

# Chapter 17

## Virtual Human Standardized Patients for Clinical Training



Thomas Talbot and Albert “Skip” Rizzo

A virtual revolution is ongoing in the use of simulation technology for clinical purposes. When discussion of the potential use of Virtual Reality (VR) applications for human research and clinical intervention first emerged in the early-1990s, the technology needed to deliver on this “vision” was not in place. Consequently, during these early years VR suffered from a somewhat imbalanced “expectation-to-delivery” ratio, as most users trying systems during that time will attest. Yet it was during the “computer revolution” in the 1990’s that emerging technologically-driven innovations in behavioral healthcare had begun to be considered and prototyped. Primordial efforts from this period can be seen in early research and development (R&D) that aimed to use computer technology to enhance productivity in patient documentation and record-keeping, to deliver cognitive training and rehabilitation, to improve access to clinical care via internet-based teletherapy, and in the use of VR simulations to deliver exposure therapy for treating specific phobias. Over the last 20 years the technology required deliver behavioral health and medical training applications has significantly matured. This has been especially so for the core technologies needed to create VR systems where advances in the underlying enabling technologies (e.g., computational speed, 3D graphics rendering, audio/visual/haptic displays, user interfaces/tracking, voice recognition, artificial intelligence, and authoring software, etc.) have supported the creation of low-cost, yet sophisticated VR systems capable of running on commodity level personal computers. In part driven by digital gaming and entertainment sectors, and a near insatiable global demand for mobile and networked consumer products, such advances in technological “prowess” and accessibility have provided the hardware and software platforms needed to produce more usable and hi-fidelity VR scenarios for the conduct of human research and clinical intervention. Thus, evolving behavioral

---

T. Talbot (✉) · A. Rizzo

University of Southern California Institute for Creative Technologies, Los Angeles, CA, USA  
e-mail: [talbot@ict.usc.edu](mailto:talbot@ict.usc.edu)

health applications can now usefully leverage the interactive and immersive assets that VR affords as the technology continues to get faster, better and cheaper moving into the twenty-first Century.

While such advances have now allowed for the design and creation of ever more believable context-relevant “structural” VR environments (e.g. combat scenes, homes, classrooms, offices, markets), the next stage in the evolution of Clinical VR will involve populating these environments with Virtual Human (VH) representations that can engage real human users in believable and/or useful interactions. This emerging technological capability has now set the stage for the next major movement in the use of VR for clinical purposes with the “birth” of intelligent VH agents that can serve the role of virtual standardized patients (VSPs) for clinical training. One problem in trying to understand VSPs is that there are several quite distinct educational approaches that are all called a ‘virtual patient.’ Such approaches include case presentations, interactive patient scenarios, virtual patient games, human standardized patients, high fidelity software simulations, high fidelity manikins and virtual human conversational agents. The emphasis of this chapter is on virtual human conversational agents and the reader is referred to Talbot et al., (Adamo 2004) for a very clear detailing of the salient features of the wide variety of approaches that are commonly referred to as virtual patients.

## **The Rationale for Virtual Standardized Patients**

An integral part of medical and psychological clinical education involves training in interviewing skills, symptom/ability assessment, diagnosis and interpersonal communication. In the medical field, students initially learn these skills through a mixture of classroom lectures, observation, and role-playing practice with standardized patients--persons recruited and trained to take on the characteristics of a real patient, thereby affording medical students a realistic opportunity to practice and be evaluated in a simulated clinical environment. This method of clinical training was first attempted in 1963, when Dr. Howard Barrows at the University of Southern California trained the first human standardized patient (Artstein et al. 2008). Since that time, the use of live actors has long been considered as the gold standard medical education experience for both learning and evaluation purposes (Babu et al. 2006; Barrows and Abrahamson 1964). Human standardized patients (HSPs) are paid actors who pretend to be patients for educational interviews and provide the most realistic and challenging experience for those learning the practice of medicine because they most closely approximate a genuine patient encounter. HSPs are also a key component in medical licensing examinations. For example, the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills exam uses SPs and is mandatory for obtaining medical licensure in the United States [cf. <http://www.usmle.org/>]. HSP encounters engage a number of clinical skill domains such as social skills, communication skills, judgment, and diagnostic acumen in a real time setting. All other kinds of practice encounters fall short of

this because they either do not force the learner to combine clinical skill domains or they spoon feed data to the student with the practice case that turns the learning more into a pattern recognition exercise, rather than a realistic clinical problem solving experience. The HSP is the only type of encounter where it is up to the learner to naturalistically pose questions to obtain data and information about the case that then needs to be integrated for the formulation of a diagnostic hypothesis and/or treatment plan.

