Designing Useful Virtual Standardized Patient Encounters

Thomas B. Talbot, MD, Kenji Sagae, PhD, Bruce John, Albert A. Rizzo, PhD
University of Southern California Institute for Creative Technologies
Playa Vista, CA
Talbot@ict.usc.edu, sagae@ict.usc.edu, bjohn@ict.usc.edu, Rizzo@ict.usc.edu

ABSTRACT

Developers and educators have explored many different ways to create "Virtual Patients" as a method to simulate a patient encounter. Some of these attempts have been educationally useful, yet no approach taken to date has satisfactorily replicated the Patient-Doctor encounter in a way that can be generalized nor have the best developments to date been readily author-able by regular medical educators. The best simulator to date is the human standardized patient actor, which has considerable disadvantages. The manner in which a virtual standardized patient can be designed requires a breakdown of the clinical encounter into components and a strategic approach to simulating each phase. These components are compared to find the optimal approach for each part of the medical encounter. The paper proposes a blend of an artificially intelligent statistical matching dialogue system with multiple choice state machine-based sub-conversations as a way in which one may richly simulate the interview and counseling phases of the clinical encounter. Also elucidated are the steps necessary for educator author-ability and approaches that will extract rich, objective assessment data. If such integration proves to be successful, the result will be a rich conversational clinical simulation that closely approximates Patient-Doctor encounters.

ABOUT THE AUTHORS

Thomas B. Talbot, MD, MS is the principal medical expert at the USC Institute for Creative Technologies, where he researches virtual patient and medical virtual reality science. Dr. Talbot is the founder of the Armed Forces Simulation Institute for Medicine at the Telemedicine and Advanced Technology Research Center. Through these efforts, Dr. Talbot has created the nation's largest medical education focused research and development program. He is a pediatrician and veteran of the US Army with wartime and simulation development experience. Research interests include serious games, natural user interfaces, VR and augmented reality, virtual interactive humans and microcontrollers.

Kenji Sagae, PhD is a research assistant professor in the Department of Computer Science at the University of Southern California and a research scientist in the USC Institute for Creative Technologies. His main areas of research are natural language processing and computational linguistics, focusing on data-driven approaches for syntactic parsing, predicate-argument analysis and discourse processing. His current work includes the application of these techniques in various application areas, such as dialogue systems, modeling of human communication dynamics, and analysis of personal narrative in blog posts.

Bruce John is a full-time researcher and game designer at the USC Institute for Creative Technologies, while also holding a part-time position at the Headington Institute as a researcher and game designer. He has completed programs in Psychology and Law, and Video Game Design at, while also earning a BS in Policy, Management, and Planning from, the University of Southern California. His current work involves allostatic load, resilience, educationally based mobile games, and other pedagogical, interactive environments such as those using virtual humans.

Albert "Skip" Rizzo, PhD directs the Medical Virtual Reality Group at the USC Institute for Creative Technologies. He is a Research Professor in Psychiatry and Gerontology. Dr. Rizzo conducts research on the design, development, and evaluation of Virtual Reality systems targeting the assessment, training, and rehabilitation of spatial abilities, attention, memory, executive function, and motor abilities. His latest projects have focused on the translation of the graphic assets from the X-Box game, Full Spectrum Warrior, into an exposure therapy application for combat-related PTSD, and the use of those assets to also create a resilience training system for pre-deployed military personnel.

Designing Useful Virtual Standardized Patient Encounters

Thomas B. Talbot, MD, Kenji Sagae, PhD, Bruce John, Albert A. Rizzo, PhD
University of Southern California Institute for Creative Technologies
Playa Vista, CA
Talbot@ict.usc.edu, sagae@ict.usc.edu, bjohn@ict.usc.edu, Rizzo@ict.usc.edu

Drs. Barrows and Abrahamson (1964) introduced the 'the programmed patient' or human standardized patient in 1963 (Talbot & Rizzo, 2012). Since that time, there have been numerous attempts to replicate the experience on a computer for anytime-anywhere access to the experience. With the now ubiquitous use of computers in medicine, there has been a great deal of effort in the area of conversational agents and different types of virtual patient experiences (Poulton & Balasubramaniam, 2011; Cook & Triola, 2009). This article will explore dialogue-based conversational agents as virtual standardized patients and the approaches to creating them. Finally, we will offer a theoretical approach that synthesizes several educational approaches over the course of a medical encounter and recommend a strategy that may finally serve as a satisfactory approach to a Virtual Standardized Patient (VSP) experience.

THE HUMAN STANDARDIZED PATIENT

Human Standardized Patients (HSPs) have been considered to be the gold standard medical education experience for both learning and evaluation purposes (Collins and Harden, 1999; Adamo, 2004; Jack et al., 2009). HSPs are paid actors who pretend to be patients for educational interviews or patients with physical findings who are paid to receive physical examinations. HSPs provide the most realistic and challenging experience for those learning the practice of medicine because they can most closely approximate a genuine patient encounter. HSPs are without peer in the evaluation of physical examination techniques. HSPs are also a key component in medical licensing examinations.

