Virtual Patients for Clinical Therapist Skills Training

Patrick Kenny, Thomas D. Parsons, Jonathan Gratch,
Anton Leuski, and Albert A. Rizzo

Institute for Creative Technologies,
University of Southern California
13274 Fiji Way Marina Del Rey, CA 90292, USA
{kenny,parsons,gratch,leuski,rizzo}@ict.usc.edu

Abstract. Virtual humans offer an exciting and powerful potential for rich interactive experiences. Fully embodied virtual humans are growing in capability, ease, and utility. As a result, they present an opportunity for expanding research into burgeoning virtual patient medical applications. In this paper we consider the ways in which one may go about building and applying virtual human technology to the virtual patient domain. Specifically we aim to show that virtual human technology may be used to help develop the interviewing and diagnostics skills of developing clinicians. Herein we proffer a description of our iterative design process and preliminary results to show that virtual patients may be a useful adjunct to psychotherapy education.

Keywords: Virtual Humans, Virtual Patients, Psychopathology.

1 Introduction

Virtual human technology may provide mental health professionals with a powerful tool for assessment, intervention, and training. This technology offers exciting potential for rich interactive experiences. Current therapeutic training systems resort to using real people (hired actors or resident students) acting as simulated patients to portray patients with given medical problems. The problem could be physical or psychological. Whilst the use of technology to replace or augment simulated patients has not been widely applied or accepted, a search of the literature of interactive virtual characters reveals only a handful of studies. Part of the problem has been the difficulty of building complex interactive virtual characters that can act as simulated patients. An additional complication has been the technological issues involved in trying to get interactive virtual characters to act like real patients. On top of all this has been the expertise in designing effective training systems that can teach the relevant material. The work presented here is a preliminary attempt at what we believe to be a large application area. Herein we describe an initial endeavor to apply our virtual characters as virtual patients (VP).

We present an approach that allows novice mental health clinicians to conduct an interview with a virtual character that emulates an adolescent male with conduct disorder. The paper will also describe the theory and praxes involved in creating the character with psychological problems along with issues and lessons learned as to

C. Pelachaud et al. (Eds.): IVA 2007, LNAI 4722, pp. 197–210, 2007.

how it can be applied in training novice therapists to perform interviews and differential diagnosis. Aspects of preliminary subject testing of the system will be discussed along with a few example dialog sessions with the character. The final section will discuss proposals for future work and modifications to the current system to make the VP more engaging. Although we are in the early stages of developing this system, the initial outcome of the tests was favorable. The paper also illustrates how a variety of core research components developed at the University of Southern California facilitates the rapid development of mental health applications.

2 Related Work

Virtual humans are increasingly being recognized as useful tools for training, education, research, and entertainment. Work in virtual humans covers a broad array of tasks that require an integrated system for a fully embodied conversational character.

Fully embodied conversational characters have been around for since the early 90's [4]. There has been much work on full systems to be used for training [9,20,21], intelligent kiosks [16], and virtual receptionists [2]. We have had previous involvement with systems that use virtual reality for PTSD [22] and ADHD [18,23].

VPs are virtual interactive agents who are trained to simulate a patient's particular clinical presentation with a high degree of consistency and realism. VPs have commonly been used to teach bedside competencies in bioethical decision making, basic patient communication and history taking, and clinical decision making [6,10,17,26]. VPs can provide valid, reliable, and applicable representations of live patients [29]. For example, in an application from Lok's research group, instead of having novice medical students practice on professional patients, they constructed a virtual environment to represent an examination room where a VP could be interviewed verbally with speech recognition [1]. The goal in this application was to determine, via clinical interview, whether the VPs ailment was due to appendicitis. Results suggested that the virtual interaction was similar to the real interaction on content measures and participants gathered the same information from the virtual human and real patient.

Research into the use of VPs in psychotherapy training is very limited [8,11]. Beutler and Harwood [3] describe the development of a VR system for training in psychotherapy and summarize training-relevant research findings. We could not find reference to any other use of VPs in a psychotherapy course to date, despite online searches through MEDLINE, Ovid, and the psychotherapy literature. Designing VPs that have human-to-human interaction and communication skills would open up possibilities for clinical applications that address interviewing skills, diagnostic assessment and therapy training.

