Conversational Agents and Natural Language Interaction: Techniques and Effective Practices

Diana Perez–Marin
Universidad Rey Juan Carlos, Spain

Ismael Pascual–Nieto
Universidad Autónoma de Madrid, Spain
Chapter 11

Embodied Conversational Virtual Patients

Patrick G. Kenny
University of Southern California, USA

Thomas D. Parsons
University of Southern California, USA

ABSTRACT

Recent research has established the potential for computer generated virtual characters to act as virtual patients (VP) for the assessment and training of novice clinicians in interpersonal skills, interviewing, and diagnosis. These VPs are embodied interactive conversational agents who are designed to simulate a particular clinical presentation of a patient’s illness with a high degree of consistency and realism. In this chapter we describe the architecture developed for virtual patients, and the application of the system to subject testing with virtual patients that exhibit a set of clinical conditions called Post Traumatic Stress Disorder (PTSD). The primary goal of these conversational agents was evaluative: can a VP generate responses that elicit user questions relevant for PTSD categorization? The results of the interactions of clinical students with the VP will be discussed. This chapter also highlights a set of design goals for increasing the visual, physical and cognitive realism when building VP systems including the design of the language, scenarios and artwork that is important when developing these characters. Finally, future research directions and challenges will be discussed for conversational virtual patients.

1. INTRODUCTION

The development of the Eliza program by Joseph Weizenbaum (1966) which was capable of engaging humans in a natural conversation and simulated a Carl Rogers empathic psychologist was one of the first conversational agents with a medical theme. Although simple in design and driven by a script called DOCTOR that performed keyword matching and replacement, it was a very powerful mechanism that tricked some people into thinking they were talking to a real psychoanalyst. Today these kinds of conversational programs are more complex with fully embodied characters that exhibit facial expressions, gestures, animation and

DOI: 10.4018/978-1-60960-617-6.ch011
speech. However one of the more challenging aspects that still remain is the conversational dialog.

A general model for conversational agents should account for the intentional and non-intentional aspects of verbal and nonverbal communication as well as the contextually grounded natural biologically based aspects of conversation (Buck, 1994). A further refinement of this understanding may view conversation as something that occurs whenever one agent’s (i.e. sender’s) behavior influences the behavior of another agent (i.e., receiver).

Our work includes a general theoretical commitment to understanding virtual human conversation as a feedback process, in which interpretation of verbal and nonverbal data (i.e., message reception and production) alone, although potentially useful, are deficient. To make up for the incompleteness of limiting virtual human conversation and to provide more situational context to these two criteria we add psychophysiological information (e.g. heart rate) from the user into the conversational processing as we believe it reflects the conversation found in human agents.

While a full elucidation of this interactionist theory is beyond the scope of this chapter (Buck, 1984, 1989, 1994) we do mention it as the theoretical underpinning of much of what we are trying to do in our work toward an integrated view of virtual and human conversation. Further, our data analytic approach to understanding the utility of virtual human conversation requires that the assessment of the communicative efficacy of the conversational agent’s behavior involves the extent to which it reduces uncertainty in the behavior of another (Wilson, 1979). Herein we discuss the development of our virtual patients (VP) and the general assessments that we make when assessing the virtual and human conversation. For our work this involves the integration of verbal communication, nonverbal communication, and contextual factors (e.g., psychophysiological data).

This chapter provides an examination into applying embodied conversational virtual patients for medical simulation and training. These VP’s are interactive characters designed to simulate a particular clinical presentation of a patient with a medical illness with a high degree of consistency and realism. This chapter describes the virtual patient architecture used for research and evaluation, the subject testing conducted with VP’s that exhibit a clinical condition of Conduct Disorder, Post Traumatic Stress Disorder (PTSD) or assessing racial bias. The development of the characters and dialog for the scenarios will also be addressed. The primary goal of the subject testing was evaluative: can a VP generate responses that elicit user questions relevant for PTSD categorization? The results of the interactions of clinical students with the virtual patient will be discussed along with areas of further research.

2. NEED FOR CLINICIAN TRAINING IN CONVERSATION SKILLS

Developing good conversational skills is essential for clinicians to establish good doctor patient relationships. Many undergraduate and postgraduate medical education and training programs have begun to place greater emphasis on the importance of high-quality conversation skills (ACGME; 2007). Traditional approaches to training clinicians in the conversation skills needed for assessment, diagnosis, and interview performance rely upon a combination of classroom learning and role-playing with human standardized patients.

The importance of conversation is reflected in recent requirements for communication evaluation in medical schools. The Accreditation Council for Graduate Medical Education (ACGME; 2007) has emphasized the importance of interpersonal and communication skills in training clinicians. Residents are expected to: 1) create and sustain a therapeutic and ethically sound relationship with the patient; 2) use effective listening skills, eliciting...
and providing information using effective nonverbal, explanatory, questioning, and writing skills; and 3) work in an efficient manner with others.

However, evaluation studies have revealed methodological deficiencies in many cases (Chant et al., 2002) and limited positive training effects (Hulsman et al., 1999). In an effort to increase interpersonal communication assessment, standardized patients (paid human actors) have been recruited and trained to exhibit the characteristics of an actual patient, thereby affording novice clinicians a realistic opportunity to practice and be evaluated in a mock clinical environment.

Although a valuable training approach, there are limitations with the use of human standardized patients that can be mitigated through simulation technology. For example, human standardized patients are expensive and cost several thousand dollars per student. Further, given the fact that there are only a handful of sites (for over 130 medical schools in the U.S.) providing standardized patient assessments of the clinician in training’s communication ability as part of the U.S. Medical Licensing Examination (USMLE), the current model provides limited availability.

Another concern is the issue of standardization. Despite the expense of standardized patient programs, the standardized patients themselves are typically unskilled actors. As a result of common turnover, administrators face considerable challenges for offering psychometrically reliable and valid interactions with the training clinicians. A related issue is the limited scope that the actors are able to portray. As a result, there tends to be an inadequate array of developmentally, socially, racially and culturally diverse appropriate scenarios.

For example, when a clinician has a pediatric focus and needs access to children, it is difficult for the clinician to pretend that the actor is a child. Finally, many clinical cases (e.g., traumatic brain injury) have associated physical symptoms and behaviors (e.g., dilated pupils, spasms, and uncoordinated movements) that simply cannot be accurately portrayed by human actors.