In an HSP encounter, a student will be given a limited set of up front information such as a background statement, chief complaint or abbreviated medical history. The dynamic of the encounter is such that the students must introduce themselves, establish rapport and use their investigative skills to elicit verbal responses from the HSP. The HSP, who is following a script, might offer clues in their responses for the student doctor to follow up on. The general interaction pattern is: doctor asks and patient answers. Typically, HSPs do not ask questions of the student doctor. For encounters with a physical exam, HSPs rate whether something has been checked. If examination techniques are to be evaluated, this is done via means of a clinician proctor.

HSP encounters engage a number of clinical skill domains such as social skills, communication skills, judgment, and diagnostic acumen in a real time setting. When it comes to these interviewing skills, other kinds of practice encounters, in our opinion, fall short because they either do not force the learner to combine clinical skill domains or those approaches spoon feed data to the student with the resulting practice case turning more into a pattern recognition exercise rather than a realistic clinical problem solving experience.

Despite the well-known advantages of HSPs, they are employed sparingly due to the high expense of hiring and training actors (Parsons et al., 2008). Moreover, the actors themselves are typically a low skilled and high turnover population resulting in challenges for maintaining the consistency of diverse patient portrayals. In practice, HSPs will often provide verbally correct answers but the medical accuracy of their responses can be as low as 30% (Tamblyn et al., 2009). This limits the value of this approach for producing realistic and valid interactions needed for the reliable evaluation and training of novice clinicians. Also, the diversity of clinical conditions that standardized patients can characterize is limited by availability of human actors and their skills. This is even a greater problem when the actor needs to be a child, adolescent, elder, a person with a disability or in the portrayal of nuanced or complex symptom presentations.

Regarding evaluation of these encounters, the imposition of recording has been reported to have demonstrable effects that may confound the end goal of clinical training (Bogolub, 1986). Additionally, the supervisor review of raw recordings is a time consuming process that imposes a significant drain on resources. Other sources of unreliability and bias in HSPs have been reported such as intra-rater reliability with 20% score variation and significant differences between raters in student pass rates (Tamblyn et al., 1991). The most common complaints about HSPs, that the authors have heard in medical student focus group encounters, mirrors the results of these research studies. Students express the desire for more than the current limited opportunities to use HSPs and for the ability to retry an encounter. They also complain about the lack of detailed performance feedback and the students' perception of bias by the HSP or observer.

THE WORLD OF VIRTUAL PATIENTS

Many different entities with unique approaches and attributes are all often called virtual patients. Such approaches include case presentations, virtual patient interactive scenarios, virtual patient games, high fidelity software simulations and virtual human conversational agents. Salient features of these virtual patient approaches are summarized in the literature (Triola, 2006; Talbot, 2012).

Case presentations test case-specific knowledge of a condition and its treatment. Virtual patient interactive scenarios extend case presentations while allowing for student choices, and evaluate clinical decision making. Virtual patient games test team training and interventional medicine, often with a dynamic avatar and treatment environment. High fidelity software simulations train surgeries in virtual settings and advanced interventional skills. Virtual human conversational agents emulate the interactions seen in HSP encounters and consist of free-text/speech interactive dialogues or prompted dialogue systems (Talbot & Rizzo, in press).

Virtual Human Conversational Agents

Recently, seminal research and development has appeared in the creation of highly interactive, artificially intelligent and natural language capable virtual human (VH) conversational agents. No longer at the level of minimal faux interaction in a virtual world, these VH agents are designed to perceive and act in a 3D virtual world and engage in face-to-face spoken dialogues with real users. In some cases, they are capable of exhibiting human-like emotional reactions. Artificially intelligent VH agents can now be created that control computer generated bodies and can interact with users through speech and gesture in virtual environments (Gratch et al., 2002). Advanced virtual humans can engage in rich conversations (Traum et al., 2008), recognize nonverbal cues (Morency et al., 2008), reason about social and emotional factors (Gratch & Marsella, 2004) and synthesize human communication and nonverbal expressions (Thiebaux et al., 2008). Such fully embodied conversational characters have been around since the early 90's (Bickmore & Cassell, 2005) and there has been much work on full systems to be used for training (Evans et al., 1989; Kenny et al., 2007; Prendinger & Ishizuka, 2004; Rickel et al., 2001; Rizzo et al., 2011), intelligent kiosks (McCauley & D'Mello, 2006), and virtual receptionists (Babu et al., 2006).

In this regard, *Virtual Standardized Patients (VSPs)*, a specific kind of virtual human conversational agent, can be used in the role of standardized patients by simulating a particular clinical presentation with a high degree of consistency, credibility and realism (Stevens et al., 2005), as well as being always available for *any-time-anywhere* training. There is a growing field of researchers applying VSP's to training and assessment of bioethics, basic patient communication, interactive conversations, history taking, clinical assessments, and clinical decision-making (Bickmore et al., 2006, 2007; Lok et al., 2007; Kenny et al., 2007; Parsons et al., 2008; Rizzo et al., 2011). Initial results suggest that VSPs can provide valid and reliable representations of live patients (Triola et al., 2006; Andrew et al., 2007). VSP applications can likewise enable the precise stimulus presentation and control (dynamic behavior, conversational dialog and interaction) needed for rigorous laboratory research, yet embedded within the context of ecologically relevant simulations of clinical environments (Kenny et al., 2007; Parsons et al., 2008; Andrew et al., 2007).