3 Virtual Patient Domain

We choose the medical field domain for this application, specifically cognitive behavior therapy (CBT). The mental health domain offers some interesting challenges to both the design of the characters and the design of the training system to enhance skills in interviewing, differential diagnosis, and therapeutic communication. The domain also offers a plethora of modeling issues including: verbal and non-verbal behavior, cognition, affect, rational and irrational behavior, personality, and psychopathology.

For our application we choose to model a character with the history and symptoms of conduct disorder (described below). We wanted to take on a problem that wasn't too hard to model and could be structured as an interview, where one could ask questions and get responses. We endeavored to constrain the domain so that the character would only discuss certain topics, which were decided upon a priori. To approach this problem we gathered data about patients with conduct disorder so that we could include their characteristics. We also spoke with experienced psychologists to develop a library of typical questions that a novice clinicians might ask a person with conduct disorder and the kinds of responses (verbal and nonverbal) persons with conduct disorder might give. Additionally, we needed something that was relevant to the types of training we would like to exemplify in the system, specifically clinical interviewing skills and diagnosis of a problem.

The data was gathered through role-playing exercises, subject testing, consulting manuals, and soliciting knowledge from subject matter experts. The methodology used was an iterative process of data collecting, testing, and refining. While the preliminary goal of the project was to use the VP to teach diagnostic skills training, the eventual goal is to have the VP be utilized in individual trainee interviews, small group and classroom settings.

3.1 Skills Training

Teaching interviewing skills with virtual humans and VPs is still a young discipline. The common practice is to use real humans (i.e simulated patients) to play the roles of the patients. A general complication involved in teaching general interviewing skills is that there are multiple therapeutic orientations and techniques to choose from and it is not well understood how to properly implement each of these with a VP. There are no standard methods and metrics to measure what works for the different types of interviews given to patients in the multitude of different mental health problems. To alleviate this problem we are concentrating on teaching skills required to diagnose a particular type of mental disorder called conduct disorder. Our goal is to obtain objective data from an initial intake interview. An intake interview is the first interview that a clinician conducts with a patient. The clinician may have some knowledge of why the patient is there (i.e. a referral question), but needs to ask the patient further questions to obtain a detailed history to narrow down the problem for differential diagnosis and treatment planning. The system is designed to allow novice clinicians to practice asking interview questions in an effort to create a positive therapeutic alliance with this very challenging VP.

3.2 Conduct Disorder

The project involved the construction of a natural language-capable VP named "Justin." The clinical attributes of Justin were developed to mimic a conduct disorder profile as found in the Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR [5]. Justin portrays a 16-year old male with a conduct disorder who is being forced to participate in therapy by his family. Justin has a history of a chronic pattern of antisocial behavior in which the basic rights of others and age-appropriate societal norms are violated. He has stolen, been truant, broken into someone's car, been cruel to animals, and initiated physical fights.

For conduct disorder the trainee's interview questions should be guided by eliciting information regarding the four general symptom categories prevalent in conduct disorder:

- Aggressive behavior – e.g. fighting, bullying, being cruel to others or animals
- Destructive behavior – e.g. arson, vandalism
- Deceitful behavior – e.g. repeated lying, shoplifting, breaking into homes or cars
- Violation of rules – e.g. running away, engaging in non appropriate behavior for age

The VP system is designed to provide answers to questions that target each of these categories and will respond to a variety of questions pertinent to these areas. More detail of how this is accomplished is seen in the architecture section. Some responses by the VP may be on target, off target, involve "brush away" responses, and in some cases, they may be irrelevant replies. For example if the trainee asks: "How are things going at home" or "Are you having any problems at home" or "How are things going?", the system will respond with "My parents think I messed up." Further questions will lead to finding out that the patient has been running away. This will lead to marking one of the above categories true for the diagnosis in the trainees' interview. In order for the trainee to pass, it will require responses in all of the categories. One important distinction between this VP system and a general question response text or speech based system is the ability to use multimodal presentation in the interaction with the patient. The character will respond with gestures along with speech. The dialog of the embodied character is synchronized with non-verbal behavior when the patient answers questions. For example a brush off response would trigger an arm to swing out as to push the issue aside. Being able to develop more multimodal behavior for each of the categories is anticipated to make this a powerful asset over just a text interface. The total set of questions, responses, behavior patterns and interview interactions are extracted from role-playing exercises, initial subject testing, interviews with doctors and common sense. In total the question set consists of over 100 questions and 70 responses. The current set of gestures consists of over 20 distinct gestures in a sitting pose and 30 standing gestures in 4 poses; hands at sided, crossed arms and right and/or left hand on hip.