Additionally, a large part of working with clients is re-visits by the clinician to see how they are progressing. This is rarely performed with the standardized patient actors due to the inconsistency in the population. Virtual patients are not meant to replace actors, but rather augment them with 24/7 availability and more standardized training. An added bonus is the ability to capture multiple forms of data from the participant, for example, the conversation log, speech, body language, facial expressions, gaze and interactions with the patient. This data can be used to evaluate the clinician and build up a more rigorous standardized way of assessing their performance with others that have also used the system from around the country and around the world.

2.1 Conversational Characters for Medical Applications

Our response to the difficulties inherent in training clinicians with standardized patients is to use virtual humans as patients. Virtual humans (VH) are developing into powerful interfaces that can enable greatly increased intuitive human like interactions. These virtual human systems consist of characters that have realistic appearances, can think and act like humans, and can express themselves both verbally and non-verbally. Additionally, these virtual humans can listen and understand natural language and track user interactions with speech, vision and biometrics systems. Advances in simulated virtual humans afford the possibility of virtual patients that reduce cost, ensure standardization and faithfully model physical symptoms.

To address the need to teach proper conversational skills requires developing agents that act as patients or clients and can carry out realistic and relevant dialog with a practicing clinician. These conversations are constrained by the mental or physical illness found in the patient. Most questions asked by a clinician will be about the patient’s condition, their behavior, symptoms or
history; infrequently it will be about general issues like the weather or sports which are more aimed towards relationship building or rapport.

However, proper rapport is crucial towards a better relationship with the client. Dialogs can last 15 minutes in an initial visit that is aimed at medical history gathering to more detailed conversations that can last weeks or months to discuss problems in a psychological assessment. The dialog can also change over time and issues that were discussed one week should be followed up the next week to assess progress. While the standard interview of a patient takes a form of introduction, discussion of symptoms, assessment and treatment, the dialog that manifests itself in this interview can vary wildly by the type of problems and symptoms the patient is exhibiting.

One thing is certain, humans use many forms of answering the same question, which can vary by many factors such as; personality, emotion or mood. Clinicians perform a method called a differential diagnosis on the patient by asking questions to rule out certain issues and digging down to a set of criteria for the disorder. Fully embodied conversational characters are important for these medical applications as clinicians and physicians need to observe the actions and expressions of the client along with listening to the speech. Studies have shown that this helps to immerse the clinician into the setting and increases the believability of the interaction as a whole by making it more real and natural (Kenny et al., 2007).

Virtual patients are artificially intelligent virtual human agents that control computer generated avatar bodies and can interact with users through speech and gesture in virtual environments (Gratch et al., 2002). Advanced virtual humans are able to engage in rich conversations (Traum et al., 2008), recognize nonverbal cues (Morency & de Kok, 2008), analyze social and emotional factors (Gratch & Marsella, 2004) and synthesize human conversation and nonverbal expressions which are synchronized together to produce realistic actions, facial expressions, lip syncing and gaze (Thieaux et al., 2008).

Additionally, they can contain underlying physiological models that simulate blood pressure, heart rate, breathing, and blushing (De Melo & Gratch, 2009). Building virtual humans requires fundamental advances in AI, speech recognition, natural language understanding and generation, dialog management, cognitive modeling and reasoning, virtual human architectures and computer graphics and animations. All these technologies need to be integrated together into a single system that can work in unison, be expandable, flexible and plug-and-play with different components.

These VPs will need to be able to exhibit certain behaviors through dialog and physical actions. The focus of this research is to build fully embodied conversational characters capable of exhibiting the full range of human dynamics and behaviors. These characters are referred to as high fidelity characters. The higher the fidelity the closer to human behavior the avatar appears. An example of a lower fidelity agent is one that uses a text interface and does not contain a body. A character with high fidelity input would contain sensors for speech to recognize clinician’s voice, for vision to recognize faces or track gestures or postures and biometric readings to record the users heart rate or skin conductance. High fidelity output includes finer grain facial expressions, variability in speech such as intonation or emotion and more expressive body language. The types of conversational agents that we research, design and build as VPs are considered to be high fidelity. High fidelity characters are defined along several dimensions to include these properties:

- **Visual Realism:** The characters should be as realistic as possible in appearance, animation, clothes and other visual aspects.
- **Mental and Physical Behavior Realism:** The characters should contain models that exhibit the proper illness to be displayed, this will in turn drive the conversation. The
mental processes should interact with the physical processes as there is a deep connection between the two.

- **Autonomous Behavior:** The characters need to be able to express themselves and exhibit their own behavior based on underlying models, and attributes. They should not just wait for the user to ask questions and generate responses.

- **Personality, Mood and Emotion:** The characters should contain models of personality, mood and emotion and they should assist in selecting the appropriate conversational style, tone and nonverbal behavior.

- **Social Interaction:** The characters exist in social settings and should have realistic behavior that emulates interpersonal skills. For example, turn taking in conversations, proper gaze and body language.

- **Multi-Modal Input and Output:** The characters should support many forms of input and output, for example, cameras, speech, gestures, biophysiology and sound that should be used by the underlying models when generating the output behavior and conversational dialog.

- **Story Management:** There should be a proper story behind the dialog and character, this will add realism, engage the user more and can help drive or constrain the conversation.

2.2 Related Work in Virtual Patients

Virtual patients fulfill the role of standardized patients by simulating a particular clinical presentation with a high degree of consistency and realism and offer a promising alternative to standardized patients (Deladisma et al., 2008; Green et al., 2004; Hayes-Roth et al., 2004, 2009; Hubal et al., 2000; Kenny et al, 2007; Lok et al., 2006; Parsons et al., 2008; Stevens et al., 2005).

There is a growing field of research that applies VPs to training and assessment of bioethics, basic patient communication, interactive conversations, history taking, and clinical assessments (Bickmore & Giorgino, 2006; Bickmore et al., 2007; Lok et al., 2006; Parsons et al., 2008; Johnsen et al., 2009). Results suggest that VPs can provide valid and reliable representations of live patients (Kenny et al., 2007; Triola & Feldman, 2006; Andrew et al., 2007; Raij et al., 2007).

