Human Computer Interaction in Virtual Standardized Patient Systems

Patrick G. Kenny, Thomas D. Parsons, and Albert A. Rizzo

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

Abstract. Interactive computer generated characters can be applied to the medical field as virtual patients for clinical training. The user interface for the virtual characters takes on the same appearance and behavior as a human. To assess if these virtual patients can be used to train skills such as interviewing and diagnosis they need to respond as a patient would. The primary goal of this study was to investigate if clinicians could elicit proper responses from questions relevant for an interview from a virtual patient. A secondary goal was to evaluate psychological variables such as openness and immersion on the question/response composites and the believability of the character as a patient.

Keywords: Virtual Patients, Artificial Intelligence, Clinical Psychology.

1 Introduction

Humans interact with objects of all types in the world on a daily bases. These objects can be as simple as an apple or as complex as a car or a computer. However, most interactions occur with other humans. These interactions can take the form of verbal behavior such as talking or non-verbal behavior such as gaze, gestures or body language. There are many factors that drive this behavior, such as, personality, emotion, mood and cognition, culture, gender, history and education. The task of describing all this behavior embodies a huge amount of work from neuroscience and psychology to cognitive science and artificial intelligence. The virtual human project at The Institute for Creative Technologies is tasked with developing and researching all aspects of this behavior and interaction through building integrated virtual human systems [1].

Virtual humans are embodied interactive agents that represent real humans in a virtual environment. These avatar characters take on human representations in their appearance, interaction and decision making and are used in many applications that require human-like interfaces, such as guides, trainers or medical. These human like qualities add to the complexities and constraints on the way users interact with the virtual characters. The integrated virtual human systems we develop make use of speech recognition, natural language understanding, verbal and non-verbal behavior generation, speech and language generation, reasoning, task modeling and appear in a virtual environment [2].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2009, LNCS 5613, pp. 514–523, 2009.

One of the main research questions of the virtual human work is in developing these characters so they appear and act like real humans without falling into the Uncanny Valley [3]. The Uncanny Valley was defined by Masahiro Mori in the 1970's to describe robots and characters that look like humans, but don't act like humans, or what we expect them to act like. As the realism of the appearance of the character approaches human appearance, for example the face, eyes or skin, while the actions and the behavior don't, or are off just a little bit, for example blank stares, or lips that don't move with the same muscle fidelity as humans, this causes an uncomfortable feeling amongst people interacting with them and destroys their believability. Our virtual human system, although they are realistic looking and acting, have not had any problems with falling into the uncanny valley.

The focus of the virtual patient is applying these virtual humans to the medical domain to create virtual standardized patients (VP) that can be used to teach interview, diagnosis, and social-interaction skills. The primary goal of this research is to assess the technology and system in eliciting correct question/response pairs from novice clinicians in a clinical interview. A secondary goal is to investigate the impact of psychological variables such as the subjects' state, current mood, and personality traits, openness to new experiences and immersion upon the resulting question/response composites and the overall believability of the characters.

Medical students currently perform interview training with human actors acting as standardized patients. The actors portray some clinical problem in what is called an Objective Structured Clinical Examination (OSCE) [4]. These tests typically take from 20-30 minutes, a faculty member watches the student perform. The evaluation consists of self assessment rating along with faculty assessment. Although schools commonly make use of standardized patients to teach interview skills, the diversity of the scenarios that standardized patients can characterize is limited by the availability of human actors and their skills at portraying the condition. Additionally the actors most likely vary their performance from subject to subject and location to location. This is an even greater problem when the actor needs to be an adolescent, elder or portray a difficult condition. Our process is similar to an OSCE, but the actor is replaced with a virtual patient and an observer is replaced by video recording. Using virtual patients will allow standard performance assessments for all subjects.

The virtual patient system was used in a series of subject testing experiments with novice clinicians and medical students. The role of the clinician was to ask appropriate and relevant questions to elicit correct responses from the virtual patient in a structured, yet free flowing, interview for history taking and diagnosis of the character. The results of the subject testing will be discussed as well as the human interaction issues with the system.

Enabling rich and engaging interaction with virtual characters in a medical setting will ultimately allow powerful experiential learning engagements for new and experienced students on a continual basis for practice with a variety of patient cases they may get little or no training with.

1.1 Virtual Human Interaction

Human to human interaction is very complex, enough so that many people in many fields devote a lot of time trying to understand it, and people find it so interesting that most movies produced are about human behavior and relationships. The interactions can vary based on numerous factors in both the person and the social setting. One on one interaction is different than multi-party interactions. Trying to re-create and model these interactions in virtual human systems is a great challenge. The assumptions and expectations that people have while interacting or engaged with other humans is brought over when people interact with the virtual humans [5], and these need to be replicated in the virtual character. For example users of the virtual human system expect to talk to the characters as they would with real humans, sometimes forgetting, or not knowing, the complexity and limitations of the underlying system software. Our VP system interaction is not based on the use of a traditional mouse and keyboard or pull down menus, but uses natural speech recognition. Speech interaction can cause confusion in the user if the character responds in a way that the user doesn't expect. The user interacting with a person builds a mental model through dialog and non-verbal behavior of that person, if that representation is violated, then the user may lose engagement or be confused.

