related PTSD with Iraq War veterans.

Caroly Pataki Clinical Professor, Department of Psychiatry and Behavioral Sciences, Keck School of Medicine, Director of Training, USC Child and Adolescent Psychiatry Residency Program. Caroly Pataki, M.D. has focused her career on the training and education of child and adolescent psychiatrists, demonstrated through her roles for the past decade as Associate Training Director for the Child and Adolescent Psychiatry Residency at UCLA and at USC where she is currently Training Director for the Child and Adolescent Psychiatry Residency Program. Her current passions are to participate in making innovations in the training of child and adolescent psychiatrists who will go on to become leaders in their field. In addition, Caroly Pataki has demonstrated her longstanding devotion to being an educator through her role over the last decade as contributing editor of the widely read psychiatric textbook, Kaplan & Sadock’s Comprehensive Textbook of Psychiatry, in which she seeks to include chapters demonstrating the state of the art technological clinical, teaching and research advances within the field of child and adolescent psychiatry. The current project using Virtual Humans (VHs) to enhance the interview skills and clinical understanding of PTSD in medical students, psychiatry residents, child and adolescent psychiatry residents and even undergraduate students is a perfect fit with Caroly Pataki’s goals to apply technology in an innovative way to enhance psychiatric education to trainees at broad range of levels in order to promote leadership and scholarship among the trainee.

Michele Pato Associate Dean for Academic Scholarship and Professor of Psychiatry and Behavioral Science-Keck School of Medicine. Throughout my 24 year career as a psychiatrist, I have developed a passion and commitment to doing research and teaching others in both clinical care and research endeavors. My academic positions have include administrative appointments as residency training director, medical director at a community mental health center and ACOS for Research and ACOS for Education at VA’s, and my present position as Associate Dean for Academic Scholarship at the Keck School of Medicine. My grant support and publications have run the gamut from peer reviewed publications in the genetics of mental illness, to randomized clinical trials in obsessive compulsive disorder and teaching how to teach, including the development of a 14 session curriculum on basic research methods that I have taught at 5 different academic institutions and whose findings I have published to show how colleagues can be encouraged to contribute to the literature and do research in any area of medicine.

Jeffery Sugar is Chief of Child and Adolescent Psychiatry USC- Keck School of Medicine.

Cheryl St. George MSN, APRN, Instructor of Medical Student Education in Psychiatry, Keck School of Medicine. Over the past seven years, Cheryl has participated in several innovative strategies for teaching and learning as faculty in Medical Student Education at the Keck School of Medicine. She is responsible for developing and implementing various curricular innovations that utilize various teaching technologies in the Psychiatry Clerkship Curriculum that include Psychiatry and Ethics, Culture Competence, and Emergency Psychiatric Assessment. Cheryl has participated in numerous Town and Gown initiatives for more than 40 years and received many awards for leadership in academic and community based projects in Florida, New Mexico and California. She recently presented her educational research project of four years on Ethics and Psychiatry at USC's annual Innovations in Medical Education Conference; and, has been invited to join "Sigma Theta Tau", The International Nursing Honor Society, which awards membership for persons demonstrating outstanding leadership, scholarship and research in nursing.

Virtual Patients for Future Leaders

**Patrick Kenny, Thomas D. Parsons,
Albert 'Skip' Rizzo**

**Institute for Creative Technologies /
USC**

Marina Del Rey, CA

{kenny,tparsons,rizzo}@ict.usc.edu

Greg M. Reger

CTR USA

USAMEDCOM

greg.reger@us.army.mil

**Carolyn Pataki, Michele Pato, Jeff Sugar,
Cheryl StGeorge**

USC Keck School of Medicine

Los Angeles, CA

{pataki,mpato,jsugar,stgeorge}@usc.edu

INTRODUCTION

Virtual patient technology can provide mental health professionals, military leaders and student clinicians with a powerful tool for training in interviewing, assessment, and intervention. There is a growing need in the military to provide training for clinicians and leaders that recognizes mental health signs and symptoms that are indicative of social dysfunction; this is exceedingly true for returning veterans. In a recent Rand (Rand, 2008) report, eighteen percent of the 1.64 million returning veterans were found to have signs and symptoms of Post-Traumatic Stress Disorder (PTSD), major depression or traumatic brain injury (TBI). As a result, these soldiers are at a higher risk of developing familial and social problems, depression, alcohol and drug abuse, and suicide. If only half of those seeking treatment received it, there would be 148,000 soldiers in therapy. If each clinician had four patients, there would need to be 37,000 qualified and trained clinicians to deal with this problem. Currently there are not enough clinicians available. Additionally most officers and leaders obtain minimal or no training on how to recognize or deal with these issues to refer soldiers for help.