Despite the well-known superiority of HSPs to other instructional methods (Benedict 2010; Bickmore and Cassell 2005), they are employed sparingly. The reason for this limited use is primarily due to the very high costs to hire, train and maintain a diverse group of patient actors. Moreover, despite the expense of standardized patient programs, the standardized patients themselves are typically low skilled actors and administrators face constant turnover resulting in considerable challenges for maintaining the consistency of diverse patient portrayals for training students. This limits the value of this approach for producing realistic and valid interactions needed for the reliable evaluation and training of novice clinicians. Thus, the diversity of clinical conditions that HSPs can characterize is limited by availability of human actors and their skills. HSPs that are hired may provide sub-optimal variation control and are typically limited to healthy appearing adult encounters. This is even a greater problem when the actor needs to be a child, adolescent, elder, person with a disability or in the portrayal of nuanced or complex symptom presentations.

The situation is even more challenging in the training of students in clinical psychology, social work, and other allied health professions. Rarely are live standardized patients used in such clinical training. Most direct patient interaction skills are acquired via role-playing with supervising clinicians and fellow graduate students, with closely supervised “on-the-job” training providing the brunt of experiential training. While one-way mirrors provide a window for the direct observation of trainees, audio and video recordings of clinical sessions is the more common method of providing supervisors with information on the clinical skills of trainees. However, the imposition of recording has been reported to have demonstrable effects on the therapeutic process that may confound the end goal of clinical training (Bickmore and Giorgino 2006) and the supervisor review of raw recordings is a time consuming process that imposes a significant drain on resources.

In this regard, VSPs can fulfill the role of human standardized patients by simulating diverse varieties of clinical presentations with a high degree of consistency, and sufficient realism (Bickmore et al. 2007; Bitzer 1966), as well as being always available for anytime-anywhere training. Similar to the compelling case made over the years for Clinical VR generally, 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 an ecologically relevant simulated environment. Toward this end, there is a growing literature on the use of VSPs in the testing and training of bioethics, basic patient communication, interactive conversations, history taking, clinical assessment, and clinical decision-making and initial results suggest that VSPs can provide

valid and reliable representations of live patients (Adamo 2004; Bitzer 1966; Bogolub 1986; Campbell et al. 2011; Cheek 2012; Collins and Harden 1999; Cook and Triola 2009; Dev and Heinrichs 2012; Dunne and McDonald 2010).

## 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 a prop to add context or minimal faux interaction in a virtual world, these VH agents are designed to perceive and act in a 3D virtual world, engage in face-to-face spoken dialogues with real users (and other VHs) and in some cases, they are capable of exhibiting human-like emotional reactions. Previous classic work on virtual humans in the computer graphics community focused on perception and action in 3D worlds, but largely ignored dialogue and emotions. This has now changed. 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. 2008a, b), recognize nonverbal cues (Morency et al. 2008), reason about social and emotional factors (Gratch and Marsella 2004) and synthesize human communication and nonverbal expressions (Thiebaut et al. 2008). Such fully embodied conversational characters have been around since the early 90's (Bickmore and 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 and Ishizuka 2004; Rickel et al. 2001); Rizzo et al. 2011a, b), intelligent kiosks (McCauley and 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 anytime-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 and Giorgino 2006; Bickmore et al. 2007; Kenny et al. 2007; Lok et al. 2007; Parsons et al. 2008); Rizzo et al. 2011a, b). 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).

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 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 reports for 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. 2011a, b). 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).

## USC Efforts to Create Virtual Standardized Patients

### *Early Work in Psychiatry*

The USC Institute for Creative Technologies began work in this area in 2007 with an initial project that involved the creation of a virtual patient, named “Justin” (see Fig. 17.1]. Justin portrayed a 16-year old male with a conduct disorder who was being forced to participate in therapy by his family. The system was designed for novice clinicians to practice asking interview questions, to attempt to create a positive therapeutic alliance and to gather clinical information from this very challenging VSP. Justin was designed as a first step in our research. At the time, the project was unfunded and thus required our lab to take the economically inspired route of recycling a virtual character from a military negotiation-training scenario to play the part of Justin. The research group agreed that this sort of patient was one that could be convincingly created within the limits of the technology (and funding) available to us at the time. For example, such resistant patients typically respond slowly to therapist questions and often use a limited and highly stereotyped vocabulary. This



**Fig. 17.1** “Justin”

allowed us to create a believable VSP within limited resources for dialog development. As well, novice clinicians have been typically observed to have a difficult time learning the value of “waiting out” periods of silence and non-participation with these patients. The system used voice recognition technology to translate speech to text, upon which the system would match questions to a limited bank of VSP responses. We initially collected user interaction and dialog data from a small sample of psychiatric residents and psychology graduate students as part of our iterative design process to evolve this application area. The project produced a successful proof of concept demonstrator and generated interest in the local medical community at Keck School of Medicine at USC that subsequently led to the acquisition of funding that supported the development of our next VSP.

