Virtual human patients for training of clinical interview and communication skills

T D Parsons, P Kenny and A A Rizzo

Institute for Creative Technologies, University of Southern California,
Los Angeles, California, USA

tparsons@usc.edu, *arizzo@usc.edu*

http://vrpsych.ict.usc.edu/

ABSTRACT

Although schools commonly make use of standardized patients to teach interview skills, the diversity of the scenarios standardized patients can characterize is limited by availability of human actors. Virtual Human Agent technology has evolved to a point where researchers may begin developing mental health applications that make use of virtual reality patients. The work presented here is a preliminary attempt at what we believe to be a large application area. Herein we describe an ongoing study of our virtual patients. We present an approach that allows novice mental health clinicians to conduct an interview with virtual character that emulates 1) an adolescent male with conduct disorder; and 2) an adolescent female who has recently been physically traumatized.

1. INTRODUCTION

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. The effects of trauma exposure manifest themselves in a wide range of symptoms: anxiety, post-traumatic stress disorder, fear, and various behavior problems. Trauma exposure is associated with increased risk of psychological problems in adulthood. Effective interview skills are a core competency for the clinicians who will be working with children and adolescents exposed to trauma.

Developing effective interviewing skills for the clinicians, residents and psychotherapists who will be working with children and adolescents exposed to trauma is a necessary skill. A clinician needs to ask various questions relating to the trauma and its effect to properly assess the patient's condition. Current therapeutic training systems resort to using real persons (hired actors or resident students) acting as standardized patients to portray patients with a given mental health problem in what is called an Objective Structured Clinical Examination (OSCE). The problem portrayed by the actor could be physical or psychological. Although schools commonly make use of standardized patients to teach interview skills, the diversity of the scenarios standardized patients can characterize is limited by availability of human actors and their skills. This is an even greater problem when the actor needs to be an adolescent. The potential of using computer generated virtual humans as standardized virtual patients (VPs) for use in clinical assessments, interviewing and diagnosis training is becoming recognized as the technology advances (Bernard et al., 2006; Bickmore, Pfeifer, & Paasche-Orlow, 2007). These VPs are embodied interactive agents who are designed to simulate a particular clinical presentation of a patient with a high degree of consistency and realism (Kenny et al., 2007). VPs have commonly been used to teach bedside competencies of bioethics, basic patient communication, interactive conversations, history taking, and clinical decision making (Bickmore, & Giorgino, 2006).VPs can provide valid, reliable, and applicable representations of live patients (Triola et al., 2006). Research into the use of VPs in psychotherapy training is in its nascent stages (Johnson et al., 2007; Parsons et al., 2008). Since virtual humans and virtual environments can allow for precise presentation and control of conversations and interactions, they can provide ecologically valid assessments that combine the control and rigor of laboratory measures with a verisimilitude that reflects real life situations.

The current project aims to improve child and adolescent psychiatry residents, and medical students' interview skills and diagnostic acumen through practice with a female adolescent virtual human with post-traumatic stress disorder (PTSD). This interaction with a virtual patient provides a context where immediate

 ISBN 07 049 15 00 6

feedback can be provided regarding trainees' interviewing skills in terms of psychiatric knowledge, sensitivity, and effectiveness. Use of a natural language-capable virtual character is beneficial in providing trainees with exposure to psychiatric diagnoses (e.g. PTSD), 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. In order to be effective, virtual humans must be able to interact in a 3D virtual world, must have the ability to react to dialogues with human-like emotions, and be able to converse in a realistic manner. The combination of these capabilities allows them to serve as unique training tools whose special knowledge and reactions can be continually fed back to trainees. The goal of this virtual patient was to focus on a character with PTSD, our previous effort was on a character with Conduct Disorder. The eventual goal is to build a library of characters with a variety of psychiatric diagnoses to train residents and students at multiple levels.

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 (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).

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 I/ITSEC 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.