Additionally VPs enable a precise presentation and control of dynamic perceptual stimuli that increases ecological validity. Hence, VPs offer the veridical control and rigor of laboratory measures and a verisimilitude that reflects real life situations (Parsons et al., 2009a; Andrew et al., 2007). In addition, some groups have developed complex cognitive models of patients that more deeply simulate a VPs decision making process for a person with gastroesophageal reflux disease (Nirenburg et al., 2009).

2.3 Virtual Patients Issues

Virtual patients are a growing area of interest for research and development and application to increase and develop users’ interpersonal and diagnostic skills in the medical field. Since it is such a young field there are no common definitions in what constitutes a VP, how they should be used or the best architecture for them. There are no standards in the underlying models, physical or psychological, and how that information should be used, or what attributes or qualities are important for VP characters. There is an enormous amount of research being conducted in the biological fields, neuroscience and brain studies in understanding the cognitive thought processes and how to build models of them (Edelman, 2006). Applying this work to build fully embodied virtual characters and complete architectures that can support them has not been attempted in detail, although it has always been a major goal in the field of Artificial Intelligence.
Most groups focus on creating text based VP systems such as (CASUS, 2010) and CAMPUS (Rudericha et al., 2004) some with very detailed cognitive models (Nirenburg et al., 2009) such as the Maryland Virtual Patients. There are some groups that create systems with 2D like characters, (Bickmore & Giorgino, 2006; Bickmore et al., 2007) and some with 3D characters, (Lok, 2006; Kenny et al., 2007a).

The European electronic virtual patients project (eVIP 2010) is attempting to create a set of standardized clinical cases that can be used by all, however integration of these cases into more complex 3D virtual characters and architectures has not been undertaken. These electronic cases mainly concentrate on the dialog design and not with other attributes like emotion, mood and non-verbal behavior. There is also a great deal of emphasis on creating VPs with physical related illnesses such as stomach problems, eye problems, trauma or military related injuries (Johnston & Whatley, 2005).

The work of our Virtual Patient Simulation Lab focuses is on creating VP’s with psychological illnesses such as suicide, depression, post traumatic stress, Alzheimer’s and other related disorders. The work here lays out a solution and a modular architecture that supports building complex embodied conversational virtual characters that can be used for medical simulation, interpersonal skills training and diagnosis.

These VPs rely heavily on being dialog driven so that a clinician can have an in-depth conversation with the patient, however they are fully embodied characters that synchronize the dialog with the non-verbal behavior and expressions. The work of the Neuroscience and Simulation Lab enables us to understand and add to the VP architecture with supporting cognitive functions and through assessing how users interact with the system with physiological input and response patterns associated from the users that is feed back to the VP to develop more adaptive VPs to augment the dialog and interaction.

3. VIRTUAL PATIENT SYSTEM ARCHITECTURE

Virtual training environments with VP characters are complex systems to construct with many disparate technologies that need to be integrated together to provide the required functionality. The various technologies allow for everything that a person would expect to encounter while interacting with a real human in a real environment, from the dialog and body expressions to real world object interactions like a stethoscope or digital thermometer. The architecture design should be able to support the multiple dimensions of fidelity addressed above.

3.1 Virtual Patient Technology

The VP system and architecture we use for research, development, subject testing and evaluation consists of many distributed modules that communicate through message passing. The underlying architecture used to drive the characters can be seen in Figure 1.

While not every module is used in our virtual patient system, the architecture supports building virtual humans of various complexities and has been used to do so (Kenny et al., 2007b). Building large systems like these to drive virtual humans are vast software engineering projects that are hard to manage and develop without proper practices and separating the functionality into distributed modules makes this task easier.

Separate modules allows for easier testing, as this can be performed on individual modules before whole system testing, the modules are easier to upgrade individually and new modules can replace older ones or be added into the architecture easily to add additional functionality. As this current system is a research platform, separating out the functionality into modules allows different researchers to develop components independently without having to use the whole system.
Embodied Conversational Virtual Patients

For example, the architecture supports using various Natural Language Understanding Modules. An important requirement for this system is message passing of data and communication between the modules; information should not just be passed in one direction, i.e., from the brain to the body, there should be feedback from lower level modules to higher ones. For example, interruption of speech or gestures if the clinician talks while the patient is responding or if the patient’s breathing increases then it should cause the heart rate to increase which could cause changes in the selected dialog. The mind should be able to know what the body is doing and vice versa.

The flow of information through the system can be seen in Figure 2 and involves three phases; Input, Reasoning and Output. Data is gathered at each phase and is used to assess the system and evaluate a user’s interaction with the character. Each of the modules in the phases requires domain data that is built beforehand. The Interaction works as follows:

3.1.1 Input Phase Modules

In the Input phase a user talks into a microphone in plain English (or language). The speech recognition engine records the audio signal which is converted to text string that is passed to a reasoning engine.

Speech Input – For Speech input the SONIC speech recognition engine from the University of Colorado, Boulder (Pellom, 2001) is used. The engine’s acoustic and language models are customized for the domain of interest (Sethy et al., 2005). The domain for the virtual patient consists of a corpus 10,000 words of general and specific
medical terminology. The engine requires no training of a user voice and is a push to talk system.

3.1.2 Reasoning Phase Modules

The reasoning phase involves making decisions based on the input data and internal state to generate action states and dialog for the character. The text from the speech engine is used by the natural language understanding system to select an appropriate response text string. This response text is further parsed to select appropriate non-verbal behavior and animations which is then sent to the output phase.

Natural Language Understanding System (NLU) – The NLU contains a corpus of questions and answers. The NLU parses the text string produced by the speech recognition module and uses a statistical text classifier to select the best question that matches the input string (Leuski et al., 2006). The question has an associated answer which is sent to the Non-verbal behavior module. Since the dialog plays a critical role in the characters interaction, the next section will describe this in more detail.

Non-Verbal Behavior Generator (NVBG) – Gestures and postures play a key role in realizing expressive, interpretable behavior in general and communicative intent specifically. For example, they qualify information in the speech such as a shrug when saying, “I don’t know.” They also emphasize important words by using, for example, a beat gesture (e.g., short chopping movement) synchronized with the word. The timing of gestures to the speech is critical, and small changes can alter an observer’s interpretation of the utterance of the speaker. Without gestures, or with poorly timed gestures, a character will look unnatural. The NVBG (Lee & Marsella, 2006) applies rules based on theoretical foundations of movement space to select the appropriate gesture animations, postures, facial expressions, and lip synch timing for the virtual character. Once the NVBG selects the appropriate behavior for the text, it then packages this up into a Behavioral Markup Language (BML) (Kopp et al., 2006) structure and sends it to the procedural animation system, SmartBody in the output phase.
3.1.3 Output Phase Modules

The output phase involves the character generating the appropriate gestures, animations and behaviors. A procedural animation system that synchronizes the gestures, speech and lip syncing and plays a pre-recorded or generated voice of the text for the character for final output to the screen is accomplished in this phase. The user then listens to the response and interacts more with the character.