Following the Justin proof of concept, our 2nd VSP project involved the creation of a teenage female sexual assault victim, “Justina” to more formally assess student views towards interacting with a VSP in a training context (see Fig. 17.2). We also aimed to explore the potential for creating a clinical interview trainer that could evaluate students in terms of their ability to ask questions relevant for assessing whether Justina met the criteria for the DSM-4r diagnosis of PTSD based on symptoms reported during the clinical interview. The interaction were also informally reviewed to get a sense as to whether students would interact with the VSP in a “sensitive” fashion as one would expect with a real-life clinical interaction with someone who had experienced significant personal trauma.



**Fig. 17.2** “Justina”

For the PTSD content domain, 459 questions were created that mapped roughly 4–1 to a set of 116 responses. The aim was to build an initial language domain corpus generated from subject matter experts and then capture novel questions from a pilot group of users (psychiatry residents) during interviews with Justina. The novel questions that were generated could then be fed into the system in order to iteratively build the language corpus. We also focused on how well subjects asked questions that covered the six major symptom clusters that can characterize PTSD following a traumatic event. While this approach did not give the Justina character a lot of depth, it did provide more breadth for PTSD-related responses, which for initial testing seemed prudent for generating a wide variety of questions for the next Justina iteration.

In the initial test, 15 Psychiatry residents (6 females, 9 males, mean age = 29.80, SD 3.67) participated in the study and were asked to perform a 15 min interaction with the VSP to take an initial history and determine a preliminary diagnosis based on this brief interaction with the character. The participants were instructed to speak normally, as they would to a live standardized patient, but were informed that the system was a research prototype that uses an experimental speech recognition system that would sometimes not understand them. They were instructed that they were free to ask any kind of question relative to a clinical interview and the system would try to respond appropriately, but if it didn't they could ask the same question in a different way.

From post questionnaire ratings on a 7-point Likert scale, the average subject rating for believability of the system was 4.5. Subjects reported their ability to understand the patient at an average of 5.1, but rated the system at 5.3 as frustrating to talk to due to speech recognition problems, out of domain answers or inappropriate responses. However most of the participants left favorable comments that they thought this technology will be useful in the future, and that they enjoyed the experience of trying different ways to talk to the character in order to elicit an relevant response to a complex question. When the patient responded back appropriately to a question, test subjects informally reported that the experience was very satisfying. Analysis of concordance between user questions and VSP response pairs indicated moderate effects sizes for Trauma inquiries ( $r = 0.45$ ), Re-experiencing symptoms ( $r = 0.55$ ), Avoidance ( $r = 0.35$ ), and in the non-PTSD general communication category ( $r = 0.56$ ), but only small effects were found for Arousal/Hypervigilance ( $r = 0.13$ ) and Life impact ( $r = 0.13$ ). These relationships between questions asked by a novice clinician and concordant replies from the VSP suggest that a fluid interaction was sometimes present in terms of rapport, discussion of the traumatic event, the experience of intrusive recollections and discussion related to the issue of avoidance. Low concordance rates on the arousal and life impact criteria indicated that a larger domain of possible questions and answers for these areas was not adequately modeled in this pilot effort.

### *Social Work Standardized Virtual Patients*

The next USC VSP project involved collaboration with the USC School of Social Work, Center for Innovation in Research (CIR). This MSW program is novel for its focus on preparing social workers for careers working with military Service Members, Veterans and their families. This project resulted in the creation of a VSP named “Sgt. Castillo” (see Figs. 17.3a and b) designed to help social work trainees gain practical training experiences with VSPs that portray behavior more relevant to military culture and common clinical conditions that Service Members and Veterans experience. This work also supported our first effort to create a limited authoring system that would allow for the creation of new VSP dialog that would support the flexible modification of the training goals. The vision was to build an interface that allowed clinical educators to create a virtual patient with the same ease as creating a PowerPoint presentation. If such authoring could be done by clinical educators, it would be possible for subject matter experts (social work educators in this case) to create VSPs that could represent a wide range of clinical conditions with the ability to manipulate the intensity and complexity of the clinical presentation and subsequent training challenge. Unfortunately, the resulting authoring system was somewhat difficult to learn without a deeper understanding of dialog management. Consequently, the authoring system was poorly adopted by our collaborators in social work and only a few VSP instantiations were created. A sample video



**Fig. 17.3** a. Sgt. Castillo military VSP. b. in use, projected on wall with trainee

of a social work trainee interviewing one of these military VSPs can be found here: <http://www.youtube.com/watch?v=PPbc18Z-8Ec>.