4 Virtual Patient Architecture

The VP system is based on our existing virtual human architecture [9,27]. 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. For example in the larger virtual human architecture the natural language section is divided into three components: a part to understand the language, a part to manage the dialog and a part to generate the output text. This is all combined into one component for the VP system.

Interaction with the system works as follows and can be seen in Figure 1. A user talks into a microphone which records the audio signal that is sent to a speech recognition engine. The speech engine converts that into text. The text is then sent to a statistical response selection module. The module picks an appropriate verbal response based on the input text question. The response is then sent to a non-verbal behavior generator that selects animations to play for the text, based on a set of rules. The output is then sent to a procedural animation system along with a pre-recorded or a generated voice file. The animation system plays and 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.

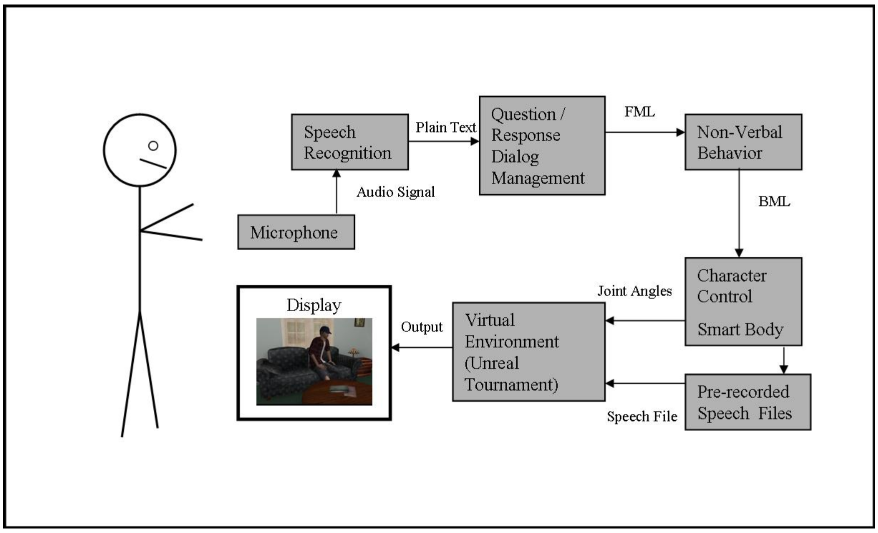

Fig. 1. This is a picture of the Virtual Patient Architecture and data flow

4.1 Human Speech Input

A human user talks to the system using a head-mounted close-capture USB microphone. The user's speech is converted into text by an automatic speech recognition system. We used the SONIC speech recognition engine from the University of Colorado, Boulder. [19]. We customized the engine’s acoustic and language models for the domain on interest [25]. In general a language model is tuned to the domain word lexicon. We collect user’s voice data during each session, it allows us to go over the data to collect words not recognized to enhance the lexicon and also to get error rates to compare the input speech with the processed output speech. The speech recognition engine processes the audio data and produces the text of the user’s utterance. It then packages the text into a message and sends it to the response selection module.

4.2 Response Selection

The response selection module receives a text message from the speech recognition module, analyzes the text, and selects the most appropriate response. The virtual human system has a number of pre-defined response lines and given a user's utterance, the response selection module has to choose a single answer among them. This response selection process is based on a statistical text classification approach developed at the Natural Language group at ICT [14]. The approach requires a domain designer to provide some sample questions for each system response. When a new question comes from the speech recognition module, the system uses the mapping between the answers and sample questions as a "dictionary" to "translate" the question into a representation of a "perfect" answer. It then compares that representation to all known text answers and selects the best match. This approach was developed for the SGT Blackwell virtual human project [15] and has been shown to outperform traditional state-of-the-art text classification techniques.