Procedural Animation System – This system is called SmartBody and takes as input a 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 (Thiebaux et al, 2008). It is capable of using either generated speech or pre-recorded speech. Smartbody is also capable of using controllers that perform specific actions based on rules or timing information, such as head nods. The controllers are seamlessly blended in with the animations specified in the input message. A Motex, which is a looping animation file, can be played for the character to give it a bit of sway, finger tapping, or some repetitive movement.

Speech Generation- The text to speech voice generation is performed by a commercial product called cerevoice, additionally pre-recorded voice has been used for some of the characters. The language models for cerevoice are generated from an actor that speaks a large set of utterances, machine learning techniques are then used to create a dynamic model that is capable of creating any utterance. The output text string is converted into an audio file by the speech synthesis software and is played and synchronized by the Smartbody system when the character speaks. An alternative approach is to use a set of pre-recorded voice files that can be used by the character when it speaks, however this limits the set of utterances and is not dynamic, the advantage is that the voice actor can add more emotion and realism to the recorded voice.

Graphical Game Engine – The game engine is used to visualize the embodied conversational characters, animations of those characters, the environments and objects in the environment, such as chairs, tables, sounds, lighting, and medical devices. This is a commercial game engine, (GameBryo, 2010), with many of the advanced features in computer graphics such as shaders, bump mapping and environmental lighting that enhance the visual realism of the setting and characters.

3.2 Dialog and Discourse management for virtual patients

This section describes the Dialog system in more details as is an important part of interacting with VP characters since a clinical interview relies heavily on the types of questions asked and the responses from the patient. The current VP system uses a classifier system called the NPCEditor for Non-Player-Character Editor, a term used in computer games to represent characters in the game that are not controlled by another human player (Leuski et al., 2006; Leuski & Traum, 2010).

The NPCEditor is a user friendly software package for editing and adding to the corpus of questions and responses, it allows a developer to test the text classifier. The classifier is trained on the set of questions, responses and links between them, and will generate a threshold value that is used for the matching and classification process. This response selection process is based on a statistical text classification approach (Leuski & Traum, 2010). There is no limit to the number of responses or sample questions, but it is advisable to have at least two or three sample questions for each response. The system allows for several response categories described below. Sometimes the system combines the text from different categories to produce the final response.
The category types are as follows:

• **On-Topic**: 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/her questions, s/he may get a different response from the system.

• **Off-Topic**: 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.

• **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**: 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.

The VP question and response system provides a powerful mechanism to quickly develop domains and build the dialog for the scenarios by experts who are not familiar with the underlying technology. The advantages are that it works quite well even with long utterances and misrecognized speech input. The disadvantages are that the characters do not respond well to follow on questions, to general questions such as “can you tell me more” referring to the context of the last question or response, and the system does not have initiative to ask questions by itself.

Other approaches to natural language processing involve building goal directed agents with dialog managers (Gratch et al., 2002) or cognitive models (Nirenburg et al., 2009). These are effective, but can be quite complex to build. The focus of this current virtual patient research was to build up a corpus of the kinds of questions and interactions that a clinician would have with the system and character and gather data with real clinicians or student clinicians in a bottom up approach before moving on to more complex systems.
3.3 Psychophysiology to Enhance Conversational Aspects of Virtual Humans

Cognitive and affective models can enhance conversational aspects of virtual humans in that they enhance human computer interactions through incorporation of the emotional state of the user (Picard, 1997; Lisetti and Schiano, 2000). Current cognitive and affective models found in virtual human research tend to use appraisal models that specify how events, agents and objects are used to elicit an emotional response depending on a set of parameters (e.g., goals, standards and attitudes) representing the subject (Gratch and Marsella, 2001, 2003, 2004).

The appraisal theories that are used draw upon the work of Smith and Lazarus’s (1990) cognitive-motivational-emotive theory. As such, they view affect as something that arises from appraisal and coping. By “appraisal”, virtual human researchers mean the methods by which persons assess their overall (i.e., events leading to the then current state, the current state itself, and future prospects) relationship with the environment. By “coping”, virtual human researchers are referring to that which determines how persons respond to the appraised significance of events. Said researchers argue that best practices in the design of symbolic systems includes appraisal theories of affect that emphasize the cognitive and symbolic influences of affect and the underlying processes that lead to this influence (see Lazarus, 1991). For these researchers, this is a practice that is to be understood as in contrast to models that emphasize lower-level processes such as drives and physiological effects (Velásquez, 1998).

In principle, it is possible to model appraisal processes for conversational aspects found in virtual human research that follow outdated appraisal models that assert specific patterns of physiological changes that may be observed in affect occurrence after the subjective experience of affect within a conversation. Research in psychophysiology has not supported these cognition first models (Cox & Harrison, 2008). In fact, a common frustration to attempts at developing an adequate scientific approach to emotion has been to focus on constructing theories of subjective appraisals. Again studies of the neural basis of emotion and emotional learning have instead focused on how the brain detects and evaluates
Embodyed Conversational Virtual Patients

emotional stimuli and how, on the basis of such evaluations, emotional responses are produced (Magnenat-Thalmann & Thalmann, 2005).

Our preferred approach to enhancing conversational aspects of virtual humans is to develop cognitive and emotional virtual human models that include psychophysiological inputs that are sent in real-time interactions from the user (e.g., heart rate, respiration, skin conductance) to the virtual human. The additional input of the user’s psychophysiological states offers contextual information to cognitive and emotional models. It is believed that these additional inputs can be developed into affect-sensitive VP interfaces that go beyond conventional virtual human models designed by pure (i.e., devoid of psychophysiological metrics) cognitive appraisal principles.