In view of these difficulties with authoring, the ICT/CIR project changed direction in order to meet the immediate need to provide clinical training to social work students currently enrolled in the CIR program. Instead of a focus on authoring and modification of the characteristics of the VSP, the emphasis shifted to training a specific psychotherapeutic approach that could involve concurrent individual and

group/classroom practice. This resulted in the development of the Motivational Interviewing Learning Environment and Simulation (MILES) to provide future social workers with the opportunity to practice Motivational Interviewing (MI) skills in a mixed-reality setting with a VSP. MILES was designed as an instructor-facilitated experience that enables an individual student to practice a MI-oriented interaction with a military veteran VSP while a classroom of students observes real time video of the student/client interaction. The individual student trainee “speaks” to the virtual human through a microphone, selecting what he or she says from a multiple-choice list of carefully constructed statements. The MILES VSP (see Fig. 17.4) has the ability to understand the spoken dialog and responds to the student in a lifelike, natural manner with realistic voice, body language, gestures and facial expressions. As the single student progresses through the scenario, a branching dialog system can lead to various successful and unsuccessful outcomes depending on the response options selected by the individual trainee. At the same time, the rest of the class follows along viewing the real time video and selects their choice of the dialog options at each interaction juncture via individual response “clickers”. An instructor control station captures performance data, including the answers selected by the lone student and their fellow classmates, to support of instructor awareness of the class’s knowledge status to facilitate feedback in the form of an After Action Review (AAR) following an interaction. This system is currently in classroom use and learning evaluations are ongoing. A sample video of the MILES project can be found here: <http://www.youtube.com/watch?v=Sg8x1rttBho&feature=youtu.be>



Fig. 17.4 MILES virtual patient

## Standardized Virtual Patients for Medical Training

After a number of prototypes and experiments conducted by the Authors and elsewhere, it had become clear that a plateau had been reached in VSP applications and technology that left progress short of the threshold required for broader adoption of interactive conversational characters for training. The primary factors limiting further improvement in experimental VSP systems were many, with the primary cause being the considerable effort required to create a single VSP encounter. Generally, it required a team of experts about 6 months to create a VSP, including up to 200 h of expert language training (Kenny et al. 2010). Additional factors included the low performance of natural language understanding (NLU) systems needed to understand the learner’s questions and the effort involved in animating, creating voices, lip syncing and scheduling motion of a virtual human avatar. The Justina prototype had a maximum NLU accuracy of 60% (Kenny et al. 2007), with other systems achieving about 75%. That level of performance resulted in frustrating encounters, whereas NLU accuracy nearing 90% is more likely to result in a more positively received interaction that flows well as a clinical interview.

One strategy around the NLU accuracy problem is to avoid NLU altogether. Virtual human conversations are possible that include an avatar that responds to pre-selected choices; such an interview is called a structured encounter. There are many kinds of structured encounters. They may be linear, branching, unlocking style and state-machine/logic-based. Structured encounters can be employed for patient interviews, surrogate interviews, counseling sessions, difficult conversations, persuasive conversations and many other purposes (Figs. 17.5 and 17.6). Learner choices are

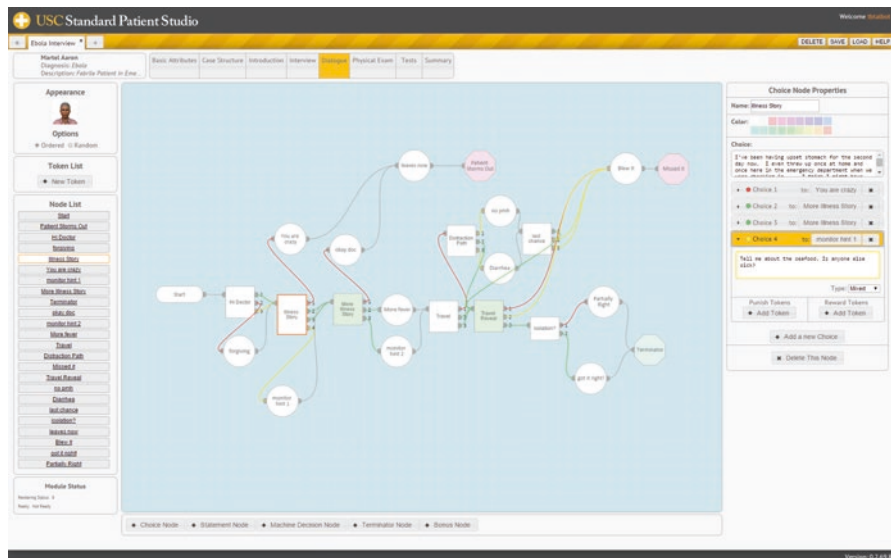
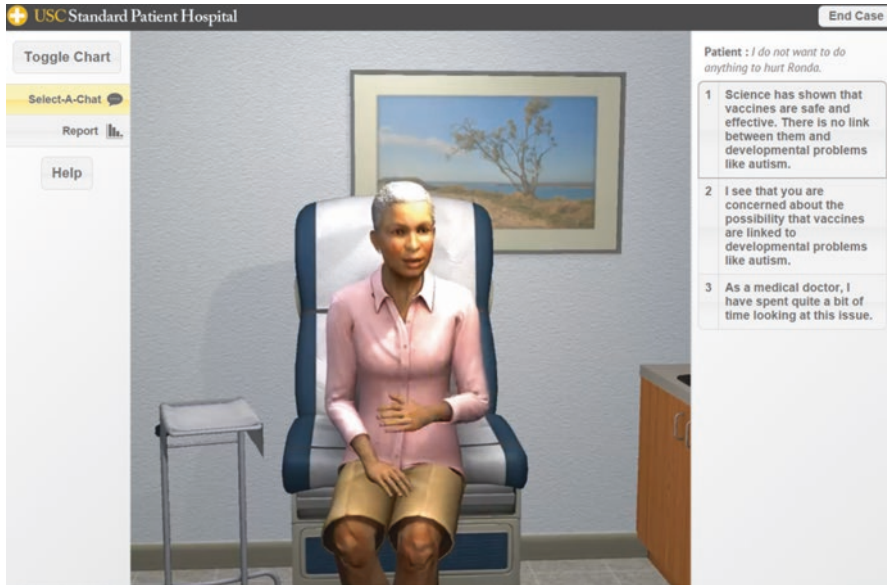