How Virtual Human Technology Functions

VSP systems require a complex integration of technologies. A general VSP architecture can be created to support a wide range of verbal interaction levels from simple question/answering to more complex approaches that contain cognitive and emotional models with goal-oriented behavior. Such architectures are modular distributed systems with many components that communicate by message passing. Each module may contain various sub-components. For example, the natural language section is divided into three components: a part to understand the language, a part to manage the dialog and a part to generate the output text. This is all combined into one statistical language component. Interaction with the system might require that a user enters text as input or talks into a microphone that records the audio signal that is sent to a speech recognition engine. With voice recognition, the speech engine converts that into text. The text is then sent to a statistical response selection module. The module picks an appropriate verbal response based on the input text question. The response is then sent to a non-verbal behavior generator that selects animations to play for the text, based on a set of rules. The output is then sent to a procedural animation system along with a pre-recorded or a generated voice file. The animation system plays and synchronizes the gestures, speech and lip-syncing for the final output to the screen. The user then listens to the response and asks more questions to the character.

Due to strengths of their dialogue system AI, VSPs excel at interview and counseling skills applications. Additionally, VSPs can be constructed so that they provide features not found in human standardized patients such as reliable, bias free assessments with detailed reporting to the learner and the possibility of repeated performances. Extensive work has been conducted on full feature VSPs by the USC Institute for Creative Technologies MedVR group (Rizzo et al, 2011). The Virtual Experience Research Group (http://verg.cise.ufl.edu) at the University of Florida also builds dialogue AI systems and virtual patients (Rossen et al., 2010).

Achieving Natural Language Dialogue with Virtual Humans

Despite several decades of work on artificial conversational agents, building machines that can come close to mimicking the human language ability in general unconstrained conversation remains largely out of the reach of current artificial intelligence and natural language processing. This is one of the problems that came to be known informally as AI-complete, and is also a central piece of the Turing Test (Turing, 1950). Substantial progress has been made, however, in constrained scenarios where the vocabulary and topics of conversation are limited, allowing for the creation of artificial agents that interact with users for specific purposes, such as providing information or participating in training simulations. Fortunately, Virtual Standardized Patient interviews fit well into a constrained scenario.

Although early work on conversational agents produced entertaining results, such as the well-known ELIZA program (Weizenbaum, 1966) and its various descendants, chat-bots based on these techniques act based on simple pattern matching rules, without a purpose or general coherence that takes user input into account over the course of a dialogue. Current work on intelligent conversational agents far surpasses the capabilities of early chat-bots, endowing machines with language and reasoning abilities based on statistical modeling, natural language processing, theories of discourse structure and various forms of artificial intelligence. Much of this work falls within the field of dialogue systems research, which overlaps with artificial intelligence (AI) and natural language processing.

Artificial Intelligence Dialogue Systems

Within the broad spectrum of dialogue systems, which includes for example telephone-based voice-enabled airline reservation systems, Google Voice and Apple's SIRI, conversational virtual human dialogue systems are of particular relevance to VSPs. Virtual humans are embodied agents that communicate through verbal (language) and nonverbal (gestures, facial expressions) channels. The natural language and dialogue technology that powers virtual humans is varied and often selected or developed to match the characteristics of particular applications. If the interaction between the virtual human and human user is tightly scripted, such that user utterances are restricted to a finite set of choices that is presented by the system, approaches such as finite state control using call flow graphs (Pieraccini and Huerta, 2005) and branching narrative for interactive games (Tavinor, 2009) are sufficient. On the other hand, if more varied natural language input is to be expected by the virtual human, one general technique that is widely used and is well matched for virtual humans that provide information by answering questions about a certain topic (e.g. a product, a company, or a website) is to model the correspondence between arbitrary questions within the relevant topic to a finite set of predefined answers. A specific technique that accomplishes this modeling task is that of cross-lingual information retrieval, which is applied to this question-answer mapping by treating questions as information queries in one language, and answers as documents in a different language (Leuski and Traum, 2010). This approach generalizes over the set of questions so that questions that the system has never encountered before can be matched to the most relevant answer from the predefined set, much like queries to a search engine are matched to documents in a collection. This has been used successfully in deployed virtual human systems ranging from museum guides at the Museum of Science in Boston (Swartout et al., 2010) and virtual patients (Kenny et al., 2010) to promotional characters for the US Army (Artstein et al., 2008; Leuski et al., 2006) and guides in virtual worlds (Jan et al., 2009), such as Second Life.