Most leaders don't receive any training other than a one page list of things to watch for, in the best case current clinical therapeutic training resort to using real people (hired actors or resident students) acting as simulated patients to portray patients with given medical problems. Such patients are often referred to as Standardized Patients. The problem could be physical or psychological. However the availability of actors and the kinds of conditions they can portray in a relatively standard way is an area of concern. The use of virtual patient technology is not meant to replace these human standardized patients but augment the live actor program with virtual characters that are available

24/7 and can portray a multitude of conditions that might be difficult for actors to represent or repeat with success, additionally being able to have a variety of characters available from elderly and young persons in different genders and cultures will be a benefit.

There are many challenges in creating interactive virtual characters that can be used as virtual standardized patients. Constructing virtual standardized patients requires a considerable amount of Artificial Intelligence technology from speech recognition and natural language to non-verbal behavior animations and 3D computer graphics. An additional complication are the issues involved in building the knowledge and behaviors for the virtual characters to act like human standardized patients with specific mental health problems. Additionally, there is the issue of expertise needed for the designing of effective training systems that can teach the relevant material.

In this paper we present some virtual standardized patient characters that we have designed and developed in our lab. We also describe a series of studies in which we have made use of the virtual patient system for interactions with medical students. The results report the subject testing findings and in the discussion we evaluate the usefulness and effectiveness of virtual standardized patients as a medium to communicate with students, along with an evaluation of the system technology as a whole.

We will also address several of the challenges in building virtual standardized patients and layout some of the research issues. We conclude with future directions in this domain and the ways in which virtual standardized patients may be applied to the training of future leaders.