To facilitate rapid development and deployment of virtual human agents with similar capabilities the Natural Language group created NPCEditor, a user friendly software package for editing the system answers, adding sample questions, training and running the text classifier. There is no limit to the number of answers or sample questions, but it is advised to have at least two or three sample questions for each answer. The system allows for several answer categories. Sometimes the system combines the text from answers of different categories to produce the final response. The only category required is on-topic responses, the others are optional, but make the system more interactive and realistic. The category types are as follows:

- On-topic – These are answers that are relevant to the domain of the conversation. These are the answers the system has to produce when asked a relevant question. Each on-topic answer should have a few sample questions and single sample question can be linked to several answers. The text classifier generally returns a ranked list of answers and the system makes the final selection based on the rank of the answer and whether the answer has been used recently. That way if the user repeats his questions, he may get a different response from the system.
- Off-topic – These are answers for questions that do not have domain-relevant answers. They can be direct, e.g., "I do not know the answer", or evasive, e.g., "I will not tell you" or "Better ask somebody else". When the system cannot find a good on-topic answer for a question, it selects one of the off-topic lines. More details on off-topic answer classification can be found elsewhere [24].
- Repeat –If the classifier selects an answer tagged with this category, the system does not return that answer but replays the most recent response. Sample questions may include lines like "What was that?" or "Can you say that again?" Normally, there is at most one answer of this category in the domain answer set.
- Alternative – If the classifier selects an answer tagged with this category, the system attempts to find an alternative answer to the most recent question. It takes the ranked list of answers for the last question and selects the next available answer. Sample questions may include lines like "Do you have anything to add?" Normally, there is at most one answer tagged with this category in the answer set.

- Pre-repeat – Sometimes the system has to repeat an answer. For example, it happens when a user repeats a question and there is only one good response available. The system returns the same answer again but indicates that it is repeating itself by playing a pre-repeat-tagged line before the answer, e.g., "I told you already." There is no need to assign sample questions to these answer lines.
- Delayed – These are the lines from the system that prompt the user to ask about a domain related thing, e.g., "Why don't you ask me about…" Such a response is triggered if the user asks too many off-topic questions. The system would return an off-topic answer followed by a delayed-tagged answer. That way the system attempts to bring the conversation back into the known domain. This category has no sample questions assigned.

Once the output response is selected, it is packaged up into a FML (Functional Markup Language) message structure. FML allows the addition of elements such as affect, emphasis, turn management, or coping strategies. For the VP, the response selection module does not add any additional information besides the text.

4.3 Behavior Generation

The FML message is sent to the non-verbal behavior (NVB) generator which applies a set of rules to select gestures, postures and gazes for the virtual character. Since the VP in this application was sitting down, the animations mainly consisted of arm movements, wave offs and head shakes or nods. The VP character does not do any posture shifts or go into standing posture, although in the next version more gestures and postures will be added. Once the NVB selects the appropriate behavior for the input text, it then packages this up into a Behavioral Markup Language (BML) [12] structure and sends it to a procedural animation system. For more detail on how the NVB works, see the paper on this at last years IVA06 [13].

4.4 Character Control and Output

This last part of the process is the execution and display of the characters multimodal behavior. This is accomplished with a procedural animation system called Smartbody [28]. Smartbody takes as input the BML message that contains the set of behaviors that need to be executed for the head, facial expressions, gaze, body movements, arm gestures, speech and lip syncing and synchronizes all of this together. These behaviors need to be in sync with the output speech to look realistic. Smartbody is capable of using generated or pre-recorded speech. The VP used pre-recorded speech.

Smartbody is hooked up to a visualization engine, in this case the Unreal Tournament game engine for the graphics output. Smartbody controls the character in the game engine and also specifies which sound to play for the characters speech output. Smartbody is also capable of having controllers that perform specific actions based on rules or timing information, such as head nods. The controllers are seamlessly blended in with the input animations specified in the BML. The VP does not make extensive use of controllers, however, future work is to design controllers for certain behavior patterns such as gaze aversion or ticks the character might have. A motex, which is a looping animation file, can be played for the character to give is a bit of sway, or in the VP case, finger tapping.