Although much can be accomplished in practice using agents that either limit user input in exchange for coherence over a long dialogue, or with question-answer agents that handle free input, but only one question at a time, these approaches sacrifice important conversational aspects of dialogue to achieve robustness in different ways. More sophisticated approaches to dialogue management take a more nuanced approach, making these trade-offs in more graded fashion to achieve acceptable levels of both understanding of natural language without a set of fixed choices and coherence over longer conversations about a specific topic. This is accomplished in part by a Dialogue Manager, which serves as the brain of a virtual human, tracking its state through the interaction and its knowledge, and deciding what it should do or say depending on user input and the virtual human's own goals. By combining techniques used for the question-answer agents discussed above with suitable dialogue management, virtual humans can be created for tasks such as training in interrogation and questioning related to specific incidents, as in the TACQ virtual human dialogue system (Gandhe et al., 2008). These types of interaction can be both rich and coherent, but are heavily driven by user questions. The addition of interaction goals and inference capabilities to the dialogue manager allow for interactions with mixed initiative, where topics of conversation can be brought up by either the user or the system. One system that uses this style of dialogue management is the SimCoach virtual human system (Rizzo et al., 2011; Morbini, in press), where virtual characters attempt to create rapport with military users to encourage exploration of health care options related to PTSD and depression. In this system, it is important for the virtual human to take the initiative and lead the conversation when necessary, but to give users enough room to change the topic or express themselves more freely than in a system with limited choices for user input. Other more sophisticated systems aim to model human language, planning and reasoning skills, as well as emotion, using dialogue managers based on a cognitive architecture (Traum et al., 2008). Such systems allow for rich behavior and multi-party interaction, but authoring of new content and new interaction scenarios is a challenging task that can be performed only by dialogue system experts. In comparison, new characters for new scenarios under the TACQ and SimCoach platforms mentioned above can be authored by domain experts, as opposed to dialogue system experts, or simply by creative writers (Gandhe et al., 2011).

Low Technology Approaches to Dialogue Play

One method of making a conversation much easier to author while avoiding the complex technology of dialogue systems is to employ multiple choice prompts that either drive video clips or a virtual character dialogue. The Virtual Child Witness (VCW) project at USC Institute for Creative Technologies modeled the rapport building phase of an investigative interview. Users select a series of multiple choice prompts that vary in their open-endedness and therefore in their productivity in eliciting narrative reports. The project was intended to demonstrate, through a virtual child, the varying effectiveness of different question types at inducing detailed accounts during the narrative rapport phase of an investigative interview. Pilot data has supported the program's efficacy in assessing the user's interviewing skills and in serving as an engaging device for training. Students are able to adjust their specific questions in real time in response to the answers and behavior of the virtual child. Both real-time and post-interaction feedback can help guide the individual student toward an optimal interview approach (Lamb et al., 2007). While VCW is a linear scenario it flows well due to a well written narrative along with mutually exclusive and intentionally imperfect choices. Scenarios have great replay value despite the simple AI used to create them (Campbell et al., 2011).

Another method of driving simulated conversations without a natural language dialogue engine is to employ menu based prompts. Multiple choice interactions can be selected by topic, allowing for conversations about selected topics in any desired order. This approach makes the conversations appear more flexible and under control by the user, but their suitability for simulating an HSP encounter is limited because the student doctor does not generate the inquiries but chooses between options. Thus, the exercise becomes one of judgment more than investigation. Nevertheless, many guided learning experiences and clinical judgment scenarios will fare better in a constrained choice environment (Gandhe, 2011).

BUILDING A BETTER VIRTUAL PATIENT FOR CLINICIAN LEARNERS: A THEORETICAL APPROACH

The physician-patient encounter has a number of different elements, as do similar encounters performed by allied health professionals. The phases of a physician clinical encounter include the interview, the physical examination, test data review, diagnosis & treatment plan, procedures, and patient counseling. Clinicians in training may be expected to present the patient case or be quizzed on aspects of it during or just after the patient encounter. Most human standardized patient encounters are centered on the interview or physical examination.

The Interview

When participating or observing HSP encounters, it becomes obvious that the interview is highly dependent on social skills and diagnostic acumen. The interview tends to follow the 'doctor asks, patient answers' model. Fortunately, this style of interaction is very well suited to AI dialogue systems because they are human question / computer answer oriented making it is easy to provide boundaries to the encounter. It is also noteworthy that the student doctor is responsible for rapport by providing an introduction and by building rapport through social statements and inquiries. More experienced clinicians will begin with broad, open questions to elicit a narrative from the patient and obtain clues for further follow up. Based upon these clues, the clinician will then follow up with more specific questioning. Novice clinicians tend to neglect open ended questions while favoring a comprehensive list of closed ended queries. The ability to handle rapport building, open ended and close ended questions within a single encounter normally requires considerable effort to develop.

The challenge to creating an optimal VSP encounter is to create an interview experience that allows for natural language input with simultaneous support for rapport, open questioning and closed questioning phases. Additionally, the system should provide for robust assessment, possibly pedagogical assistance and be author-able by medical educators. The key challenge to populating a dialogue manager is coming up with the questions that the user may ask, often because similar queries can be phrased many ways. This difficulty can be mitigated by the fact that doctors are mostly going to ask similar questions regardless of the patient case. Rapport and open ended questions are universally the same regardless of the VSP medical diagnosis and the closed ended questions will tend to follow a medical 'review of systems' or body system focused model. So in theory, most interview questions only need to be placed into the dialogue manager once, for nearly every specific VSP one would want to create. As an example, rapport statements (Figure 1) are limited in number and not case specific. The same can be said of open ended questions (Figure 2).

