Work-in-Progress—Testing of a Virtual Patient: Linguistic and Display Engagement Findings

Fei Chen
Department of Anesthesiology
University of North Carolina
Chapel Hill, NC, USA
fei_chen@med.unc.edu

Yueh Lee
Department of Radiology
University of North Carolina
Chapel Hill, NC, USA
yueh_lee@med.unc.edu

Robert Hubal
Eshelman School of Pharmacy
University of North Carolina
Chapel Hill, NC, USA
hubal@unc.edu

***Abstract*—This work-in-progress paper reports the development and testing of an innovative virtual patient simulation application for medical education. The goal was to make the conversation-based simulation experience both natural and authentic, with a focus on the language processing and visualization components. Metrics measuring student engagement and session flow were collected from 115 sessions completed by 57 participants. Preliminary analysis found average conversational turn time longer during development compared to testing (p=0.044) and greater number of topic matches for participants viewing the patient on a holographic display compared to participants viewing the patient on a monitor (p=0.004).**

***Index Terms*—virtual patient, engagement, natural language, display technology**

I. Introduction

An ongoing shift toward managed care and community-based medicine, shorter hospital stays, and a new emphasis on preventive medicine requires increased levels of interpersonal competency in students [1], [2]. In healthcare, nearly all medical training centers integrate role-play with standardized patients as part of a problem-based medical curriculum [3]. These activities are a means to allow nascent clinicians to become more comfortable with screening questions and interviewing techniques. Among many training techniques is the use of simulated encounters with standardized patients. A subset is use of virtual patients using responsive virtual humans.

Virtual humans simulate physical, psychological, linguistic, social, and other behavioral qualities of people in computer-based environments [4]–[6]. As role-players—specifically for this paper, as virtual standardized patients—virtual humans give students authentic practice of target skills and feedback on their performance within a representative sample of realistic situations [7]–[10].

There are, however, at least two limitations to most virtual patient efforts. First, the realism of behaviors exhibited by virtual patients is still lacking. As with virtual human behaviors in games and other training simulations, the most effective training requires appropriate and consequential responses exhibited by virtual humans given students' activity (or inactivity). None of the virtual patients created to date, for instance, have life histories; in current terminology they do not exhibit realistic patterns of life (POL [11]; see also [12]). Virtual patients should have family, social networks, economic resources, intentions, and other goals—a 'backstory' that could influence its behavior in the moment. Second, the user interface and restrictions imposed by standard interactions in a virtual environment are somewhat at odds with the realistic kinetic nature of interacting with patients [13]. Typical human-machine interaction is through point-and-click interfaces that require somewhat forced means of engagement with virtual patients and that themselves can compete with the portrayed situation for the user's attention. For medical simulation to reach its potential, users must be able to exert natural control, engaging with a realistic character portrayed using engaging means. In this experiment, we aimed to develop and test an innovative application that addresses these two limitations in virtual patient simulations.

II. Methods

A. Procedures

The experiment was conducted in a quiet office setting. At the beginning of each experimental session the experimenter presented the purpose of the experiment to the participant of the session. The participant sat in front of a screen during the entire interaction with the virtual patient. For some participants the patient was projected onto a 60" monitor mounted about head height on a wall approximately ten feet from the participant. For other participants the patient was portrayed using a 15.6" lightfield display, with a holographic projection from Looking Glass Factory on a desk in front of the participant. (See Figure 1.) The virtual environment included a clinic room and virtual patient, preceded by front-end screens that presented background information on the patient such as is found on an electronic health record.

B. Application

The application is built so that the participant converses with the patient. In an approach similar to that adopted by other researchers (e.g., [14]), the application shows a dynamic and randomly-ordered menu of available topics for the participant to address; on average, at each conversational turn, there are some five choices. Though selectable, these menu options are meant to guide the participant through the conversation, and thus only present a brief description of the available topics. For instance, a typical set of options for the participant at the beginning stages of the interaction might include "Ask permission to counsel", "Confirm identity", "Ask how patient is doing", and "Ask about purpose of visit". Each topic enters into a series of subtopics, and once a topic is exhausted it is no longer presented as a menu option.