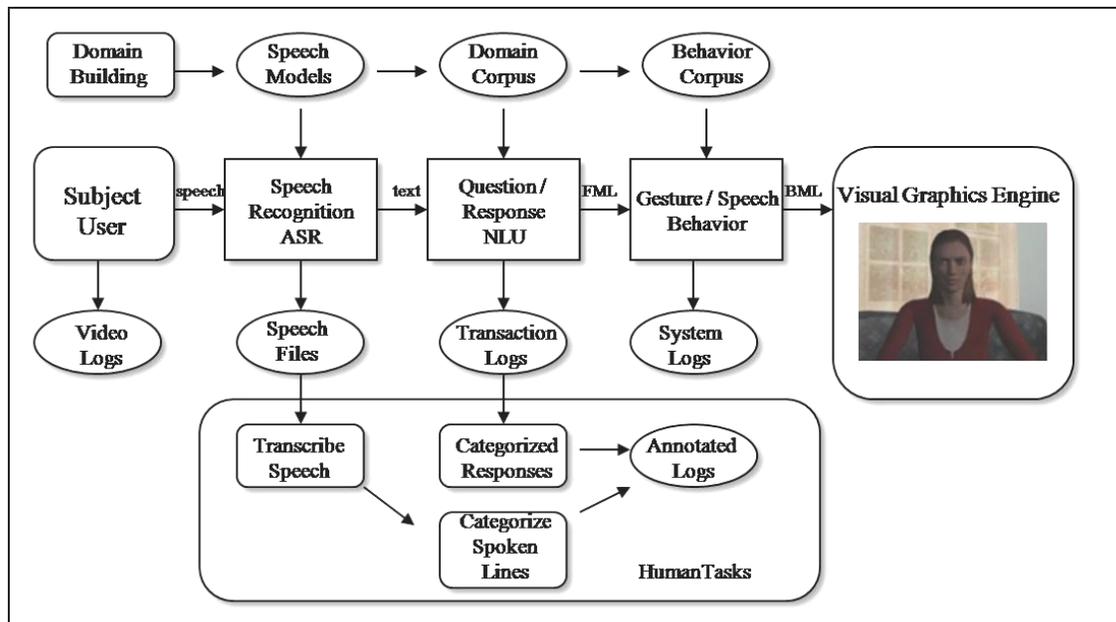
Background and Related Work

Virtual Patients (VPs) are embodied interactive characters (Cassell, 1998) which are designed to simulate a particular clinical presentation of a human patient with a high degree of consistency and realism (Stevens, 2005). There is a growing field of applying virtual reality and virtual patients to issues such as therapy, telehealth, rehabilitation, training and prevention (Rizzo, 2006). VPs have commonly been used to teach bedside competencies of bioethics, basic patient communication, interactive conversations, history taking, and clinical decision making. (Bickmore 2006), (Bickmore, Giorgino, 2006), (Lok, 2006). Results suggest that VPs can provide valid, reliable, and applicable representations of live patients (Andrew, 2006), (Triola, 2006). Virtual patients enable a precise presentation and control of dynamic perceptual stimuli (visual, auditory, olfactory, gustatory, ambulatory, and haptic conditions), along with conversational dialog and interactions, they can provide ecologically valid assessments that combine the veridical control and rigor of laboratory measures with a verisimilitude that reflects real life situations

VIRTUAL PATIENT TECHNOLOGY

Building virtual humans to be used as patients requires a large integrated effort with many components. These patients are 3D computer generated characters that act, think and look like real humans. The users can interact with them through multi-modal interfaces such as speech and vision. The components in the system together form the body, the brain and the environment that the virtual humans exist in.

The VP system is based on our existing virtual human architecture previously presented at IITSEC 2007 (Kenny, 2007). The general architecture supports a wide range of virtual humans from simple question/answering to more complex ones that contain cognitive and emotional models with goal oriented behavior. The architecture is a modular distributed system with many components that communicate by message passing. Because the architecture is modular it is easy to add, replace or combine components as needed.



(Johnsen, 2007), (Parsons, 2007). As a result, VPs provide a reliable testbed from which to perform experiments and better understand the intricacies needed to effectively design and develop the underlying technology. Further, a great deal of work has been accomplished in constructing virtual human technology that allows these characters to implement an extensive array of interactivity (Hubal, 2004).

Interaction with the system works as follows and can be seen in Figure 1. A user talks into a microphone that records the audio signal which is sent to a speech recognition engine. The speech engine converts the signal into text. The text is sent to a statistical question/response selection module. This module picks an appropriate verbal response based on the input text question. The selected response is then sent to a non-verbal behavior generator that selects output gestures, based on a set of rules. That gesture output is combined with the output text to be spoken, pre-

recorded or a computer generated voice, and played through an animation system which synchronizes the gestures, speech and lip syncing for the final output to the screen. The user then listens to the response and asks more questions to the character in an iterative process. Data is logged at each step to help with the evaluation of the technology. We have built two virtual patient characters with this system, but each has different domains. Justin, Figure 3, is an adolescent boy that has conduct disorder (Kenny 2007) and Justina, Figure 2, is a female assault victim character that has PTSD (Parsons 2008).

Research Issues in Virtual Patients

There are many research areas in developing virtual humans and virtual patient technology such as speech recognition, verbal and non-verbal behavior, autonomous agents and tools to build the domains. This research is best done in an iterative process with subject testing to help inform the development and identify the problem areas. The main research areas can be broken down into the following categories:

Speech Recognition

The user has to interact with the system in some fashion. There are several ways this is commonly done; with a pull-down menu interface with scripted choices or with a speech or text interface for more natural interactions. With pull down or scripted interfaces, there are challenges in developing the set of items the user can choose. With a speech interface, where the input is more unconstrained, it is difficult to parse what the user says as speech technology is not 100% accurate. However in limited domains they can still function properly enough. The benefit is that it creates a much more natural interface for the user to interact with the virtual patient as they would with real patients. The drawback is that users tend to think of the technology as working at capacity and don't understand when it does not perform well. There are two major research areas in speech recognition (Narayanan, 2004), proper recognition of voices, the speech model, and the size of the lexicon of words, the language model. Proper speech processing requires understanding the voice from people of all genders, ages and cultures. There is a trade off between having a general speech model that will recognize most voices, but have lower accuracy vs. a specific one, i.e. male or female, that will be better for that specific gender, but needs to be switched for different users. There is also a trade off between the accuracy of words recognized in the language model and the speed the database can be searched to figure out a correct response. It takes several weeks to build and train speech models and deciding on the proper words for

the domain of interest to be put into the language models. This is best done in an iterative process.

Natural Language Dialog

Interactive characters need to engage the user in a realistic conversation. One important aspect of this engagement is the dialog that the subjects and the characters use, both on what they hear, or interpret what they hear, and what they say. There has been much work on natural language understanding and generation for characters (Traum 2007); however it is still a hard problem. One method that seems to work well is a statistical approach (Leuski, 2006), where a corpus of dialog for likely questions and appropriate responses is built. The input questions are matched with output responses by a user when building the dialog, and the system statistical picks the best one during run-time. It is still a challenge to build the dialog corpus for the domain and authoring tools are desperately needed.