When it comes to specific patient complaints, doctors tend to organize their line of questioning according to body or physiological systems. They will ask about general medical history or social history then proceed to the neurological system, respiratory system, et cetera and so on. The comprehensive review of systems and the closed ended questions to address them are globally consistent, though the different systems that may be addressed in a particular encounter depend on the diagnoses that the physician is trying to rule in or out (Figure 3). For example, a patient complaining of diarrhea will be asked questions about constitutional symptoms like fever and the digestive system. A patient with headaches may be asked about their sleep history and the neurological system. Both patients, however, will be asked about their dietary history. The questions may be the same, but the answers will differ greatly based upon the actual VSP history. It is fully possible to construct a comprehensive list of likely questions (Swartz, 1994) that cover a general medical history and a full review of systems.

If all these questions are placed into a dialogue manager, they can represent the basis of a 'default patient' question set. This comprehensive question list would also have to account for multiple ways to ask many of these questions (Have you had fever? vs. was your temperature high?). Given that approach and the extreme difficulty in accounting the many different close ended questions that can be asked about fever (Is your fever high? Do you have fever? How high was the fever? et cetera.), any questions referring to the topic of fever may be directed to a single response like "I've had fever for two days, up to 102.5 last night".

Hello, I am Doctor XXXX
Good Morning.
How do you do today?
Tell me something about yourself
It's nice to see you today
I'm glad to meet you
I like your XXXX
What do you prefer I call you?
Thanks for seeing me
This must have been sad
I understand this must be hard for you

Figure 1 - Common greeting, empathy and rapport building statements.

What brings you in today?
What kind of problem are you having?
How can I help you?
What can I help you with?
Why did you come to the doctor?
What's wrong?
Why are you here?
Tell me more
Tell me more about that?
How is your health?
And then what happened?
What else can you tell me?
Can you think of anything else?
Anything else I should know?
Is there more?

Figure 2 - Open ended questions include basic questions and follow on queries.

Do you have cough?
- Is it productive?
- Are you coughing anything up?
- Any nighttime cough?
- Are you wheezing?
- Are you short of breath?
- Is it keeping you awake?

Do you have fever?
Where does it hurt?
Are you sleeping okay?

Figure 3 - Closed ended questions. The indented questions relating to cough are 'local context' questions that refer to either cough or the respiratory system. With local context, asking 'is it keeping you awake' will evoke a response about the cough, rather than knee pain.

The advantage of the default question set approach is that most of the work and complexity of intelligent agent creation can be focused on a single effort, then duplicating the product of that effort many times over each time a new VSP case is needed. Additionally, many of the responses will be the same or similar between patient cases, especially in regards to review of systems type answers. The default patient will benefit from having default 'normal' health answers pre-authored as well. The diarrhea patient will have different gastrointestinal system answers and the headache patient will have different neurological answers, but both will probably report the same normal respiratory or musculoskeletal system answers. Thus, pre-authoring normal review of system (ROS) patient responses for the default patient becomes a necessity that vastly reduces the burden for VSP case authors who can now modify the desired responses to address the particular case being created.

Although it has been mentioned that true bidirectional conversations are difficult to author with virtual human technology, some cheats are possible. For example, following up a patient verbal response with a multiple choice option for a follow up statement allows for the creation of provocative and rich scenarios if employed sparingly. Student doctor selections can affect the subsequent reply and assessment variables. Such a technique will not pose an undue burden to dialogue management tools such as those used by the authors.

To summarize an optimal design for a dialogue based VSP interview, questions types such as introductions, rapport, open-ended questions and close-ended review of system questions can, for the most part, be pre-authored by system developers. Most of the responses can also be pre-authored with 'normal' or negative symptom answers. Authors who write VSP encounters then have the reduced burden of authoring only the patient history, narrative (open-ended response) and any ROS items that will be different for that patient. The difference between this much more facile approach versus the need to write encounters from scratch is that the economies of scale should then permit VSP encounters to be authored by medical educators, as compared to teams of computer scientists. This next step will greatly advance, simplify and democratize the ability to employ VSPs for medical education and represents the near term focus of the authors' research.

Assessment

As mentioned previously, a significant limitation of HSP assessment concerns intra-rater reliability, inter-rater reliability, perceived objectivity and reliability by the student and detailed feedback on the student doctor's performance (Tamblyn, 1991). VSP authors should be able to tag the dialogue responses with assessment classifications such as 'must ask item', 'secondary importance item', 'rapport builder' and 'critical error' categories. By tying assessment triggers to responses instead of the questions a student may ask, the number of items that need to have assessments attached is greatly reduced. The 'must ask item' category refers to critical items needed to determine diagnosis and rule out more serious conditions. The 'secondary importance items' category of tags represents information that would fill out the story and represent a thorough interview but are not critical to establishing the diagnosis. 'Rapport builder' tags, probably already placed by default, refer to assessment tags that increase rapport scores. The 'critical error' category refers to responses that are triggered by grossly inappropriate interactions on part of the student.