The same confusion can happen with the virtual patient system and performance could suffer. For the medical domain we have some leeway as patients can, and usually do, act in non-traditional manners, thus if the patient responds in an off topic manner then this may not be thought of as incorrect. For our testing of the system we want to allow freedom of interaction and not constrain the user to specific bounds of what they can or can't say or how they should say it. We usually do not know what kind of questions the clinicians may ask the patient, which makes it hard to design the domain. Additionally, clinicians have varied training and there are multiple approaches to Interviewing and Diagnosis [6] that the system will ultimately have to take into account. However, this freedom of exchange is a good thing to capture for it will allow us to evaluate the technology and character interaction to provide methods for improving or automating the system for a more natural human computer interaction in the future.

2 Virtual Patient System

This virtual patient system consists of a computer generated 3D character that interacts with a human through natural speech. The character responds in kind through speech and gestures. Creating virtual patients that interact falls into two main areas; the technology and the domain. The technology needs to support what the character should do and how the user interacts with it, and in this case supporting the medical domain. In our virtual patients we have been concentrating on building characters with psychological problems in contrast to physical problems, they are more dialog based then motion or action based. One of the challenges of building characters for this domain is designing what needs to go into the patients in terms of the dialog, behaviors and actions to fit the patient profile.

2.1 Psychological Medical Domain

The role of the clinician during an initial meeting and engagement with a patient is to capture a history of the person, find out what is going on, and try to narrow down the problem in what is called a differential diagnosis by ruling out issues and problems not relevant to the case. The virtual patient interaction should mimic the real patient interview as much as possible with the goal of allowing the clinician to ask appropriate and relevant questions to elicit correct responses from the virtual patient in a structured, yet free flowing, interview for history taking and diagnosis of the character.

Fig. 1. Justina Virtual Patient

The virtual patient for this research is an adolescent female character with Post-Traumatic Stress Disorder (PTSD) called Justina, Figure 1. PTSD usually happens to people after some kind of traumatic event, such as a military engagement or assault and causes changes in behavior of the person. 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.

According to the most recent revision to the American Psychiatric Association's DSM Disorders, PTSD is divided into six major categories; refer to the DSM-IV category 309.81 [7] for a full description and subcategories;

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

One of the challenges of designing and building these interactive virtual standardized patients has been in enabling the characters to act and carry on a dialog like a real patient that has the specific mental condition for the domain of interest. Additional issues involve the breadth and depth of expertise required in the psychological domain to generate the relevant material for the virtual character.

The PTSD domain for the virtual patient allows exploration of the interaction with real end users while still being able to constrain the dialog and character behavior to a small and manageable corpus. The system needs to respond appropriately based on the six categories mentioned above and requires developing dialog for the virtual character to say around each of the categories. Table 1 is a set of example questions the clinician may ask and responses the virtual patient may say. Two additional categories were added for questions that involved building rapport or general questions, i.e. 'what's your name?' and for technical issues, like accidental button presses while thinking about a question to ask.

Table 1. Question / Response Categorization for PTSD

Category	User Question	Virtual 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?	Sometimes 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) Communication	Hi Justina, I'm Doctor…	Hello
(H) Other	'Button Press'	I don't get what you mean.

2.2 Technology

The virtual patient system consists of a set of distributed modules of which only a few are interfaces to the user. The user interacts through speech recognition to talk to the virtual character and through the 3D graphics that shows the character's animation in a virtual environment on a large monitor. The distributed set of components that make up the system form a pipeline that is the information flow from the input of the user to the output of the character. The main components can be divided into three areas:

User Input

- Speech Input – This component takes the user input from a microphone and translates that into a string of text to be used by the natural language system. The speech recognizer requires a speech and language model. Since everyone has a different voice, i.e. male, female, child, elderly, a different speech model is required and is changed for each user. In our case the users are male or female adults. A language model is required that defines the possible set of words in the domain that can be recognized. The corpus for the virtual patient consistes of 20K words.

Processing

- Natural Language – The text from the speech recognizer is sent to a statistical question/response module [8] that picks a response based on the input question. The question/response pairs are matched by hand prior to deployment. For the virtual patient PTSD domain there were 500 questions with 100 responses. If a question is asked that is not in the domain, an off topic response would be given such as; "I don't know" or "I would rather not talk about that". The questions and responses were acquired through expert knowledge, roleplaying and best guesses.
- Behavior Generation – After a response is selected then gesture animations are applied based on a set of rules that govern the non-verbal behavior [9]. Gestures can be hand movement, body posture, gaze or the like and is only limited by available animations.