Fig. 17.5 USC Standard Patient “Select-a-Chat” structured virtual human encounter authoring tool



**Fig. 17.6** A structured virtual human encounter depicting a vaccine resistant parent. (USC Standard Patient)

definite and appropriate responses are guaranteed. Assessments are based on accurate data and have no potential for assessment bias.

The use of structured virtual humans for training is established; it has been successfully integrated into routine training with the previously mentioned MILES being an example. Another MILES variant, ELITE Lite has been accredited by the US Army for training. According to the accreditation document (Lamb et al. 2007), ELITE Lite survey feedback reported 88.7% of respondents indicated practice exercises provide a sufficient representation of an informal interaction between a counselor and counselee. Subjects (87%) indicated the training experience was engaging and effective while 77% indicated they have a better understanding of the counseling process after using ELITE Lite. Most users indicated they would rather use ELITE Lite vs. lecture and PowerPoint instructional method (85%).

Another compelling structured encounter prototype is Virtual Child Witness (VCW—Fig. 17.7). VCW is a structured virtual human encounter intended to assess forensic interviewing skills. This effort focused on questioning strategy and compared “experts”, a group of professionals who completed a forensic interviewing course with novices. The study, designed to see if the virtual human encounters could be an effective assessment tool, showed significantly higher performance in the expert group compared with novices. Analysis of the study data also revealed a strong training effect with subjects who unexpectedly played the structured encounter multiple times (Leuski et al. 2006). Of interest, VCW was created with very small budget on the SimCoach virtual human platform. SimCoach shortened



**Fig. 17.7** Virtual Child Witness—a structured encounter

the development time because it handled all the tasks required to create animated virtual humans and provided an online delivery mechanism (Bitzer 1966).

Although structured encounters are a useful tool for many training applications, there is still a desire to simulate the medical interview with a VSP. The expense and limited access to human standardized patients coupled with the potential for objective assessments and repeatable, low cost encounters makes a compelling case for the success of VSPs. Fortunately, recent technology advances have succeeded in breaking the VSP plateau to the point where the major problems inhibiting VSP creation and adoption are being addressed.

The USC Standard Patient (USP) project is a freeware open-source VSP community ([www.standardpatient.org](http://www.standardpatient.org)) that has applied considerable resources to improving natural language random access (NLRA) VSPs – the kind that mimic typical conversations with human patients (Fig. 17.8). The improvements (Leuski and Traum 2010) include creation of an automated online virtual human tool, an improved medical NLU system, a universal VSP taxonomy, and a new approach to assessing human-computer conversations.

An automated online virtual human tool, SimCoach, was created first. SimCoach enables the rapid creation of cloud-based online virtual humans. SimCoach VSPs work on current-generation web browsers and greatly simplify the development burden for virtual human (VH) creation. SimCoach automates speech actions,



**Fig. 17.8** Natural language capable VSPs permit learners to ask questions in a natural manner through speech or typed input. (USC Standard Patient)

animation sequencing, lip synching, non-verbal behavior, NLU integration, and AI processing and interaction management. With assets in place, new VHs can be created by providing text content. SimCoach was initially employed for training for VCW and is now the virtual human technology platform for USP.

The next impediment to be addressed is the fact that most prior NLRA VSPs were authored by creating a language focus around a specific medical problem or diagnosis. Questions would be compiled and answers associated to create a case that receives training data. This labor-intensive process needed for every patient case. Additionally, off-topic questions were poorly handled and caused such VSPs to appear inflexible. The Standard Patient project adopted a unified medical taxonomy (UMT) instead. UMT provides a common patient description regardless of actual patient condition. This makes new patient cases much more easily authorable and provides a fixed NLU training domain. Every Standard Patient VSP is represented by the complete unified taxonomy. Baseline and non-authored case elements are filled in by the UMT system based on age/sex appropriate default responses.