Rapport and critical errors aside, the simple act of tagging desired VSP responses as 'must ask' or 'secondary importance' can yield a wealth of assessment and feedback data to the student. Firstly, it can be used to develop a performance score and missed 'must ask' items can be reported back to the student and faculty. Patterns of what the student misses over several encounters can also be compiled from one or multiple cases to determine a pattern of missed information that can guide the student. Second, the number of questions posed by the student doctor since the last score earning response up to the end of the encounter can be divided into the total number of questions posed to create a perseveration index. A perseveration index is a measure of how long the student keeps in the interview going after obtaining all the necessary information. Calculating the number of scoring responses vs. the number of total questions posed can be used to produce a diagnostic efficiency rating. Finally, since the data collection in VSP sessions is both hidden and automatic, it is hopeful that effects related to performance recording will not alter student performance like they can in videotaped HSP encounters.

Pedagogical Guides

The VSP learning experience can benefit from intelligent guides for the student doctor. Symbols on an on-screen overlay that depict assessment item achievements can be employed to show progress towards completing the encounter, for example. Another pedagogical guide could be a virtual attending physician (VAP) agent. Such an agent is simply a control that when selected, asks a proper question chosen randomly from the correct assessment items for the current case. This sort of help creates scaffolding for the student doctor while maintaining the conversational flow of the VSP natural language encounter.

Physical Examination

The physical examination requires cognitive knowledge of when it is appropriate to conduct a particular exam as well as procedural knowledge to correctly perform the examination maneuvers and to interpret exam findings. The gold-standard is a real patient encounter or a human standardized patient. No virtual patient can yet substitute for training or testing the procedural knowledge of the physical exam, though a limited subset of the exam can be practiced on high-fidelity manikins. VSPs can address the cognitive aspects and the interpretation of results. The best virtual patient platforms to address these are Virtual Patient Games, using their high fidelity graphics systems, as well as simple multimedia that can show photographs of physical exam findings.

Physician Decision Making Tasks

Physician tasks after the interview and before patient counseling include test result interpretation, rendering the diagnosis and formulation of a treatment plan. These tasks do not really benefit from interactive virtual human encounters. Optimal methods to simulating these features include computer-based forms, multimedia control panels, short answer tests and multiple choice questions. The technology threshold to successfully emulate this is very low.

Patient Counseling

Patient counseling is the phase of the encounter when the physician explains the diagnosis and treatment plan. He will ask for and respond to patient questions, sometimes bargain with the patient and will finally try to estimate patient understanding of the treatment plan. This is a very difficult phase to replicate with a virtual standardized patient natural language dialogue system. Other approaches that might be more practical include multiple choice style dialogues that guide the experience on a predictable trajectory.

The reason that the counseling phase of the VSP interaction is so much more difficult to author with a dialogue management system is because counseling involves more extended exposition on part of the student doctor and because the interaction tends to be more of a bidirectional conversation. Current artificial intelligence systems work much better with short sentences that are about a single topic as well as conversations that tend to be driven by just one of the two conversation parties. Bidirectional conversations challenge the dialogue system. Including such features dramatically increases scenario authoring difficulty (Traum et al., 2008). Counseling interactions using natural language agents is not likely to be mastered before VSP interviews are, as interview sessions play into many more of the strengths of today's dialogue-based systems.

CONCLUSION

It is entirely possible to satisfactorily replicate a robust patient interview such as with an actual patient or human standardized patient with a virtual standardized patient. We have introduced the technology that can make this possible and provided a vision for organizing natural language dialogue systems in a way that can facilitate the multiple phases of interview interactions within a single virtual human session.

When attempting a robust VSP interview, it is absolutely essential to provide for rapport building, open ended questions that lead to narrative responses and a full review of systems based set of short answers.

A major step towards making cases author-able by medical educators will be to pre-populate most possible questions and to provide for a full set of normal/negative responses to the review of systems questions. This combination of pre-authored items and guided authoring wizards will reduce the likely single case authoring burden by more than 95 percent as a conservative estimate. This estimate is based upon our experience creating VSPs, knowing the time involved in authoring steps and what work could be supplanted with the approaches described in this paper. With this approach, most cases will merely require the authorship of patient history details, history of present illness narrative and pertinent changes to review of systems responses as is appropriate to that case. Future research will seek to validate the efficacy of this approach and obtain specific data to quantify actual authoring time and effort reductions.

It will be essential to develop a new type of authoring tool that is specific to medical educator authorship of VSP cases that provides for simplified data input for unique case data and modification of existing pre-authored responses. Assessment tagging must also be easy to perform in such as system.

The combination of virtual human natural language dialogue AI, the strategies mentioned herein for VSP authoring, robust assessment and a simple, educator friendly case editor may very well prove to be the right break-through combination needed to make quality virtual standardized patient encounters a widely available tool for medical education. This approach embodies our teams' near term research objectives for VSPs.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Defense Health Medical Research Program, the Defense Advanced Research Projects Agency, the Telemedicine and Advanced Technology Research Center, the US Army Medical Research and Materiel Command, the Army Research Laboratory and US Army RDECOM for funding our virtual human and virtual patient work.