Non-Verbal Behavior

Along with the verbal output of the character is the non-verbal behavior. This behavior consists of all the actions, animations, facial expressions, gaze and movements that the character will perform during the dialog exchange. The output gestures all need to be synchronized together with the dialog otherwise the timing will not properly match and it will look unnatural. This non-verbal behavior is difficult to design for virtual patients, because some of the actions that patients with mental health conditions perform may not be natural or normal, for example a muscle or eye tick or looking away for no apparent reason. To generate the non-verbal behavior output in our system is a two phase process. The first phase parses the output text and applies some rules to select animations, for example when the character says; "I don't know", an animation that points at itself is selected. (Lee, 2006) These rules can be designed by the user to match the desired condition. The second phase is synchronizing the selected animations, the output speech, and lip syncing of the text for the character, this is done by a procedural animation system developed called Smartbody (Thiebaut, 2008)

Autonomous Agents

The virtual patient characters should act like real people with real mental conditions. This means they should have their own intrinsic behavior and remember past dialogs or subject areas talked about. They should be able to follow the conversation and add input or behavior on their own with initiative, emotion and personality. (Gratch, 2002) They current virtual patient system does not have an autonomous agent driving the

underlying behavior, the behavior and dialog are driven by the input question. While this is effective, it does not create depth to the character. Developing agents with cognitive models and parameterized behaviors is a good foundational research area. As the mental conditions are better understood, deeper levels of cognitive models can be applied to the characters to give them rich behavior.

Domain Building and Tools

One of the challenges of building interactive VPs that can act as simulated patients has been in enabling the characters to act and carry on a dialog and behavior like a real patient that has the specific mental condition for the domain of interest. This dialog and behavior has to be inputted into the system and this can be a challenge for non-technical people. The process of knowledge acquisition and knowledge design is still a challenge as it requires breadth and depth of expertise in the psychological domain to gather the relevant material for the character and is usually constrained by the underlying technology. Tools to acquire the knowledge and build the domains and still in their infancy and need to be user friendly.

THE PTSD DOMAIN

This current virtual patient project aims to improve the interview skills and diagnostic acumen of psychiatry residents, military leaders and medical students. This is accomplished through interactions with VPs with various signs and symptoms indicative of a given mental health classification. Subjects interview a male patient with conduct disorder, or a female adolescent virtual human with post-traumatic stress disorder (PTSD). We are currently developing a Sgt. Justina for a more ecologically valid interaction within a military relevant environment. (See Figure 6). This military relevant Sgt. Justina would present with signs and symptoms indicative of PTSD following an assault on a military base. The participant's task would be to interview her and glean an adequate history for a preliminary differential diagnosis. The interaction with a military relevant VP provides a context where immediate feedback can be provided regarding the trainees' interviewing skills in terms of psychiatric knowledge, sensitivity, and effectiveness. Use of an embodied natural language-capable virtual character is beneficial in providing trainees with exposure to psychiatric diagnoses such as PTSD that is prevalent in their live patient populations and believed to be under-diagnosed due to difficulty in eliciting pertinent information. Virtual reality patient paradigms, therefore, will provide a unique and important format in which to teach and refine trainees' interview skills

and psychiatric knowledge. Additionally there is a growing need in the current military setting for training leaders to recognize signs of mental problems from returning veterans.

In our first attempt to design a VP 'Justin', Figure 3, we choose conduct disorder as the domain of interest, in which the patient's responses were reflective of someone that would be somewhat resistant to answering questions. Inappropriate or out of domain responses were seen as part of the disorder and this did not negatively impact the interview process. The current domain of PTSD is less forgiving and requires the system to respond appropriately based on certain criteria for PTSD. For the PTSD domain we built an adolescent girl character called Justina, see Figure 2. Justina has been the victim of an assault and shows signs of PTSD.



Figure 2. Justina Virtual Patient



Figure 3. Justin Virtual Patient

Although there are a number of perspectives on what constitutes trauma exposure in children and adolescents, there is a general consensus amongst clinicians and researchers that this is a substantial social problem (Resick, 1997). The effects of trauma exposure manifest themselves in a wide range of symptoms: anxiety, post-trauma stress, fear, and various behavior problems. New clinicians need to come up to speed on how to interact, diagnose and treat this trauma.