The resulting affect-sensitive VP interfaces would be similar to brain-based-devices (BBDs) that are being designed based on biological principles and are programmed to alter their behavior to the environment through self-learning (Edelman, 2006). An example of such research is found in work to develop intelligent robots. A series of devices with sensors and computer-simulated brains have been built in Gerald Edelman’s (2006) Neurosciences Institute in La Jolla. The brains are modeled on human anatomy, complete with versions of visual cortex, inferotemporal cortex, and hippocampus. They are not pre-programmed, but evolve neuronal connections in response to experience. These devices can learn to recognize patterns and navigate novel environments.

Although the development of such computational models of emotion for conversational aspects of virtual humans can be difficult, researchers (Magnenat-Thalmann & Thalmann, 2005) have pointed out that computational approaches to emotional processing are both possible and practical. It is important to note, however, that there is a tendency of virtual human researchers to rely upon modalities such as facial expression, vocal intonation, gestures, and postures.

Unfortunately, this tendency results in limitations due to “communicative impairments” (both nonverbal and verbal) inherent in the technology. This is very much the case regarding expression of affective states. Although these vulnerabilities place limits on traditional conversational and observational methodologies found in much virtual human research, psychophysiological signals are continuously available; and 2) are arguably not directly impacted by these difficulties. As a result, psychophysiological metrics may proffer an approach for gathering robust data despite potential virtual human technology limitations.

Furthermore, there is evidence that psychophysiological activity of persons immersed in virtual environments is associated with 1) trait differences (immersability; Macedonio, Parsons, & Rizzo, 2007) and 2) state differences (intensity of the environment; Parsons et al., 2009a, 2009b). These findings from virtual reality research reflect the finding that transition from one affective state to another is accompanied by dynamic shifts in indicators of autonomic nervous system activity (Bradley, 2000).

Individual (Parsons et al., 2009a) and cohort (Parsons et al., 2009c) differences have been shown to impact results gleaned from psychophysiological assessments using virtual environments, which reflects the need for psychophysiological assessment of persons interacting with VPs. In addition to extending the validation of VPs, this author’s lab uses psychophysiological metrics to develop a psychophysiological interface (Neuroscience and Simulation Interface; NSI) for VPs that can adapt VP scenarios to the user’s psychophysiological processing of information.

More specifically, psychophysiological measures such as heart rate, skin conductance responses, facial electromyographic response recordings, respiration, electroencephalographic recordings, and eyetracking can be continuously recorded while subjects interact with the VP. These recordings can be processed in real time to
gain information about the user’s current state of emotional and cognitive processing. This information can then be relayed to the virtual human in order to change for example, the behavior of the VP. If the user is distressed by the current state of the interaction, a psychophysiological pattern of increased heart rate, increased skin conductance levels, a more rapid rate of respiration, increased corrugator muscle activity, decreased alpha wave activity and diversion of gaze may develop. Allowing for access to this information about the user’s current emotional state offers the VP an increased capability to understand the dynamics of the current interaction and develop a context appropriate response or behavior change.

4. DESIGN AND EVALUATION OF PATHOLOGIES

One of the challenges of building complex interactive VPs that can act as simulated patients has been in developing the specific mental condition in the domain of interest as it requires breadth and depth of expertise in the psychological domain to generate the relevant material for the character and dialog. This section discusses the domain and scenarios created for the subject testing and evaluation of the characters with real clinicians talking to the automated system. The results are described in the next section.

4.1 Domain Building

The domain consists of the story, the dialog, characters and settings and the illness and diagnosis criteria for the medical application. This data is generated through a combination of methods; discussions with subject matter experts, role playing exercises, user testing and textbooks or manuals. For the psychological medical illness that we want to exhibit we refer to the Diagnostic and Statistical Manual on Mental Disorders (DSM IV-TR, 2000). This manual is used by all professional clinicians and lists the criteria for a proper diagnosis for a particular mental disorder.

Building the domain involves developing the proper set of dialog and words that will be uttered by the character for the scenario. It is important to capture the kinds of questions that a clinician will typically ask during an interview or diagnosis session. The questions can vary from introductions and rapport to ones that elicit specific kinds of information. This is initially done through using experts to role play out the scenario, where one expert plays the patient and the other plays the clinician.

The role play scenarios are then transcribed and the set of questions and responses are put into a corpus database. This is an iterative process that will capture about 80 to 90 percent of the dialog. It is still a challenge to predict all the questions that a clinician will ask of a patient. Another method to capture additional questions and to assess the proper responses for questions is to use a Wizard of Oz method (Kelley, 1983), this is where a subject user talks to the virtual character, but the character is controlled by a human selecting appropriate responses from an interface. Similar methods have been used effectively to gather and evaluate spoken user data in tutoring systems (Litman & Silliman, 2004; Litman & Forbes-Riley, 2010).

This corpus is used to customize the engine’s acoustic and language models for the domain of interest. 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 domain corpus and to get word error rates to compare the input speech with the recognized speech from the speech engine.

4.2 Artwork and Character Design

The artwork plays a crucial role in defining the characters behavior, attitude and condition. People
Embodied Conversational Virtual Patients

Figure 3. Justin-VP

are able to make a judgment about someone within the first few seconds. The VP project involved the development of several characters. For one scenario we developed “Justin”, (see Figure 3) 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.

Another scenario was the development of “Justina” an adolescent female patient, (see Figure 4). This character was developed with more thought and time and the goals are to increase the realism with successive iterations. The artwork makes use of game technological advances such as shaders that apply graphical properties to the characters to make them look more realistic, like the skin, wrinkles or blushing to increasing the number of polygons and textures to make them look less blocky.

The project also involved the development of a clinical virtual environment that was modeled after a typical clinician’s office and was meant to represent a place that would make the patient feel at home. Another important aspect is the animations that the character will use. Since these characters are in seated positions, then most of the gestures will involve hand movement or upper body alterations, head movement and facial expressions. A library is built that is used by the non-verbal generator to apply the appropriate animations for the spoken text and state of the character.
4.3 Virtual Patient: Conduct Disorder

The first scenario involved the construction of a natural language-capable VP named “Justin”. The clinical attributes of Justin were developed to emulate a conduct disorder profile as found in the Diagnostic and Statistical Manual of Mental Disorders. Justin (see Figure 3) portrays a 16-year old male with conduct disorder who is being forced to participate in therapy by his family.