Fig. 2. Virtual Patient "Justin" in the clinician's office

4.5 Artwork

The artwork plays a crucial role in defining the characters behavior, attitude and condition. People are able to make a judgment about someone within the first few seconds. The project involved the development of a VP named "Justin", see Figure 2. We wanted a sixteen year old boy that has some kind of mental problem, but we wanted to keep the character design general so that the artwork would not be tied to a specific medical condition, for example giving him a broken arm. For the boy we wanted a typical teenager with a T-shirt, blue jeans and baseball hat. One must be careful with the design of the character as everything can lead to questions by the users. For example the character has a rip in the pants. This was seen, but not realized until one of the subject testers asked the patient where he got the rip in the pants. Since this was not anticipated, there were no appropriate responses except a brush-off.

The project also involved the development of a clinical virtual environment in which the trainees learn interviewing techniques. The environment was modeled after a typical clinician's office and was meant to represent a place that would make the patient feel at home.

5 Evaluation and Discussions

The purpose of the evaluation was two fold: 1) General assessment of the system's capacity for interactive response to questions posed to the VP; and 2) Specific application of the system's performance when interacting with a novice clinician as he or she conducted an intake interview and relevant psychiatric history of the VP. The sample of participants included six persons from the University of Southern

California's Keck School of Medicine. Two staff members associated with the project administered the evaluations. Initial inclusion criteria for these evaluations required that participants have previous clinical therapy skills. The desired level of clinical competence, however, was not readily available. As a result, our sample size was limited to an N of 3.

The method of evaluation was conducted in three phases: 1) Pre-Questionnaire Phase: an initial pre-questionnaire assessed the participant's skill level for spoken dialog systems and clinical skill set relevant to differential diagnosis of adolescent behavior problems; 2) Interview Phase: a 30 minute interactive interview with the VP (i.e. Justin); and 3) Post-Questionnaire Phase: following the interview a post-questionnaire assessed the participant's knowledge of the VPs condition, the interaction with the VP and the system in general. The post-questionnaire had questions ranked on a 7 point Likert scale with some area for comments. In total there were 35 questions that made up the pre and post forms. The initial results presented here will be used to adjust the system for a more formal evaluation.

5.1 Assessment of the System

Assessment of the system was completed by 1) experimenter observation of the participants as they communicated with the VP in addition; 2) the questionnaires. In order that the system be adequately evaluated, we determined a number of areas that needed to be addressed.

- The behavior of the VP should match the behavior one would expect from a patient in such a condition. Specific behaviors included: verbalization, gesture, posture, and appearance.
- Adequacy of the communicative discourse between the VP and the participants.
- Proficiency (e.g. clarity, pace, utility) of VPs discourse with the participant.
- Quality of the speech recognition of utterances spoken.

The results from the survey suggest that the system performed well and participants reported that the system simulated the real-life experience (i.e. ranked 5 or 6). The verbal and non-verbal behavior also ranked high (i.e. between 5 and 7).

Some participants found aspects of the experience "frustrating", mainly because they were not able to receive anticipated responses and the system's tendency to repeat some responses too many times. This was due to the speech recognition's inability to evaluate some of the stimulus words. Further, there were too many "brush off" responses and questions asked that were outside the dialog set previously constructed. The VP seems to be slightly more resistant than anticipated. As a result, it was difficult for participants to get the VP to elaborate on matters that were initially presented. This occurred, for example, when a participant attempted to acquire more detailed information related to the VPs peer-relations, drug use, and familial relations. The level of depth in a topic area limited the VPs response set to only a cursory level and as a result, further details were not forthcoming. When the participants tried to ask questions about the familial relations from a different vantage point the same responses or "brush off" responses were chosen. Whilst this is desirable in that it pushes the participant to apply multiple interviewing tactics, the system may also seem rigid and unresponsive if there is not enough feedback from the VP, This could be alleviated if an observer or instructor was watching the trainee and provided necessary feedback in those instances in which the system cannot.