One of the challenges of building complex interactive VPs that can act as simulated patients has been in enabling the characters to act and carry on a dialog like a real patient with the specific mental issues present for that condition in the domain of interest. The domain of PTSD requires the system to respond appropriately based on certain criteria for PTSD as described in the DSM manual (309.81; DSM American Psychiatric Association, 2000) . PTSD is divided into six major categories as described in the DSM-IV:

- A. Past experience of a traumatic event and the response to the event.
- B. Re-experiencing of the event with dreams, flashbacks and exposure to cues.
- C. Persistent avoidance of trauma-related stimuli: thoughts, feelings, activities or places, and general numbing such as low affect and no sense of a future.
- D. Persistent symptoms of anxiety or increased arousal such as hyper vigilance or jumpy, irritability, sleep difficulties or can't concentrate.
- E. Duration of the disturbance, how long have they been experiencing this.
- F. Effects on their life such as clinically significant distress or impairment in social or educational functioning or changes in mental states.

Diagnostic criteria for PTSD includes a history of exposure to a traumatic event in category A and meeting two criteria and symptoms from each B, C, and D. The duration of E is usually greater than one month and the effects on F can vary based on severity of the trauma. Effective interviewing skills are a core competency for the clinicians, residents and developing psychotherapists who will be working with children and adolescents exposed to trauma. Rather than assessing for all of the specific criteria, we focused upon the major clusters of symptoms following a traumatic event. Next, we developed two additional categories that we felt would aid in assessing user questions and VP responses that are not included in the DSM:

- G. A general category meant to cover questions regarding establishing rapport, establishing relations, clarifications, opening and closing dialog.
- H. Another category to cover accidental mouse presses with no text, the user is required to press the mouse button while talking.

A clinician needs to ask questions in each of these categories to properly assess the patient's condition.

Table 1 is the types of questions that a typical subject would ask in each of the DSM categories and the kinds of responses that the virtual patient would convey.

Table 1. Question / Response Categories for PTSD

DSM	Question / Response Categories	
	User Questions	Patient Response
(A) Trauma	So, what happened to you that night?	Something really bad happened.
(B) Re-experience	Do you still think about what happened?	I feel like the attack is happening all over again
(C) Avoidance	Do you go out with your friends?	I just stay away from everyone now.
(D) Arousal	Do you feel jumpy?	I feel like I have to watch my back all the time.
(E) Duration	How long has this been going on?	A few months
(F) Life Effect	Are you upset?	Sometimes I don't do anything but stay in my room and cry.
(G) Rapport	Hi Justina, I'm Doctor..	Hello
(H) Other	Button Press	I don't get what you mean.

SUBJECT TESTING

We conducted subject testing of the Justin and Justina characters. Justin was a pilot test to assess the feasibility of the system, while Justina was a more in-depth assessment of the dialog and interaction. This paper will concentrate on Justina, results from Justin can be found here (Kenny, 2007). Participants were asked to take part in a study of novice clinicians interacting with a VP system. They were not told what kind of condition the VP had if any. Two recruitment methods were used: poster advertisements on the university medical campus; and email advertisement and classroom recruitment to students and staff. A total of 15 people (6 females, 9 males; mean age = 29.80, SD 3.67) took part in the study. Ethnicity distribution was as follows: Caucasian = 67%; Indian = 13%; and Asian = 20%. The subject pool was made up of three groups: 1) Medical students (N=7); 2) Psychiatry Residents (N=4); 3) Psychiatry Fellows (N=4). For participation in the study, students were able to forgo certain medical round times.

Measures:

Virtual Patient Pre-Questionnaire. This scale was developed to establish basic competence for interaction with a virtual character that is intended to

be presented as one with PTSD, although no mention of PTSD is on the test.

Justina Pre-questionnaire. We developed this scale to gather basic demographics and ask questions related to the user's openness to the environment and virtual reality user's perception of the technology and how well they think the performance will be. There were 5 questions regarding the technology and how well they thought they might perform with the agent.

Justina Post-questionnaire. We developed this scale to survey the user's perceptions related to their experience of the virtual environment in general and experience interacting with the virtual character in particular the patient in terms of it's condition, verbal and non-verbal behavior and how well the system understood them and if they could express what they wanted to the patient. Additionally there were questions on the interaction and if they found it frustrating or satisfying. There were 25 questions for this form.