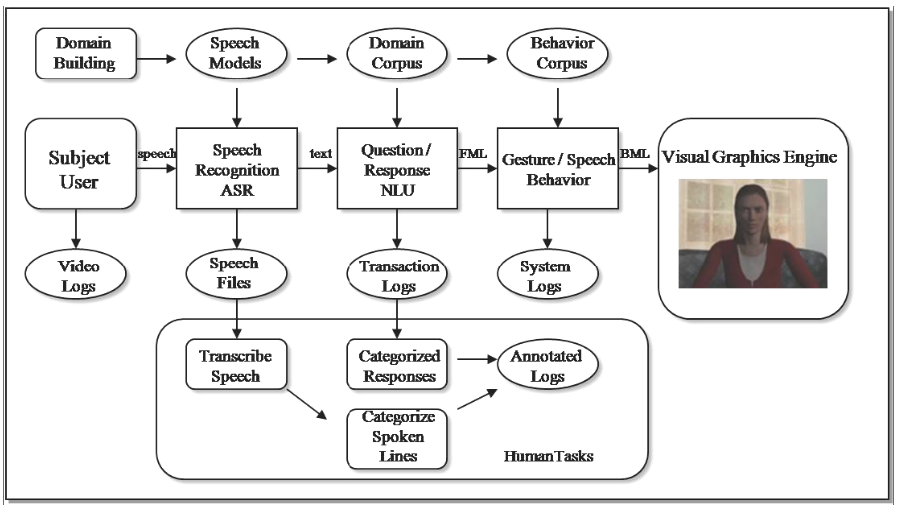

Figure 1. *Diagram Describing a Participant's Interaction with the System.*

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

©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

appropriate verbal response based on the input text question. The selected response is then sent to a nonverbal 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 2, is an adolescent boy that has conduct disorder (Kenny 2007) and Justina, Figure 3, is a female assault victim character that has PTSD (Parsons 2008).

Figure 2. *Justin Virtual Patient.*

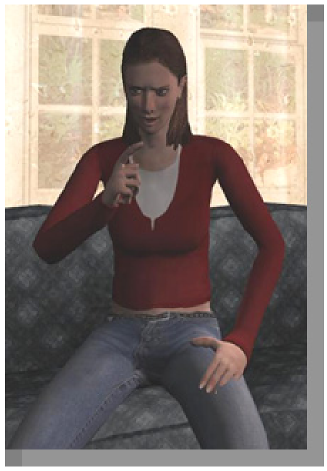

Figure 3. *Justina Virtual Patient.*

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 patents may be applied to the training of future clinicians.

2. 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:

2.1 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.

2.2 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.

2.3 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 synching of the text for the character, this is done by a procedural animation system developed called Smartbody (Thiebaux, 2008)

2.4 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.

2.5 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.

3. 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). The interaction with a 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 underdiagnosed 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.

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.

4. 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 indepth 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.

4.1 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 its 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.

4.2 Procedures

For the PTSD domain we used Justina, who has been the victim of an assault. 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.

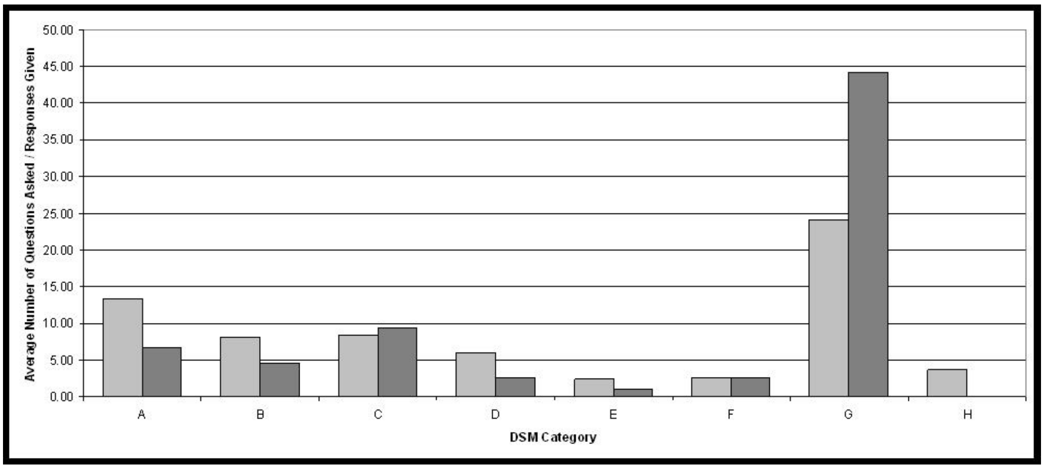

Figure 4. *Categorized Questions/Responses.*

5. 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 4 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.