REFERENCES

Adamo, G. (2004). Simulated and standardised patients in OSCEs: achievements and challenges 1992–2003. *Medical Teaching*, 25(3): 262–70.

Andrew, R., Johnsen, K. et al., (2007). Comparing Interpersonal Interactions with a Virtual Human to those with a Real Human. *IEEE Transactions on Visualization and Computer Graphics. 13*(3), 443-457.

Artstein, R., Gandhe, S., Leuski, A. & Traum, D.R. (2008). Field Testing of an interactive question-answering character. *Proceedings of the ELRA workshop on evaluation*, 36-40. Marrakech, Morocco.

Babu, S., Schmugge, S., Barnes, T., & Hodges, L. (2006). What Would You Like to Talk About? An Evaluation of Social Conversations with a Virtual Receptionist. In J. Gratch et al. (Eds.): *IVA 2006*, LNAI 4133, Springer-Verlag, Berlin, Germany. pp. 169-180.

Barrows, H.S. & Abrahamson, S. (1964). The programmed patient: A technique for appraising student performance in clinical neurology. *Journal of Medical Education*, 39: 802-805.

Bickmore, T. & Cassell, J. (2005). Social Dialogue with Embodied Conversational Agents. In J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (eds.), *Advances in Natural, Multimodal Dialogue Systems*. New York: Kluwer Academic.

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

Bogolub, E.B. (1986). Tape Recorders in Clinical Sessions: Deliberate and Fortuitous Effects. *Clinical Social Work Journal.* 14(4), 349-360.

Campbell, J., Core, M., Artstein, R., Armstrong, L., Hartholt, A., Wilson, C., Georgila, K., Morbini, F., Gomboc, D., Birch, M., Bobrow, J., Lane, H. C., Gerten, J., Leuski, A., Traum, D., Trimmer, M., DiNinni, R., Bosack, M., Jones, M., Clark, R. & Yates, K. (2011, March) Developing INOTS to Support Interpersonal Skills Practice. *IEEE Aerospace Conference.*

Collins, J. P. & Harden, R. M. (1999). The Use of Real Patients, Simulated Patients and Simulators in Clinical Examinations. *Association for Medical Education in Europe.* Accessed on Jan. 4, 2011 at: http://www.medev.ac.uk/resources/features/AMEE_summaries/Guide13summaryMay04.pdf

Cook, D.A. & Triola, M.M. (2009) Virtual patients: a critical literature review and proposed next steps. *Medical Education*, 43:303-311.

Evans, D., Hern, M., Uhlemann, M. & Lvey, A. (1989). *Essential Interviewing: A Programmed Approach to Effective Communication* (3rd Ed): Brooks/Cole Publishing Company.

Gandhe, S. , DeVault, D., Roque, A., Martinovski, B., Artstein, R., Leuski, A., Gerten, J. & Traum, D.R. (2008). From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters. In Proceedings of Interspeech. Brisbane, Australia

Gandhe, S., Taylor, A., Gerten, J. & Traum, D.R. (2011). Rapid Development of Advanced Question-Answering Characters by Non-experts. In: Proceedings of the SIGDIAL 2011 Conference, *The 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue.* Portland, Oregon, USA: The Association for Computer Linguistics. p. 347-9.

Gratch, J. & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research. 5*(4): 269-306.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E. & Badler, N. (2002). Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems*. July/August: 54-61.

Jack, B., Chetty, V., Anthony, D., Greenwald, J., Sanchez, G., Johnson, A., Forsythe, S., O'Donnell, J., Paasche-Orlow, M., Manasseh, C., Martin, S. & Culpepper, L. (2009) A Reengineered Hospital Discharge Program to Decrease Rehospitalization: A Randomized Trial. *Ann Intern Med.* 150(3):178–187.

Jan, D., Roque, A. , Leuski. A., Morie, J., & Traum, D.R. (2009). A Virtual Tour Guide for Virtual Worlds. *IVA Proceedings of the 9th International Conference on Intelligent Virtual Agents.* Amsterdam, Netherlands: Springer.

Kenny, P., Parsond, T. & Garrity, P. (2010, November) Virtual Patients for Virtual Sick Call Medical Training. *Interservice/Industry Training, Simulation, and Education Conference.*

Kenny, P., Rizzo, A.A., Parsons, T., Gratch, J. & Swartout W. (2007). A Virtual Human Agent for Training Clinical Interviewing Skills to Novice

Therapists. *Annual Review of Cybertherapy and Telemedicine 2007. 5*, 81-89.

Lamb, M.E., Orbach, Y., Hershkowitz, I., Esplin, P.W. & Horowitz, D. (2007) Structured forensic interview protocols improve the quality and informativeness of investigative interview with children: A review of research using NICHD Investigative Interview Protocol. *Child Abuse & Neglect,* 31(11-12):1201-1231.

Leuski, A. & Traum, D.R. (2010). Practical language processing for virtual humans*. In Proceedings of the Twenty-Second Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-10).*

Lok, B., et al. (2007). Applying Virtual Reality in Medical Communication Education: Current Findings and Potential Teaching and Learning Benefits of Immersive Virtual Patients. *Jour. of Virtual Reality. 10*(3-4), 185-195.