Figure 4. Subject Testing Setup

Procedures:

For the PTSD domain we used Justina, who has been the victim of an assault, see Figure 4. The technology used for the system is based on the virtual human technology developed at USC (Kenny et al., 2007; Swartout et al., 2006).

The data in the system was logged at various points to be processed later. Figure 1 is a diagram of how the user interacts with the VP system, described earlier, and the data logging and annotation pipeline. There are four areas where the data is logged. 1) The user speech is recorded from what s/he says; this lets us transcribe what the speech engine processes. 2) A transcript of the entire dialog session is recorded from the question/response system is saved. 3) System logs are stored to allow us to reconstruct what happened in the system if needed. 4) Cameras recorded participant's

facial expressions and system interaction with the patient to be analyzed at a later time.

The set of questions from the user and responses from Justina in the dialog interaction were classified into one of the DSM categories from above. This allowed us to study the responses of the system to questions asked by the subjects to see if they covered all the DSM categories.

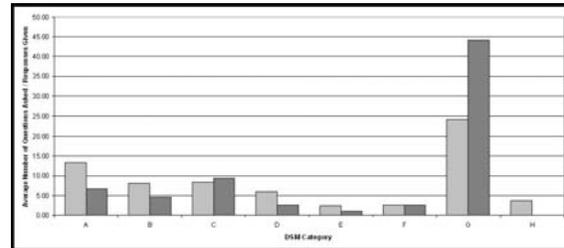


Figure 5. Categorized Questions/Responses

RESULTS

Assessment of the system was completed with the data gathered from the log files in addition to the questionnaires. The log files were used to evaluate the number and types of questions that the subjects were asking, along with a measure to see if the system was responding appropriately to the questions. For a 15 minute interview the participants asked on average, 68.6 questions with the minimum being 45 and the maximum being 91. Figure 5 is a graph showing the average number of questions, asked by the subjects, lighter color, and responses by the system, darker color for each of the 8 DSM categories. It is interesting to note that most of the questions asked were either general questions (Category #G, Average 24 questions) or questions about the Trauma (Category #A, Average 13 questions), followed by category #C and #B, 8. The larger number of questions asked in #G was partially due to clarification questions, however we did not break down the category further to try to classify this. The distribution of questions in each category for each participant was roughly equivalent, which meant in general people asked the same kinds of questions

There are several areas in the system that can be problematic due to technological issues which would cause the system to mis-recognize the question as out of domain, something the natural language system did not know about, and generate an inappropriate response. One such area was speech recognition. We used a speaker independent speech recognizer that did not contain all of the words or phrases asked by the subjects, as it was not known all the questions they would ask. Additionally the system did not perform as

well for women voices as with men. The natural language system deals with out of domain questions by responding with an off topic response, in our case the phrase 'I don't get what you mean'. This was a particular issue, based on the questionnaires, where the subjects got frustrated, as the system responded with this phrase too many times and there was not enough variability with out of domain responses. This response was said in total 411 times across all subjects, comparing that to the total responses of, 1066, the ratio was one in every 2.5 responses. While there is no standard for a reasonable set of questions to out of domain responses, this ratio at least gives us a measure as to how well the system was performing. While this value may seem high and did frustrate some subjects, most subjects were able to continue with questioning and get appropriate responses to perform a diagnosis. Future analysis on the speech recognition word error rate and accuracy will yield data as to what words and questions are needed to improve the speech models. It is clear from the transcriptions that the domain we built was not sufficient to capture all of the questions people were asking, the results from this study will be added to the domain for future testing. The interviewing method that people used to ask questions varied by individual; there were many different styles and personality factors that influenced the length and type of question, for example some people asked multiple segment questions, like 'hi how are you, why did you come here today?'. This is hard to recognize by the system, as it does not have natural language understanding. There are many novice assumptions by the subjects in how well this technology performs.

From the post questionnaires on a 7 point likert scale, the average value subjects rated the believability of the system to be 4.5. Subjects were also able to understand the patient, 5.1. People rated the system at 5.3 as frustrating to talk to, due to speech recognition problems, out of domain questions or inappropriate responses. However most of the participants left favorable comments that they thought this technology will be useful, they enjoyed the experience and trying different ways to talk to the character and also trying to get an emotional response for a difficult question. When the patient responded back appropriately to a question they found that very satisfying.