There is a concern that participants ascribe characteristics to the VP which in fact are not present. For example, although the VP responded “Yes” to a question about whether the VP hurt animals, in actuality the system did not recognize the input speech. This can lead to confusion later on if the character then responds differently. In fact one of the most substantial lessons learned was the amount of conversation state that needed to be tracked for the topics and questions asked by the participant so that the VP’s responses would be consistent throughout the session.

5.2 History Taking and Interviewing Assessment

If the system was to be used in helping to train clinicians then we would want to know if the subjects could identify a behavior problem in the VP. Specifically, we would want to assess the clinician’s ability to differentially diagnose the VPs conduct disorder. We asked several questions about what condition the participants thought the VP had, what led to this conclusion, how early they diagnosed the condition, and if the condition was believable to them. Additionally we had them select from conduct disorder categories, what they gathered from interaction with the character, and what problems they thought the VP had. These were multiple choice questions where participants could choose from home, school, parents, friends or life in general. Most participants selected school, parents, and life in general. Although there were responses in all categories, the ones the participants selected were the ones with the most responses. The participants tended to focus more on a single area, rather than asking questions that would allow them to have a broad understanding of the VPs psychiatric history. This may reflect the fact that these were novice clinicians, easily swayed by a desire for a quick diagnosis instead of a full clinical picture for differential diagnosis. The data from the initial intake interview suggests that overall the participants were able to establish that the VP was resistant, had problems with his family, school and life, but could only attribute this to a behavior problem and not conduct disorder. The participants did not ask appropriate questions, which would have elicited information regarding the four general symptom categories prevalent in conduct disorder. As a result, they were unable to prompt the system to offer the correct responses. Again, this reflects the novice level of these clinicians and the need for a qualified supervisor to help direct their development. A more experienced clinician would know the importance of seeking a full clinical picture so that she may then explore in more depth the referral issues.

5.3 Results and Lessons Learned

Based on the subject testing the question and response set needs to be expanded to include more detail about the conditions in each of the conduct disorder categories, and have the system offer up more information that would lead them to asking more questions about topics that are covered in the domain.

People unfamiliar with speech systems tend to have run-on or multi-subject sentences, i.e. ask many questions in one speech utterance. The current system does not deal with this well and usually ends up with a brush off or incorrect response choice. This can frustrate the user as they expect the system to be able to do more then it can. This can lead to negative training, or it could also be seen as good training to get the user to think about the questions they are asking. There are some proper ways to conduct an interview as described in [7], but everyone has their own questioning style they are comfortable with and the system would need to be able to handle each type. The subjects enjoyed the experience, thought it could be useful tool, and it gave them freedom to think and explore the problem space.

One question that should be discussed is; Can you participate in a clinical interview without discourse memory, or the ability for the patient to remember more then the last one or two questions. One of the complaints that we saw in subject using these kinds of question response systems are the number of repeat responses. Having a way to track the discourse would ultimately benefit the usability, however it would make authoring the dialog more complex as this would increase the questions and response lines required. It is unclear how much discourse memory is needed in this application and future studies are required for that.

6 Conclusion

This study was our initial prototype of building an interactive VP that was capable of discourse with novice clinicians so that they may establish the VPs clinical history and differential diagnosis. We described the domain, the architecture, the subject testing and evaluation conducted. The plan is to take the results and lessons learned from the evaluation and apply those to a more formal study. For the current study, we acknowledge that this may have some bearing on the overall interpretation of results. Furthermore, these findings are based on a fairly small sample size. As a necessary next step, the reliability and validity of the test needs to be established using a larger sample of participants. This will ensure that the current findings are not an anomaly due to sample size.