NLU, one main impediment to fluent learner-patient interactions, was addressed through the creation of a new medical NLU system called LEXI Mark 1. LEXI is a vastly improved NLU system specifically developed for medical interactions. The system is closely tied to the UMT and includes lexical assessment, probabilistic

modeling and content matching approaches. Lexi is capable of improving performance through human-assisted and machine learning. The implication of an approach that trains the NLU for the UMT rather than a specific case is that NLU training affects improvements in all cases on the system. Lexi has demonstrated better than 92% NLU accuracy in testing with a well-trained taxonomy under training conditions.

A new approach to conversational assessment, INFERENCE-RTS, was then developed. INFERENCE is an advanced game-based assessment engine that is capable of analyzing human conversations in real-time and associating learner speech acts with effects on the UMT. With this system, case authors annotate patient utterances in the case-authoring tool with assessment tags. Such tags are employed to indicate information that is of critical importance or moderate importance to the diagnosis. Tags exist for every UMT taxonomy item. The feedback intervention system encapsulates diagnostic performance and provides learners with concrete improvement tasks, a mind-map case taxonomy visualization (Fig. 17.9) and a learning-curve tool. INFERENCE was designed for deliberate practice at the proximal level of learner development. Future research will establish if such a system is practical and efficacious.

Tagging the universal taxonomy and providing feedback in the form of the mind-map has finally provided a workable solution to automating the assessment of conversational interviews. Such automated assessments can be accurate within a few percent. The use of feedback from this type of display has been employed and has demonstrated a strong training effect in real world use.

The combined effect of all these recent improvements results in a practical system that maintains ease of use, allows content creation in a timely manner, and

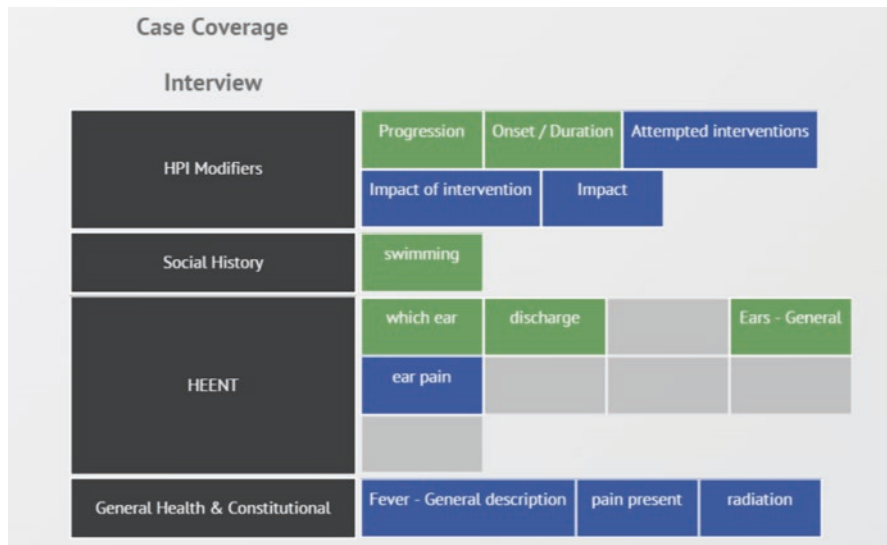


Fig. 17.9 VSP interview mind-map

provides practical assessment feedback to learners and educators. Researchers have yet to conduct the necessary validations to determine the educational impact of VSP systems that employ combinations of these recent advances. In the near future, this information will be available and will determine the next course of action to advance VSPs for medical and psychological education purposes. If these combined technologies prove efficacious, it will be of great interest to see how this influences the milieu of medical and professional training.

Most VSPs attempted to date have been on traditional computers. With the increased prevalence of mobile devices, it is logical to consider the migration of VSP technology to phones and tablets. Regardless, there are significant usability barriers to adoption of VSPs on mobile platforms. The limitations are more human factors-based rather than caused by technical limitations. For example, how will a person interact with a conversational VSP? Will people talk to their phones? Will people type on tablet screens? Computers have excellent keyboards and when speech recognition is performed, this is usually with the benefit of a headset microphone to isolate speech. Phone and tablet microphones capture surrounding sound and this may result in too many speech recognition errors. It may also present a more awkward interaction. Structured encounter-style VSPs do not suffer from these limitations and are much more readily adaptable to mobile device adoption.