Justin’s history is significant for 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. Our goal was to obtain objective data from an initial intake interview by the clinician. An intake interview is where the clinician asks more general questions and tries to get an overview of what the problems are and make an initial diagnosis, not trying to apply therapy or ‘cure’ the patient.

The trainee’s interview questions were guided by the need to determine if the patient is internalizing or externalizing behaviors and for eliciting information regarding the four general symptom categories prevalent in conduct disorder:

- **Aggressive behavior**: fighting, bullying, being cruel to others or animals.
- **Destructive behavior**: arson, vandalism.
- **Deceitful behavior**: repeated lying, shoplifting, breaking into homes or cars.
- **Violation of rules**: 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. Some responses by the VP may be on target, off target, involve “brush off” responses, and in some cases, they may be irrelevant replies.

The probability of a specific response being emitted is rated to the question asked. 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 the test will require responses in all of the categories. The total set of questions and responses are extracted from role playing exercises, initial subject testing, interviews with doctors and common sense for specific responses. In total a response set would consist of over 800-1200 lines of text and 200-300 questions. The matching of user questions to an appropriate patient response in the NPCEditor Dialog tool is a manual process performed by a domain designer and can be done over a period of a few days, but requires a few weeks to perfect with lots of testing as there can be many ways to ask the same question that will generate a single response.

4.4 Virtual Patient: Trauma Exposure

For the next VP scenario, our lab constructed an adolescent female character called “Justina” that had been the victim of an assault and showing signs of PTSD (see Figure 4). The technology used for this character and the “Justin” character is the same underlying VP architecture. The changes in this scenario were in the speech recognizer and natural language understanding module for the new scenarios to include a more in-depth and larger set of dialogs.

The experience of victimization is a relatively common occurrence for both adolescents and adults. However, victimization is more widespread among adolescents, and its relationship to various problem outcomes tends to be stronger among adolescent victims than adult victims. Whilst much of the early research on the psychological sequelae of victimization focused on general distress or fear rather than specific symptoms of PTSD, anxiety, or depression, studies have consistently found significant positive correlations between PTSD and sexual assault, and victimization in general and violent victimization in particular (Norris et al., 1997).

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 & Nishith, 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.

According to the most recent revision to the American Psychiatric Association’s DSM Disorders (2000), PTSD is divided into six major categories; refer to the DSM-IV category 309.81 for a full description and subcategories.

1. Past experience of a traumatic event and the response to the event.
2. Re-experiencing of the event with dreams, flashbacks and exposure to cues.
3. Persistent avoidance of trauma-related stimuli: thoughts, feelings, activities or places, and general numbing such as low affect and no sense of a future.
4. Persistent symptoms of anxiety or increased arousal such as hyper vigilance or jumpy, irritability, sleep difficulties or cannot concentrate.
5. Duration of the disturbance, how long have they been experiencing this.
6. 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 (re-experiencing), C (avoidance), and D (hyper vigilance). The duration of E is usually greater than one month and the effects on F (social functioning) 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.

Table 1 is an example of the types of questions asked for each of the DSM criteria for PTSD. Below that is an example dialog between a real student clinician and the automated VP character, “Justina”.

**Example Interchange between a Clinician and the Virtual Patient “Justina”**

**Clinician:** What, a, what, brings you in the office today?

**Justina:** Something bad happened that night

**Clinician:** Sorry to hear that, can you tell me what happened?

**Justina:** I was in the car with Eddie and he stopped the car and wanted me to kiss him. But when I tried to stop him, he threatened me with a knife.

**Clinician:** When did this happen to you Justina?

**Justina:** I was so scared

**Clinician:** That must have been very tough for you.

**Clinician:** Do you find yourself, ah, still being bothered by what happened?

**Justina:** Whenever thoughts of the event pop into my head I put my ipod on real loud.

**Clinician:** Ah, you mentioned thoughts pop into your head, what kind of thoughts pop into your head?

**Justina:** That horrible time replays in my mind over and over.

**Clinician:** Can you tell me, ah, what happens in your mind when you think of it over and over?

**Justina:** Something really bad happened

**4.5 Evaluation and Results**

The primary goal of the subject testing was evaluative: can a VP generate responses that elicit the proper user questions relevant for categorization and representation of a clinical condition.
Additionally, can the supporting architecture be used as a mechanism to train clinicians and provide a level of fidelity for the VP’s that will make it believable by the participants? This was accomplished through evaluation of the users’ utterances, and VP responses to those utterances, through assessment of the technology underlying the VP system and through psychophysiological responses to the VP in a bias task.

4.5.1 Subject Testing Method

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.

The VP system consisted of the virtual character “Justin” and “Justina”, as seen in Figure 1 and 2, along with a headset for the speech input and mouse button to press and hold when the subject talks and releases when they stop talking. A control station was adjacent to the subject to run the system and log the data. Additionally cameras were setup to record the subjects face and the interaction with the VP from the side for later post processing, analysis and review.

There were two subject testing sessions that were performed, one with “Justin” and one with “Justina”. A total of 21 people were involved between both set of subject testing. The sample of participants for Justin included 6 persons, medical clinicians and staff, from the University of Southern California’s Keck School of Medicine. A total of 15 people (6 females, 9 males; mean age = 29.80, SD 3.67) took part in the study for Justina.

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=8); 2) Psychiatry Residents (N=6); 3) Psychiatry Fellows (N=4), Nurses (N=3). For participation in the study, students were able to forgo certain medical round time for the time spent in the interview and completing questionnaires.

The subject testing followed a standard method of a set of pre-questionnaires, next a 15 minute interactive interview of the character followed by a set of post-questionnaires. The subjects were allowed to ask anything they wanted to the patient using the speech recognition software so as to not put any constraints on how they talked or what they talked about. Since this was an initial pilot study, the goal was also to find out what they would ask the character. The subject testing adhered to the following paradigm:

Pre-Questionnaires
1. Tellegen Absorption Scale (TAS) (Tellegen & Atkinson, 1974).
3. Virtual Patient Pre-Questionnaire (VPQ1).
4. Justina or Justin Pre-questionnaire (JPQ1).

15 Minute Virtual Patient Interview by the clinician with the automated system.

Post-Questionnaires
2. Justina Post-questionnaire (JPQ2).
3. Virtual Patient Post-questionnaire (VPQ2).