This is an initially prototype system that we are currently using as an assessment tools to make a differential diagnosis of a virtual patient. In order to build a system like this to be effective for general clinical interviewing a minimal set of requirements should be considered. It's hard to state all that is needed without more subject testing and evaluations. But these requirements should address the technology, the learning objectives and the virtual patients. The technology needs to support face to face interaction, ideally using speech recognition. The system needs to respond in a natural amount of time for the character, response times that are too slow distract from the process. The system needs to allow the characters to respond and express themselves with verbal and non-verbal behavior. Understanding the learning objectives and how best to enable that in the system is valuable, however it can be quite a complex and difficult process. The technology needs to support those objectives. The characters need to be believable and responsive for the type of mental disorder they posses and designing the surface level verbal and nonverbal behavior is still an area of research.

Are embodied characters better then just speech or text interfaces? We believe that having the ability for a character to show verbal and non-verbal behavior is a powerful mechanism. Clinicians rely on certain nonverbal cues to understand the behavior of the patients and to make a diagnosis. Increasing the believability of these characters will draw the clinician into a closer engagement then just analyzing text.

Future work with the system should include: 1) addition of a camera for more user input into the system. This would enable the character to see the users' movement, e.g. if they are gazing at them. The more interaction the better; 2) addition of more personality to the character; this would allow training with different types of characters, for example an aggressive vs. a passive one. The kinds of questions the user asks may change based on some personality characteristics. The personality would be reflected in the verbal and non-verbal behavior; 3) maintain more conversation and discourse memory about what is being discussed in the interview to reduce brush off or incorrect responses; 4) addition of intonation, prosody, and affect to the speech output (e.g given that persons with Conduct Disorder are prone to anger, it may be appropriate to have him shout his responses when the same question is asked several times); 5) addition of tools to build characters that have several different behavior problems, the dialog they would use and the non-verbal behavior they would manifest; 6) Compare an interview of a real clinician and a real patient to a VP and see how the questions asked to the real patient respond when asked to the VP.

For interview training a series of small vignettes that guide one particular interviewing technique such as reflective listening and following would be useful. Additional enhancements to the system would include the building of an agent that tracks the dialog, affective behaviors, and other state information of interviewers. Such enhancements would also augment the VPs behaviors based on the system's capacity for representing various psychological problems.

Characters should be built that don't fit neatly into one specific category, as no one ever does. They should have the core criteria of a normal person but can deviate from that norm into one of the mental disorder categories. Ascribing this kind of verbal and non-verbal behavior to a fully embodied multimodal character will ultimately increase the believability, interactivity and effectiveness of the system. In most therapy session talking to the patient is the cure, a clinician that tries to understand the problem and offer up a solution will lead to a clinician that can enhance their skill set.

Acknowledgments. 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. We wish to thank Keck School of Medicine at USC and Cheryl StGeorge for helping with the test subjects.

References

1. Andrew, R., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Stevens, A., Bernard, T., Oxendine, C., Wagner, P., Lind, S.: Comparing Interpersonal Interactions with a Virtual Human to those with a Real Human. IEEE Transactions on Visualization and Computer Graphics (2006)
2. Babu, S., Schmugge, S., Barnes, T., Hodges, L.: What Would You Like to Talk About? An Evaluation of Social Conversations with a Virtual Receptionist. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 169–180. Springer, Heidelberg (2006)