Character Output

- Speech Generation – The speech output can be generated with an automated speech generation system or with pre-recorded voice overs that match the response text. For the virtual patient pre-recorded voice was used.
- Animation Output – The animations that were selected for the non-verbal behavior are combined with the generated speech to synchronize and play out together through a procedural animation system called Smartbody [10]. This drives the 3D character in a realistic fashion.

3 Subject Testing

Subject testing of the virtual patient system was conducted with medical students to evaluate the interaction, systems usefulness, effectiveness and usability as a medium to communicate with the students in performing the interview task. The evaluation consisted of an assessment of the system as a whole through questionnaires and data collection during the interview. The human computer interaction factor evaluation examined the technology underlying the speech recognition, dialog interaction and behavior of the character for this task with the user.

An important issue in the study of intelligent virtual agent interaction is to take into account the users openness to the interaction, new experiences and novel technologies. It has been suggested in a recent study [11] that physiological arousal appeared to be moderated by participant openness. High-absorption individuals may be more capable of imagining that the VP has PTSD when it is suggested. The users' openness will be compared to their performance to assess if they did better on the interview task.

3.1 Participants

Participants were asked to take part in a study where they would interact with an advanced prototype technical virtual patient system, in a similar matter to how they currently perform an OSCE. They were not told what kind of condition the VP had if any. Recruitment methods were by poster advertisements on the university medical campus, and classroom recruitment. 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 time with the time spent in the interview and questionnaires, which took approximately 45 minutes.

3.2 Method

The subject testing was divided into three phases, a pre-test and pre-questionnaire, the interview and a post-questionnaire. The pre-questionnaire was performed in a separate room from the interview and took about 10 minutes. For the interview the participants were asked to perform a 15 minute interaction with the VP and assess any history or initial diagnosis of a condition of the character. The participants were seated in front of a large monitor that had the virtual patient sitting on a couch in the therapists' room. The subjects used a head mounted microphone and were required to press the mouse button, talk, and then release the mouse button. The participants were asked to talk normally as they would to a standardized patient, but were informed that the system uses speech recognition and was a research prototype. They were free to ask any kind of question and the system would try to respond appropriately. At the end of the 15 minute exchange they would be sent to another room to take the post-questionnaire. Video, system logs and data from the various modules was logged as follows:

- First, the user speech was recorded from the automated speech recognition (ASR) engine, the speech before and after the engine processed it was captured. The before speech was later transcribed to compare with the processed speech.
- Second, the text from the natural language (NL) statistical question/response system was saved. The NL system records a transcript of the entire dialog session, this is used later to help analyze the question/response interaction.
- Third, system log files of the messages between the modules were captured and could be used to reconstruct what happened during the interaction.
- Fourth, cameras recorded participant's facial expressions and body language during interaction with the virtual patient system to be used for future studies.

3.3 Measures

The following measures were used to assess the impact of openness (absorption and immersiveness) upon the "believability" of the system. Prior to the experiment itself, the subjects were required to fill in the following questionnaires:

1. Tellegen Absorption Scale (TAS). The TAS standardized questionnaire aims to measure the subject's openness to absorbing and self-altering experiences. The TAS is a 34-item measure of absorption [12].
2. Immersive tendencies questionnaire (ITQ). The standardized ITQ measures individual differences in the tendencies of persons to experience "presence" in an immersive virtual environment (VE). The majority of the items relate to a person's involvement in common activities. While some items measure immersive tendencies directly, others assess respondents' current fitness or alertness, and others emphasize the user's ability to focus or redirect his or her attention. The ITQ is comprised of 18 items, and each is rated on a 7-point scale [13].

3. Virtual Patient Pre-Questionnaire (VPQ1). This unstandardized scale was developed to establish basic clinical competence for interaction with a person that is intended to be presented with PTSD, although no mention of PTSD is on the test.
4. Justina Pre-questionnaire (JPQ1). 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.

After the experiment the subjects filled in the following questionnaires:

1. Presence questionnaire (PQ). The Presence Questionnaire is a common measure of presence in immersive virtual reality. Presence has been described of as comprising three particular characteristics: sense of being within the VE; extent that the VE becomes the dominant reality for users; and extent to which users view the VE as a place they experienced rather than simply images they observed. The PQ is a widely used questionnaire [12].
2. Justina Post-questionnaire (JPQ2). We developed this unstandardized 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.
3. Virtual Patient Post-questionnaire (VPQ2). This scale was exactly the same as the Virtual Patient Pre-questionnaire and will be used in the future for norming of a pre-post assessment of learning across multiple interactions with the VP.