The TAS questionnaire aims to measure the subject’s openness to absorbing and self-altering experiences. The TAS is a 34-item measure of absorption. The ITQ measures individual differences in the tendencies of persons to experience “presence” in an immersive VE. The VPQ1, VPQ2 scale was developed to establish basic diagnosis competence for interaction with a patient that is intended to be presented as one with PTSD, although no mention of PTSD is on the test. We
developed the JPQ1, JPQ2 scale to gather basic demographics and ask questions related to the user’s experience and perception of the technology and how well they think the performance will be and was. The PQ is a common measure of presence in immersive virtual reality.

4.5.2 Justin Evaluation

Research has been completed to assess the system by 1) experimenter observation of the participants as they communicated with the VP; and 2) questionnaires. To adequately evaluate the system, a number of areas were used as a basis for the evaluation that included:

- **Consistency**: The behavior of the VP should match the behavior one would expect from a patient in such a condition (e.g. verbalization, gesture, posture, and appearance)
- **Adequacy**: The discourse between the VP and the participants should provide adequate verbal and nonverbal communication
- **Proficiency**: The clarity, pace, utility of VPs discourse with the participant
- **Quality**: The quality of the speech recognition of utterances spoken.

Basic usability findings revealed that the VP had high-quality overall system performance. Participants reported that the system 1) simulated real-life experience; and 2) the verbal and non-verbal behavior was satisfactory. However, results also revealed that some participants found aspects of the experience “frustrating”. For example, some participants complained that they were receiving un-anticipated responses and the system tended to repeat some responses too frequently. This was due to the speech recognition’s inability to evaluate certain of the stimulus words. Further, there were too many “brush off” responses from the VP when participant questions were outside the VP’s dialog set.

4.5.3 Justina Evaluation

The primary goal in this study was evaluative: can a virtual 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, discussion of the traumatic event, and the experience of intrusive recollections. Further, there appears to be a satisfactory amount of discussion related to the issue of avoidance. These results comport well with what one may expect from the VP (Justina) system.

Assessment of the system was completed with the data gathered from the log files in addition to the questionnaires to evaluate the number and types of questions being asked. Figure 5 is a graph showing that for the 15 minute interview the 15 subjects asked on average, 68.6 questions, lighter color, and responses by the system, darker color for each of the 8 DSM categories.

A summary of relations (measures as effect sizes “r”) was developed between each 1) DSM PTSD Category cluster of user questions; and 2) each (corresponding) cluster of responses from the VP representing the same DSM PTSD Category. These are “clusters” of Question/Response pairs that reflect categories used for differential diagnosis. The present focus is on effect sizes that describe the strength of association between question and response pairs for a given diagnostic category. For this experiment an effect size of 0.20 was regarded as small, 0.50 as moderate, and 0.80 as a large effect.

Moderate effects existed between User Questions and VP Response pairs for Category A ($r = 0.45$), Category B ($r = 0.55$), Category C ($r = 0.35$), but only small effects were found for Category D ($r = 0.13$) and Category F ($r = 0.13$). After controlling for the effects of the users openness and feeling of “being with” the virtual environment and character through the TAS, increased effects were found for Category A ($r = 0.48$), Category C ($r = 0.37$), Category D ($r = 0.15$), and Category
F (r = 0.24). The “believability” of the VP as a relationship showed strong effects existed between the TAS and ITQ (r = 0.78).

4.5.4 Virtual Patient: Assessing Bias

As mentioned above, cognitive and affective models can enhance conversational aspects of virtual humans in that they enhance human computer interactions through an incorporation of the emotional state of the user which can increase the believability of the system as a whole. This can be accomplished through evaluation of a user’s emotional state in a bias task while interacting with a VP.

In addition to the experiments mentioned above, this author’s lab has also run subjects through protocols in which we measured the activation and control of affective bias using 1) psychophysiological startle eye blink responses; and 2) self-reports as human participants interacted with VPs, “Justin and Justina” variations, representing both the human’s race and another race (Parsons et al., 2009d. We also assessed the differences in psychophysiological responses of humans interacting with VPs representing same and different sex groups. By measuring eyeblink responses to startle probes occurring at short and long latencies following the onset of the same compared with other ethnicity VPs, we were able to examine affective processes associated with both the activation and potential control of bias.

The initial participant pool included 5 adults at the University of Southern California. Four were Caucasian, one was African-American. Our findings revealed a difference score between the median blink amplitude to African American VPs and Caucasian VPs was determined. A one-way ANOVA was performed to determine if high vs. low external motivation related to physiological responses to different racial groups of virtual patients. The difference score tended to be lower in those with high external motivation to behave in a non-racist manner. Those who were lower in external motivation had a larger difference score between startle amplitudes while looking at African American vs. Caucasian VPs. The larger difference score reflects larger startle amplitudes to African American VPs, suggesting an implicit negative bias towards that group. We are currently

Figure 5. VP-DSM question / responses

![Bar chart showing average number of questions asked by DSM category (A-F).]
running additional subjects to enhance the modeling through the inclusion of other modalities such as facial expression, vocal intonation, gestures, and postures, which may be amalgamated with psychophysiology to increase the complexity of the user representation.

5. FUTURE RESEARCH DIRECTIONS

The field of creating conversational virtual patients is an emerging area that requires advanced speech and language understanding and dialog systems that are tied to the underlying physical and mental models for embodied conversational characters. Additionally, developing tools to assist in dialog and scenario building is of a great need as is standardizing clinical cases that can be used by various architectures.

To expand on the VP system architecture additional multi-modal input will allow for powerful sensing for the characters and increase the realism of the interaction. These modules that will be included in the future iterations of the VP architecture include:

- **Bio Physiological Input**: biophysiological input to the system used the Biopac system to capture heart rate, skin conductance and eye blinks. The collected data is used to evaluate the user in performing the tasks with the VP; additionally the input would be feed back into the virtual patient so that it can sense the participant and react accordingly. A version was used to assess racial bias with the VPs (Parsons, 2009b).

- **Vision Input**: The vision input consists of a users head and eye gaze information, i.e. what they are looking at, the orientation of the head and eye blink information that is used to evaluate the user’s non-verbal behavior and rapport (Morency & de Kok, 2008).

- **Natural Language Understanding (NLU)**: More complex forms of natural language parsing that form a semantic representation by matching it to semantic frames that are part of a large framebank generated from an ontology for the domain can be used to have a better understanding of the dialog by the agent. In addition to the core semantics, these frames could also includes information like speech act and modality intonation and prosody generated by a more advanced speech recognizer.