3. Beutler, L.E., Harwood, T.M.: Virtual reality in psychotherapy training. Journal of Clinical Psychology 60, 317–330 (2004)
4. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H.: An Architecture for Embodied Conversational Characters. In: Proceedings of the First Workshop on Embodied Conversational Characters, Tahoe City, California, October 12-15 (1998)
5. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR 4th Ed, (Text Revision), American, Psychiatric Association (1994)
6. Dickerson, R., Johnsen, K., Raij, A., Lok, B., Hernandez, J., Stevens, A.: Evaluating a script-based approach for simulating patient-doctor interaction. In: Proceedings of the International Conference of Human-Computer Interface Advances for Modeling and Simulation (2005)
7. Evans, D., Hern, M., Uhlemann, M., Lvey, A.: Essential Interviewing: A Programmed Approach to Effective Communication, 3rd edn. Brooks/Cole Publishing Company (1989)
8. Frank, G., Guinn, C., Hubal, R., Pope, P., Stanford, M., Lamm-Weisel, D.: JUSTTALK: An application of responsive virtual human technology. In: Proceedings of the Interservice/Industry Training, Simulation and Education Conference, USA (2002)
9. Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E.: Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems (July/August 2002)
10. Johnsen, K., Dickerson, R., Raij, A., Harrison, C., Lok, B., Stevens, A., et al.: Evolving an immersive medical communication skills trainer. Presence: Teleoperators and Virtual Environments 15(1), 33–46 (2006)
11. Kiss, B., Szijarto, G., Benedek, B., Simon, L., Csukly, G., Takacs, B.: CyberTherapy: Applications of virtual reality and digital humans in clinical psychology. In: 2nd International Conference on Computer Animation Geometric Modeling, Hungary (2003)
12. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thorisson, K., Vilhjalmsson, H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA (August 21-23, 2006)
13. Lee, J., Marsella, S.: Nonverbal Behavior Generator for Embodied Conversational Agents. In: 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA (2006)
14. Leuski, A., Patel, R., Traum, D., Kennedy, B.: Building effective question answering characters. In: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, Sydney, Australia (2006)
15. Leuski, A., Pair, J., Traum, D., McNerney, P.J., Georgiou, P., Patel, R.: How to talk to a hologram. In: Edmonds, E., Riecken, D., Paris, C.L., Sidner, C.L. (eds.) Proceedings of the 11th international conference on Intelligent user interfaces(IUI'06), Sydney, Australia, pp. 360–362. ACM Press, New York (2006)
16. McCauley, L., D'Mello, S.: MIKI: A Speech Enabled Intelligent Kiosk. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 132–144. Springer, Heidelberg (2006)
17. McGee, J.B., Neill, J., Goldman, L., Casey, E.: Using multimedia virtual patients to enhance the clinical curriculum for medical students. Medinfo 9(2), 732–735 (1998)
18. Parsons, T.D., Bowerly, T., Buckwalter, J.G., Rizzo, A.A.: A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. Child Neuropsychology (2007)
19. Pellom, B.: Sonic: The University of Colorado continuous speech recognizer. Technical Report TR-CSLR-2001-01, University of Colorado, Boulder, CO (2001)

20. Prendinger, H., Ishizuka, M.: Life-Like Characters - Tools, Affective Functions, and Applications. Springer, Heidelberg (2004)
21. Rickel, J., Gratch, J., Hill, R., Marsella, S., Swartout, W.: Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences. In: AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, Stanford University, CA (2001)
22. Rizzo, A.A., Pair, J., Graap, K., Treskunov, A., Parsons, T.D.: User-Centered Design Driven Development of a VR Therapy Application for Iraq War Combat-Related Post Traumatic Stress Disorder. In: Proceedings of the 2006 International Conference on Disability, Virtual Reality and Associated Technology, pp. 113–122 (2006)
23. Rizzo, A., Bowerly, T., Buckwalter, J., Klimchuk, D., Mitura, R., Parsons, T.D.: A Virtual Reality Scenario for All Seasons: The Virtual Classroom. CNS Spectrums 11(1) (2006)
24. Patel, R., Leuski, A., Traum, D.: Dealing with out of domain questions in virtual characters. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, Springer, Heidelberg (2006)
25. Sethy, A., Georgiou, P., Narayanan, S.: Building topic specific language models from webdata using competitive models. In: Proceedings of EUROSPEECH, Lisbon, Portugal (2005)
26. Stevens, A., Hernandex, J., Johnsen, K., et al.: The use of virtual patients to teach medical students communication skills. The Association for Surgical Education Annual Meeting, April 7-10, New York, NY (2005)
27. Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel, J., Traum, D.: Toward Virtual Humans. AI Magazine 27(1) (2006)
28. Thiebaux, M., Marshall, A., Marsella, S., Fast, E., Hill, A., Kallmann, M., Kenny, P., Lee, J.: SmartBody: Behavior Realization for Embodied Conversational Agents. In: IVA07 (2007)
29. Triola, M., Feldman, H., Kalet, A.L., Zabar, S., Kachur, E.K., Gillespie, C., et al.: A randomized trial of teaching clinical skills using virtual and live standardized patients. Journal of General Internal Medicine 21(5), 424–429 (2006)