CONCLUSIONS

The primary goal in this study was evaluative: can a virtual standardized patient generate responses that elicit user questions relevant for PTSD categorization? Findings suggest that the interactions between novice

clinicians and the VP resulted in a compatible dialectic in terms of rapport (Category G), discussion of the traumatic event (Category A), and the experience of intrusive recollections (Category B). Further, there appears to be a pretty good amount of discussion related to the issue of avoidance (Category C). These results comport well with what one may expect from the VP (Justina) system. Much of the focus was upon developing a lexicon that, at minimum, emphasized a VP that had recently experienced a traumatic event (Category A) and was attempting to avoid (Category B) experiences that may lead to intrusive recollections (Category C). However, the interaction is not very strong when one turns to the issue of hyper-arousal (Category D) and impact on social life (Category F). While the issue of impact on social life (Category F) may simply reflect that we wanted to limit each question/response relation to only one category (hence, it may have been assigned to avoidance instead of social functioning), the lack of questions and responses related to hyper-arousal and duration of the illness (Category E) reflects a potential limitation in the system lexicon. These areas are not necessarily negatives for the system as a whole. Instead, they should be viewed as potential deficits in the systems lexicon.

FUTURE WORK

We presented an approach that allows novice mental health clinicians to conduct an interview with a virtual character that emulates an adolescent female with trauma exposure and a male with conduct disorder. The work presented here builds on previous initial pilot testing of virtual patients and is a more rigorous attempt to understand how to build and use virtual humans as virtual patients along with the issues involved in building domains, the speech and language models and working with domain experts.

We will continue to perform more rigorous subject testing with both professional medical students and with non-experts to evaluate how well the different populations perform and further studies in comparing live actor interactions with the virtual patient interactions.

Additional analyses of the data will include more depth comparisons of questions and responses for on and off topic subjects and ways to automate the classification of utterances and improve the speech recognition models. Building an agent framework that will be able to recognize conversation attributes (e.g. opening or closing statements, empathy, topic areas, follow-up and clarification questions) and give the character more autonomous behavior (e.g. asking

questions to the clinician, assertiveness and, initiative levels).

Persons have multifarious interviewing and personality styles. Whilst some prefer a direct and explicit style of interaction, others prefer a more subtle and implicit interaction. The system needs to be able to recognize these styles and adjust its responses accordingly. Studies are needed that incorporate learning objectives into the interview session and that investigate whether the virtual patient system has a learning impact. This is something that we consider to be of high valuable in applying these systems to leadership and clinical training.

It is our belief that the addition of more questions and responses the accuracy of the system will increase along with substance of the conversions interactions between the clinician and the VP. In order to be effective VPs must be interactive, affective (have the ability to react to dialogues with human-like emotions), and communicative—be able to converse in a realistic manner with behaviors and facial expressions that match the clinical condition of interest. The combination of these capabilities allows them to serve as unique training and learning tools whose special knowledge and reactions can be continually fed back to trainees. Whilst our initial goal of this study was to focus on a VP with PTSD, a similar strategy could be applied to teaching a broad variety of psychiatric diagnoses to trainees at multiple levels: medical students, psychiatry residents, child and adolescent psychiatry residents and leaders in the military. Future efforts will be to build the dialog domain for Sgt. Justina and perform subject testing with military leaders. We would like to perform subject testing with a broad range of leaders and clinicians at many military hospitals and bases to have them acquire basic skills needed to assess returning veterans and help improve the leadership of the future.

ACKNOWLEDGEMENTS

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. This work was sponsored in part by a USC Provost grant for Teaching with Technology and the Virtual Humans Project.



Figure 6. Sgt. Justina Character

REFERENCES

- Andrew, R., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Stevens, A., Bernard, T., Oxendine, C., Wagner, P., Lind, S. (2006).: Comparing Interpersonal Interactions with a Virtual Human to those with a Real Human, *IEEE Transactions on Visualization and Computer Graphics*.
- Bernard, T., Stevens, A., Wagner, P., Bernard, N., Schumacher, L., Johnsen, K., Dickerson, R., Raj, A., Lok, B., Duerson, M., Cohen, M., Lind, D.S. (2006). A Multi-Institutional Pilot Study to Evaluate the Use of Virtual Patients to Teach Health Professions Students History-Taking and Communication Skills. *Proceedings of the Society of Medical Simulation Meeting*.
- Bickmore, T., Pfeifer, L., Paasche-Orlow, M. (2007) "Health Document Explanation by Virtual Agents" *Proceedings of the Intelligent Virtual Agents Conference, Paris*.
- Bickmore, T.,Giorgino, T. (2006) "Health Dialog Systems for Patients and Consumers" *Journal of Biomedical Informatics*, 39(5): 556-571.
- Cassell, J., Bickmore, T., Billingham, M., Campbell, L., Chang, K., Vilhjálms, H., and Yan, H.(1998): "An Architecture for Embodied Conversational Characters.", *Proceedings of the First Workshop on Embodied Conversational Characters, October 12-15, Tahoe City, California*.
- DSM, American Psychiatric Association (2000). (*DSM-IV-TR Diagnostic and statistical manual of mental disorders*, 4th edition, text revision. Washington, DC: American Psychiatric Press, Inc.
- Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E.(2002): *Creating Interactive Virtual*