Another promising idea is to imbue a manikin or task trainer with VSP capabilities. Such a capability could greatly improve the interactive potential of plastic-based physical training systems. The main technical limitation is similar to the mobile device problem; voice recognition. Specifically, voice recognition system in robots will have to work at a distance. Future distant recognition systems (DSR) will require a high level of individual speaker discrimination and will likely adopt microphone array-based acoustic beamforming technology. (Lok et al. 2007) Unfortunately, DSR technology is not yet at a sufficient level of maturity for effective use with VSPs.

## Conclusion

Virtual reality standardized patients have come a long way from faux-interactions on time-sharing mainframes starting half a century ago. Work over the last 15 years; in particular, has produced a wealth of knowledge and practical lessons in both the advance of VSP technology as well as experience with VSPs in clinical training applications. Despite these advances, VSPs have yet to see mainstream adoption in clinical training for a number of reasons. Recent work appears to have advanced sufficiently to ameliorate or overcome the most significant barriers. Thus, the age where VSPs may play a major role in training may finally be upon us. Future success may no longer be rate-limited by the pace of technology, but by the creativity and innovation of educators who will create compelling VSP experiences and curricula.

## References

- Adamo, G. (2004). Simulated and standardised patients in OSCEs: Achievements and challenges 2–2003. *Medical Teacher*, 25(3), 262–270. <https://doi.org/10.1080/0142159031000100300>.
- Artstein, R., Gandhe, S., Leuski, A., & Traum, D. R. (2008). Field testing of an interactive question-answering character. In *Proceedings of the European language resources association workshop on evaluation* (pp. 36–40). Marrakech.
- Babu, S., Schmutge, 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.), *IWA 2006* (pp. 169–180). Berlin: Springer-Verlag. [https://doi.org/10.1007/11821830\\_14](https://doi.org/10.1007/11821830_14).
- 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.
- Benedict, N. (2010). Virtual patients and problem-based learning in advanced therapeutics. *American Journal of Pharmaceutical Education*, 74(8), 143–148. <https://doi.org/10.5688/aj7408143>.
- 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. [https://doi.org/10.1007/1-4020-3933-6\\_2](https://doi.org/10.1007/1-4020-3933-6_2).
- Bickmore, T., & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Bio-medical Informatics*, 39(5), 556–571. <https://doi.org/10.1016/j.jbi.2005.12.004>.
- Bickmore, T., Pfeifer, L., & Paasche-Orlow, M. (2007). Health document explanation by virtual agents. In *Proceedings of the intelligent virtual agents conference*, Paris. Springer.
- Bitzer, M. (1966). Clinical nursing instruction via the PLATO simulated laboratory. *Nursing Research*, 15, 144–150. <https://doi.org/10.1097/00006199-196601520-00009>.
- Bogolub, E. B. (1986). Tape recorders in clinical sessions: Deliberate and fortuitous effects. *Clinical Social Work Journal*, 14(4), 349–360. <https://doi.org/10.1007/BF01892595>.
- Campbell, J., Core, M., Artstein, R., Armstrong, L., Hartholt, A., Wilson, C. et al. (2011, March 14–12). Developing INOTS to support interpersonal skills practice. IEEE aerospace conference, big sky, MT.
- Cheek, W. (2012) Direct communications with breakaway ltd. via written survey. Retrieved from <http://www.breakawayltd.com>.
- 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 Dundee. Retrieved Jan. 4, 2011, from [http://www.medev.ac.uk/resources/features/AMEE\\_summaries/Guide13summaryMay04.pdf](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. <https://doi.org/10.1111/j.1365-2923.2008.03286.x>.
- Dev, P., & Heinrichs, W. (2012) Direct communications with clinispace/innovations in learning via written survey. Retrieved from <https://www.clinispace.com>.
- Dunne, J. R., & McDonald, C. L. (2010). Pulse!!: A model for research and development of virtual-reality learning in military medical education and training. *Military Medicine*, 175(s1), 25–27.
- Evans, D., Hearn, M. T., Uhlemann, M. R., & Ivey, A. E. (1989). *Essential interviewing: A programmed approach to effective communication* (3rd ed.). Pacific Brooks: Brooks/Cole Publishing Company.
- Gratch, J., & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4), 269–306. <https://doi.org/10.1016/j.cogsys.2004.02.002>.
- Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17(4), 54–61. <https://doi.org/10.1109/MIS.2002.1024753>.
- 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*, 5, 81–89.