McCauley, L. & D'Mello, S. (2006). A Speech Enabled Intelligent Kiosk. In J. Gratch et al. (Eds.): *IVA 2006*, LNAI 4133, Springer-Verlag, Berlin, Germany. pp. 132-144.

Morbini, F., DeVault, D., Sagae, K., Nazarian, A., Gerten, J. & Traum, D.R. (in press). FLoReS: a Forward Looking, Reward Seeking, Dialogue Manager. *Submitted to the 13th SIGDIAL Meeting on Discourse and Dialogue.*

Morency, L.-P., de Kok, I. & Gratch, J. (2008). Context-based Recognition during Human Interactions: Automatic Feature Selection and Encoding Dictionary. *10th International Conference on Multimodal Interfaces*, Chania, Greece, IEEE.

Parsons, T. D., Kenny, P. et al. (2008). Objective Structured Clinical Interview Training using a Virtual Human Patient. *Stud. in Health Tech. and Informatics. 132*, 357-362.

Pieraccini, R. & Huerta, J. (2005). Where do we go from here? Research and commercial spoken dialog systems. *In Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue.* Lisbon, Portugal.

Prendinger H. & Ishizuka, M. (2004). *Life-Like Characters – Tools, Affective Functions, and Applications*, Springer.

Pulton, T. & Balasaburamaniam, C. (2011) Virtual patients: A year of change. *Medical Teacher*, 33:933-937.

Rickel, J., Gratch, J., Hill, R., Marsella, S. & Swartout, W. (2001). Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences. *The Proceedings of the AAAI Spring Symposium on AI and Interactive Entertainment*, Stanford University, CA.

Rizzo, A.A., Kenny, P. & Parsons, T. (2011). Intelligent Virtual Humans for Clinical Training. *International Journal of Virtual Reality and Broadcasting.* 8:3. Available: http://www.jvrb.org/8.2011/

Rizzo, A.A., Lange, B., Buckwalter, J.G., Forbell, E., & Kenny, P. (2011). An intelligent virtual human system for providing healthcare information and support. *In Studies in Health Technology and Informatics*.

Rossen, B. & Lok, B. (2012). A crowdsourcing method to develop virtual human conversational agents. *International Journal of Human-Computer Studies.* 70:301-319.

Rossen, B., Cendan, J. & Lok, B. (2010). Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans. In Allbeck et al (Ed.): *IVA 2010, LNAI 6356*, pp 392-398.

Stevens, A., Hernandez, J. et al. (2005). The use of virtual patients to teach medical students communication skills. *The Association for Surgical Education Annual Meeting*. NY, NY.

Swartout, W., Traum, D.R., Artstein, R., Noren, R., Debevec, P., Bronnenkant, K. & Williams, J. (2010). Ada and Grace: Toward realistic and engaging virtual museum guides. *Intelligent Virtual Agents*: 286–300.

Swartz, M.H. (1994) *Textbook of physical diagnosis: history and examination, 2nd ed.* W.B. Saunders, Philadelphia.

Talbot, T.B., Sagae, K., John, B. & Rizzo, A.A. (in press) Sorting out the Virtual Patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations. International Journal of Game and Computer Mediated Simulations.

Talbot, T.B. (in press). Balancing Physiology, Anatomy & Immersion: How Much Biological Fidelity is Necessary in a Medical Simulation? *Journal of Military Medicine.*

Talbot, T.B. & Rizzo, A.A. (2012) *Interview with Dr. Samuel Abrahamson.* Conducted in Los Angeles, CA on May 23rd, 2012.

Tamblyn, R.M., Klass, D.J. & Schnabl, G.K. (2009). The accuracy of standardized patient presentation. *Medical Education* 25(2):100-109.

Tamblyn, R.M., Klass, D.J., Schabl, G.K. & Kopelow, M.L. (1991) Sources of unreliability and bias in standardized-patient rating. *Teaching and Learning in Medicine.* 3(2):74-85.

Tavinor, G. (2009). *The art of videogames: New Directions in Aesthetics.* Wiley-Blackwell, Oxford.

Thiebaux, M., Marshall, A., Marsella, S., & Lee, J. (2008). SmartBody: Behavior Realization for Embodied Conversational Agents. *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).* Portugal.

Traum, D., Marsella, S., Gratch, J., Lee, J. & Hartholt, A. (2008). Multi-party, Multi-issue, Multi-strategy

Negotiation for Multi-modal Virtual Agents. *8th International Conference on Intelligent Virtual Agents*. Tokyo, Japan, Springer.

Traum, D.R., Marsella, S., Gratch, J., Lee, J., Hartholt, A. (2008). Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents. *In Proceedings of the conference on Intelligent Virtual Agents*, p. 117-30.

Triola, M., Feldman, H., Kalet, A.L., Zabar, S., Kachur, E.K., Gillespie, C., Anderson, M., Griesser, C. & Lipkin, M. (2006). A Randomized Trial of Teaching Clinical Skills Using Virtual and Live Standardized Patients. *Journal of General Internal Medicine*, 21:424-426.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, Vol. 59(236), 433-460.

Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9(1):36-45.