- Humans: Some Assembly Required, IEEE Intelligent Systems, July/August, 54-63.
- Hubal, R., Fishbein, D., Paschall, M., (2004) Lessons Learned using Responsive Virtual Humans for Assessing interaction Skills. *Proceedings of IITSEC*.
- Johnsen, K., Raij, A., Stevens, A., Lind, D., Lok, B. (2007). "The Validity of a Virtual Human Experience for Interpersonal Skills Education" in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM Press, New York, NY, 1049-1058.
- Kenny, P., Parsons, T.D., Gratch, J., Leuski, A., Rizzo, A.A. (2007). Virtual Patients for Clinical Therapist Skills Training. *Lecture Notes in Artificial Intelligence 4722*, 197-210.
- Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., Piepol, D.. (2007). Building Interactive Virtual Humans for Training Environments. *Proceedings of IITSEC*. Nominated for Best Paper.
- Lee, J. & Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA.
- Leuski, A., Patel, R., Traum, D., Kennedy B.(2006): Building effective question answering characters. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, Sydney, Australia.
- Lok, B., Rick F., Andrew R., Kyle J., Robert D., Jade C., Stevens A., Lind, D.S., (2006). "Applying Virtual Reality in Medical Communication Education: Current Findings and Potential Teaching and Learning Benefits of Immersive Virtual Patients", To Appear in *Journal of Virtual Reality*.
- Narayanan, S., and Alwan, A., editors. (2004). *Text to Speech Synthesis: New paradigms and advances*. Prentice Hall.
- Parsons, T.D., Kenny, P., Ntuen, C., Pataki, C.S., Pato, M., Rizzo, A.A., St-George, C., Sugar, J. (2008). Objective Structured Clinical Interview Training using a Virtual Human Patient. *Studies in Health Technology and Informatics*, 132, 357-362.
- Parsons, T.D., Bowerly, T., Buckwalter, J.G., & Rizzo, A.A.(2007):. A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*.
- Resick, P.A., and Nishith, P. (1997). Sexual assault. In *Victims of Crime*, 2d ed., edited by R.C. Davis, A.J. Lurigio, and W.G. Skogan. Thousand Oaks, CA: Sage Publications, Inc., pp. 27-52.
- Rizzo, A.A., Pair, J., Graap, K., Treskunov, A. & Parsons, T.D.(2006): User-Centered Design Driven Development of a VR Therapy Application for Iraq War Combat-Related Post Traumatic Stress Disorder. *Proceedings of the 2006 International Conference on Disability, Virtual Reality and Associated Technology*, 113-122.
- Stevens, A., Hernandex, J., Johnsen, K., et al.(2005): The use of virtual patients to teach medical students communication skills. *The Association for Surgical Education Annual Meeting*, April 7-10; New York, NY.
- Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel J., Traum, D. (2006). Toward Virtual Humans. *AI Magazine*, 27, 1.
- Thiebaut, M., Marshall, A., Marsella, S. and Kallmann, M. (2008). SmartBody: Behavior Realization for Embodied Conversational Agents *Proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS)*.
- Traum, D., Antonio Roque, Anton Leuski, Panayiotis Georgiou, Jillian Gerten, Bilyana Martinovski, Shrikanth Narayanan, Susan Robinson, and Ashish Vaswani. (2007). "Hassan: A virtual human for tactical questioning". In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 71-74, Antwerp, Belgium.
- Triola, M., Feldman, H., Kalet, A.L., Zabar, S., Kachur, E.K., Gillespie, C. (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of General Internal Medicine*, 21, 424-429.
- RAND Report MG720, (2008). Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery http://rand.org/pubs/monographs/2008/RAND_MG720.pdf