- Kenny, P., Parsond, T., & Garrity, P. (2010, November 29–December 2). Virtual patients for virtual sick call medical training. Interservice/Industry training, simulation, and education conference, Orlando.
- 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*(11-39), 1201–1231. <https://doi.org/10.1016/j.chiabu.2007.03.021>.
- Leuski, A., & Traum, D. R. (2010, July 11–15). Practical language processing for virtual humans. In Proceedings of the twenty-second annual conference on innovative applications of artificial intelligence (IAAI-10), Atlanta.
- Leuski, A., Kennedy, B., Patel, R., & Traum, D. R. (2006, November 27-30). Asking questions to limited domain virtual characters: How good does speech recognition have to be? In Proceedings of the 25th army science conference, Orlando.
- Lok, B., Ferdig, R. E., Raij, A., Johnson, K., Dickerson, R., & Coutts, J. (2007). Applying virtual reality in medical communication education: Current findings and potential teaching and learning benefits of immersive virtual patients. *Journal of Virtual Reality*, *43*(3–4), 185–195. <https://doi.org/10.1007/s10055-006-0037-3>.
- McCauley, L., & D’Mello, S. (2006, August 21–23). A speech enabled intelligent kiosk. In J. Gratch et al. (Eds.), In Proceedings of the 6th annual intelligent virtual agents conference, Marina del Ray, CA (pp. 47–144). Berlin: Springer.
- Morbini, F., DeVault, D., Sagae, K., Nazarian, A., Gerten, J., & Traum, D. R. (2013). FLoReS: A forward looking, reward seeking, dialogue manager. Submitted to the 13th annual SIGDIAL meeting on discourse and dialogue, Seoul.
- Morency, L.-P., de Kok, I., & Gratch, J. (2008, October 20–22). Context-based recognition during human interactions: Automatic feature selection and encoding dictionary. In Proceedings of the 10th international conference on multimodal interfaces, Chania.
- Parsons, T. D., Kenny, P., Ntuen, C. A., Pataki, C. S., Pato, M. T., & Rizzo, A. A. (2008). Objective structured clinical interview training using a virtual human patient. *Studies in Health Technology and Informatics*, *132*, 357–362.
- Prendinger, H., & Ishizuka, M. (2004). *Life-like characters – Tools, affective functions, and applications*. Berlin: Springer. <https://doi.org/10.1007/978-3-662-08373-4>.
- Rickel, J., Gratch, J., Hill, R., Marsella, S., & Swartout, W. (2001, March 25–27). Steve goes to bosnia: Towards a new generation of virtual humans for interactive experiences. In Proceedings of the AAAI spring symposium on AI and interactive entertainment, Stanford University.
- Rizzo, A. A., Kenny, P., & Parsons, T. (2011a). Intelligent virtual humans for clinical training. *International Journal of Virtual Reality and Broadcasting*, *8*(3). Retrieved from <http://www.jvrb.org/8.2011/>.
- Rizzo, A. A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., & Sagae, K. (2011b). An intelligent virtual human system for providing healthcare information and support. *Studies in Health Technology and Informatics*, *163*, 503–509.
- Rossen, B., Cendan, J., & Lok, B. (2010, September 20-22). Using virtual humans to bootstrap the creation of other virtual humans. In Allbeck et al (Ed.), Proceedings of the international conference on virtual agents 2010, Philadelphia (LNAI6356, pp. 392–398).
- Stevens, A., Hernandez, J., & Johnsen, K. (2005). The use of virtual patients to teach medical students communication skills. *American Journal of Surgery*, *191*, 806–811. <https://doi.org/10.1016/j.amjsurg.2006.03.002>.
- Swartout, W., Gratch, J., Hill, A. W., Marsella, S., Rickel, J., & Traum, D. (2006). Toward virtual humans. *AI Magazine*, *27*(1), 96–108.
- Swartout, W., Traum, D. R., Artstein, R., Noren, R., Debevec, P., Bronnenkant, K., et al. (2010). Ada and Grace: Toward realistic and engaging virtual museum guides. *Intelligent Virtual Agents*, 286–300.
- Talbot, T. B. (2013). Balancing physiology, anatomy & immersion: How much biological fidelity is necessary in a medical simulation? *Journal of Military Medicine*.

- Talbot, T. B., Sagae, K., John, B., & Rizzo, A. A. (2012). 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 Gaming and Computer Mediated Simulations.*, 4(4), 1–19.
- Thiebaut, M., Marshall, A., Marsella, S., Fast, E., Hill, A., Kallmann, M., et al. (2008, May 12-73) Smart body: Behavior realization for embodied conversational agents. In Proceedings of the Inter-national conference on autonomous agents and multi-agent systems (AAMAS), Estoril.
- Traum, D. R., Marsella, S., Gratch, J., Lee, J., & Hartholt, A. (2008a, September 1–3). Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In Proceedings of the 8th international conference on intelligent virtual agents, Tokyo.
- Traum, D. R., Marsella, S., Gratch, J., Lee, J., & Hartholt, A. (2008b). Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In Proceedings of the conference on intelligent virtual agents (pp. 117–130).
- Triola, M., Feldman, H., Kalet, A. L., Zabar, S., Kachur, E. K., & Gillespie, C. (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of General Internal Medicine*, 21, 424–426. <https://doi.org/10.1111/j.1525-1497.2006.00421.x>.