- **Natural Language Generation (NLG)**: maps an internal semantic representation generated by a Dialog Manager into a surface string. This can be similar to the process used by the NLU, but in reverse order, or based upon a domain dependent grammar. The resultant string sent would contain more complex information to add inflection to the speech or for selection of better animations of specific illnesses, emotions or moods. The reasoning phase could also take into account underlying models of cognition, emotion and personality to generate more autonomous and realistic behavior for the various clinical cases.

- **Intelligent Agent**: reasons about plans and generates actions based on its internal state and the input from the verbal text or other multi-modal devices. Complex agents can be created using a cognitive architecture that reason about plans and actions and integrate models of personality, emotion, mood and specific clinical illness criteria.

According to Fairclough (2009), the next generation of intelligent technology will be characterized by increased autonomy and adaptive capability. Such intelligent systems need to have ample capacity for real-time responsivity (Aarts, 2004). For example, to decrease the intensity of a conversation if a user is becoming too aroused...
Embodied Conversational Virtual Patients

while conversing with a virtual human, or to make the conversation more engaging if the user is bored. The psychophysiological computing approach proffers the VP a means of monitoring, quantifying, and representing user context and adapt in real-time.

We have collected (and continue to collect) quantitative and qualitative results. The VPs fit well into usability testing. Clinicians in training had positive responses to the VP and behaved as they normally would during a clinical encounter. It seems apparent that VPs will play an important role in the future of psychotherapy education for psychiatry residents and psychology trainees. The use of VPs could be implemented in several ways. For example, VPs could be developed to recognize the essential features and common pitfalls of an initial psychotherapy interview so that they could give more specific, relevant, and reliable verbal feedback to the residents involved. In addition, the use of VPs illustrating common problems such as acting out, transference, intrusive questions, or seductive behavior would allow residents to have an experience with these anxiety provoking situations in a simulated setting before they occur in their practice. Finally, performance in VP scenarios could be used as an additional source of data for the assessment of resident competency in the psychotherapy and educational domains.

6. CONCLUSION

In this chapter, there was a discussion of the ways in which advanced technologies (i.e., virtual patients) can move beyond traditional approaches to training clinicians in assessment, diagnosis, interviewing and interpersonal communication. The traditional approaches rely upon a combination of classroom learning and role-playing with human patients. Much of this work is done with actors that have been recruited and trained to exhibit the characteristics of an actual patient, thereby affording novice clinicians a realistic opportunity to practice and to be evaluated in a mock clinical environment.

Although a valuable training approach, there are limitations with the use of human patients that can be mitigated through simulation technology. For example, human patients are expensive and cost several thousand dollars per student. Further, given the fact that there are only a few sites providing standardized patient assessments as part of the U.S. Medical Licensing Examination, the current model provides limited availability.

In addition to issues of availability of trained actors, there is the issue of standardization. Despite the expense of standardized patient programs they are typically unskilled actors. As a result of common turnover, administrators face considerable challenges for offering psychometrically reliable and valid interactions with the training clinicians. The limited scope that the actors are able to portray tends to be an inadequate array of developmentally, socially, and culturally appropriate scenarios. For example, when a clinician has a pediatric focus and needs access to children, it is difficult for the clinician to pretend that the actor is a child. Finally, many clinical cases (e.g., traumatic brain injury) have associated physical symptoms and behaviors (e.g., dilated pupils, spasms, and uncoordinated movements) that simply cannot be accurately portrayed by human actors.

In this chapter a series of experiments were described to elucidate the usefulness and effectiveness of an affect-sensitive VP Interface System. 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 and the subject testing and evaluation conducted. The primary goal in this study was evaluative: can a virtual standardized patient generate responses that elicit user questions relevant for clinical illness categorization? Findings suggest that the interactions between novice clinicians and the VP resulted in a compatible dialectic in terms of
rapt, discussion of the traumatic event, and the experience of intrusive recollections of the event.

While self-report data are widely used in virtual human research, they are susceptible to modification by a participant’s awareness of the social desirability of particular responses, reducing the sensitivity of the measures, implicit behavioral and psychophysiological responses are automatic and thus considered less susceptible to self-conscious influences (Schwarz, 1999). A further issue discussed in this chapter was that the current cognitive and affective models found in virtual human research tend to use appraisal models generated from a cognitive point of view and do not adequately take into account the psychophysiological response.

It was contended that a preferred approach to developing cognitive and emotional virtual human models would include psychophysiological inputs from the humans to the virtual humans during interactions. It is believed that these additional inputs can be developed into affect-sensitive VP interfaces that go beyond conventional virtual human models designed by pure (i.e., devoid of psychophysiological metrics) cognitive appraisal principles. The resulting affect-sensitive VP interfaces would be similar to brain-based-devices (BBDs) that are being designed based on biological principles and are programmed to alter their behavior to the environment through self-learning.

In summary, effective interview skills are a core competency for training clinicians. Although schools commonly make use of standardized patients to teach interview skills, the diversity of the scenarios standardized patients can characterize is limited. Virtual Standardized Patient technology has evolved to a point where researchers may begin developing mental health applications that make use of virtual human patients for training.

REFERENCES


**ADDITIONAL READING**


**KEY TERMS AND DEFINITIONS**

**Discourse**: A formalized written or spoken communication.

**DSM**: Diagnostic and Statistical manual on Mental Disorders book that describes the properties and criteria of mental illnesses.

**Fidelity**: The level of human behavior which the avatar can generate.

**Multimodal interaction**: Provides the user with multiple modes of interfacing with a system.

**Psychophysiological**: Investigates at the way psychological activities produce physiological responses by measuring attributes such as blood pressure, heart rate, respiration, skin conductance, eyeblinks and others.

**Standardized Patient**: An actor trained to exhibit the characteristics of an actual patient.

**Subject Testing**: A method of testing and analyzing participants with an automated virtual patient system through speech, video, and questionnaire analysis.

**Text Classifier**: A statistical algorithm to match user questions to character responses from a corpus of utterances.

**Virtual Patient**: A computer generated character programmed to act like an actual patient.

**Wizard of Oz Testing**: A method to gather domain data with participants in a simulation system by having a person act out the virtual character and typing or saying appropriate responses to the participant who asks questions, like a virtual role play.