6. 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) experienced that my 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.

A secondary goal was to investigate the impact of psychological variables upon the VP Question/Response composites and the genreal believability of the system. After controlling for the effects of these psychological variables, increased effects were found for discussion of the traumatic event (Category A), avoidance (Category C), hyper-arousal (Category D), and impact on social life (Category F). Further, the impact of psychological characteristics revealed strong effects upon presence and believability. These findings are consistent with other findings suggesting that hypnotizability, as defined by the applied measures, appears moderate user reaction. Future studies should make use of physiological data correlated with measures of immersion to augment and quantify the effects of virtual human scenarios.

Herein we described an ongoing study of our Virtual Patient System. 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. 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 and the many issues involved in building domains, speech and language models and working with domain experts. The lessons learned here can be applied across any domain that needs to build large integrated systems for virtual humans. We believe this is a large and needed application area, but it's a small enough domain that we can perform some serious evaluations on using virtual humans in real settings.

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 in the types of questions asked. Additionally further studies in comparing to real OSCE's with real actors to the virtual patient will be performed. Additional incorporation of rapport [7,12] using the facial gestures analysis with the system will further enhance the virtual patient interaction to produce more results in this domain.

Additional analyses that need to be performed with the data include: investigate the domains questions and responses to assess how many were on-topic and how many off topic; How well did the speech recognition perform based on word error rate; How did the speech recognition, graphics and non-verbal impact the subjects, their interview experience, presence and immersion in the system?; Can we automate the process of extracting data from large corpus of speech data in this domain to build topic areas?; Can we automate the process of classifying the subjects questions into the DSM categories from the speech or transcriptions of the speech? Define further sub-categories for interactive conversions, such as; opening, closing, empathy, topic area, follow-up, query, clarification, self-disclosure to name a few and annotate the transcriptions with these categories. This will help us to build better tools to build domains and characters.

It is our belief that with more questions covered in the domain the accuracy of the system will go up along with the depth of the conversions which will further enhance the virtual patient system. In order to be effective virtual humans must be able to interact in a 3D virtual world, must have the ability to react to dialogues with human-like emotions, and be able to converse in a realistic manner with behaviors and facial expressions. 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. Our initial goal of this study was to focus on a VP with PTSD, but a similar strategy could be applied to teaching a broad variety of psychiatric diagnoses to trainees at every level from medical students, to psychiatry residents, to child and adolescent psychiatry residents.

7. REFERENCES

T Bernard, A Stevens, P Wagner, N Bernard, L Schumacher, K Johnsen, R Dickerson, A Raij, B Lok, M Duerson, M Cohen and D S Lind (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.*

T Bickmore, L Pfeifer and M Paasche-Orlow (2007), Health Document Explanation by Virtual Agents, *Proceedings of the Intelligent Virtual Agents Conference, Paris.*

T Bickmore and T Giorgino (2006), Health Dialog Systems for Patients and Consumers, *Journal of Biomedical Informatics, 39(5):* 556-571.

K Johnsen, A Raij, A Stevens, D Lind and B Lok (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.

P Kenny, T D Parsons, J Gratch, A Leuski and A A Rizzo (2007), Virtual Patients for Clinical Therapist Skills Training, *Lecture Notes in Artificial Intelligence 4722,* 197-210.

P Kenny, A Hartholt, J Gratch, W Swartout, D Traum, S Marsella and D Piepol, (2007), Building Interactive Virtual Humans for Training Environments, *Proceedings of I/ITSEC.*

T D Parsons, P Kenny, C Ntuen, C S Pataki, M Pato, A A Rizzo, C St-George and J Sugar (2008), Objective Structured Clinical Interview Training using a Virtual Human Patient, *Studies in Health Technology and Informatics, 132,* 357-362.

W Swartout, J Gratch, R Hill, E Hovy, S Marsella, J Rickel and D Traum (2006), Toward Virtual Humans, *AI Magazine, 27,* 1.

M Triola, H Feldman, A L Kalet, S Zabar, E K Kachur and C Gillespie (2006), A randomized trial of teaching clinical skills using virtual and live standardized patients, *Journal of General Internal Medicine, 21,* 424-429.