4 Results

The present focus is on effect sizes that describe the strength of association between question and response pair composites for a given PTSD category. An effect size of 0.20 was regarded as a small effect, 0.50 a moderate effect, and 0.80 a large effect.

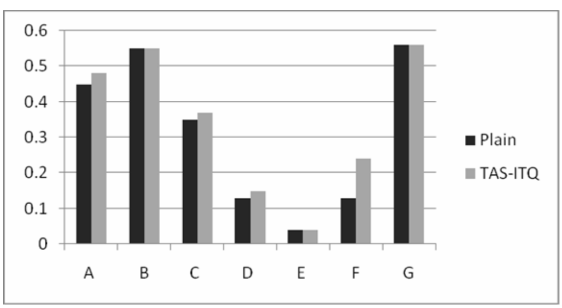

Fig. 2. Effect Size on Question / Response Composites

Figure 2 shows a chart of the data. Moderate effects existed for PTSD Category A,B,C,G, but only small effects were found for Category D and F. After controlling for the effects of the Tellegen Absorption Scale, increased effects were found for Category A, C, D, and F. To assess the impact of psychological characteristics such as absorption and immersiveness upon the "believability" of the VSP and student interaction we created a composite variable that included scores from the TAS and the ITQ. Strong effects existed between the ITQ and the PQ (0.78), and moderate effects existed between the ITQ and the VSP Post-questionnaire (0.40).

These results showed that the users were able to ask question and elicit responses in each of the PTSD categories. Additionally the findings suggest that the presence and openness appears to moderate user reaction and perform better on the task. Future studies should make use of physiological data correlated with measures of immersion to augment and quantify the effects of virtual human scenarios.

5 Conclusion and Future Work

Here we focused on effective interview skills—a core competency for psychiatry residents and developing psychotherapists. The keys aspects of the interview that we looked at were: interpersonal interaction; attention to the VP's vocal communications, as well as verbal and non-verbal behavior. Specifically, we wanted to assess whether the user (clinician in training) asked questions related to the reason for referral and received appropriate responses and also made attempts to gather information about the VP's problems. Finally, we wanted to see if the user would attempt detailed inquiry to gain specific and detailed information from the VP, separating relevant from irrelevant information. The primary goal in this study was evaluative and the Question/response composites were developed to reflect the shared variance existing between the responses of the VP and the users Questions that are necessary for differential diagnosis.

In future work we plan to compare the virtual patient system with live standard patient actors to assess if the technology constrains the communications or rapport between the clinician and patient. We will also compare a VP with PTSD against a VP that does not have PTSD to assess if a clinician can make a proper diagnosis. Additional improvements in the language and speech for the domain will allow for deeper and richer dialog. Building virtual characters is an iterative approach that improves the technology with feedback from real users to assess if these systems can be used as effecting teaching and training tools, which we believe is the case.

References

1. Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S.D.: Piepol Building Interactive Virtual Humans for Training Environments. In: Proceedings of I/ITSEC (November 2007)
2. Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E.: Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems, 54–63 (July/August 2002)

3. Mori, M.: On the Uncanny Valley. In: Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valley, Tsukuba, Japan, December 5 (2005)
4. Walters, K., Osborn, D., Raven, P.: The development, validity and reliability of a multimodality objective structure clinical examination in psychiatry. Medical Education 39, 292–298 (2005)
5. Gratch, J.N.W., Gerten, J., Fast, E., Duffy, R.: Creating Rapport with Virtual Agents. In: 7th International Conference on Intelligent Virtual Agents, Paris, France (2007)
6. Evans, D.H.M., Uhlemann, M., Lvey, A.: Essential Interviewing: A Programmed Approach to Effective Communication, 3rd edn. Brooks/Cole Publishing Company (1989)
7. DSM, American Psychiatric Association (DSM-IV-TR) Diagnostic and statistical manual of mental disorders, 4th edn., text revision. American Psychiatric Press, Inc., Washington (2000)
8. Leuski, A., Traum, D.: A Statistical Approach for Text Processing in Virtual Humans. In: Army Science Conference (2008)
9. Lee, J., Marsella, S.: Nonverbal Behavior Generator for Embodied Conversational Agents. In: 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA (2006)
10. Thiebaux, M., Marshall, A., et al.: SmartBody: Behavior Realization for Embodied Conversational Agents. In: International Conference on Autonomous Agents and Multi-Agent Systems, Portugul (2008)
11. Macedonio, M., Parsons, T.D., Rizzo, A.A.: Immersiveness and Physiological Arousal within Panoramic Video-based Virtual Reality. Cyberpsychology and Behavior 10, 508–516 (2007)
12. Tellegen, A., Atkinson, G.: Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. Journal of Abnormal Psychology 83, 268–277 (1974)
13. Witmer, B., Singer, M.: Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoperators and Virtual Environments 7(3), 225–240 (1998)