An important difference from other studies in how to manage the encounter is in how the application captures input from the participant. First, it is free speech; the participant is free to ask

978-1-7348995-0-4/20/$31.00 ©2020 Immersive Learning Research Network

how the patient is doing using any phrasing s/he feels comfortable with. The transcription of the input—that is, the determination of intent of the input—is performed by calls to a server; several of the major services were tried, from IBM to Google, and all yield recognition in the 90-95% range. Second, the intents are mapped using a comprehensive database created from past efforts and from extensive pretesting. In the mapping, the transcription is compared to all exemplars of the different available topics, and the best match is returned. The matching is fuzzy, in that the mapping allows for extra words, different ordering, synonyms, and placeholders such as proper names. However, it is both efficient and effective, in that there are a limited number of exemplars to consider at any step in the conversation because it is expected that the participant will address only of the available topics. On the occasion that there is no threshold match the virtual patient is instructed either to ask the participant to repeat the question, or to remain silent (based on separately-defined rules that govern behavior). Third, in line with past work focused not only on the content of the input but also its valence [15], the intent mapping captures values regarding the input such as politeness and sentiment, and—to improve matching—it also takes into account past actions such as having introduced oneself or not.

Participants were asked to run through multiple encounters with the same patient, but at different ages (there were seven ages available; the patient backstory evolved over the course of the available encounters). A run through a single encounter is called a session. After the encounters, the experimenter engaged the participant in an open-ended discussion of her/his experience.

C. *Participants and Design*

Participants were a convenience, volunteer sample of pharmacy and medical students.

The formal development of the virtual patient architecture ended during the course of having participants run through encounters. The major component of development at that stage was extension and refinement of the natural language processing. Sessions were thus separately analyzed for those run prior to the end of development on the project (development sessions) and those run after the end of development (testing sessions). The development sessions were those used to refine grammars that help the system interpret intention, as well as change flow of the dialog based on unanticipated questions or responses uttered by the participant. Testing sessions were those used to assess the performance of the system.

D. *Analysis*

Descriptive statistics of metrics measuring engagement and session flow (i.e., session time, number of conversational turns, turn time, and intent topic match) collected from participants' interactions with the virtual patients were summarized. Two sample *t*-tests were conducted to compare the metrics between development and testing sessions, as well as between large monitor and holographic display conditions.

III. Results

Testing extended from April 2019 into March 2020, all in-person. A total of 57 participants collectively completed 115

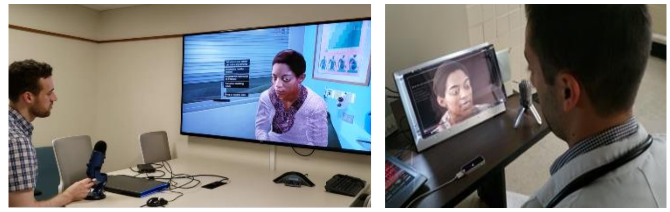

Fig.1. Two versions of display, widescreen (left) and holographic (right).

encounters (i.e., sessions), either during development or during testing. The minimum time for a session to be analyzed was two minutes and at least five conversational turns.

Overall average session time was over 20 minutes, suggesting that participants engaged in the discussion with the patient. However, there was a wide range. Across all sessions participants covered nearly 40% of all topics available, indicating a wide range of discussion although not nearly as broad as possible. Part of the reason for this finding is the skew in the number of matches per topic: Nearly all participants asked basic questions such as "What brings you in today?" and "Can you confirm your name?", but for most questions only a few participants broached the topic.

14% of the time every single turn matched a topic; 23% of the time this was nearly true (at least 90% of all turns had a match). 12% of the time there were more mismatches than matches, although the majority of such sessions had relatively few conversational turns. In other words, as the conversational length grew, so did the accuracy of the dialog exchange.

Analyzing via *t*-tests (Table I), run time was shorter during development compared to testing, though not significantly so, but the average conversational turn time was longer (p=0.044) and the percent of matches generally fewer (p=0.081) during development compared to testing. These data suggest that during testing participants ran through more of the dialog, in turn suggesting that the system was able to maintain their engagement.

One more set of between-subjects analyses were undertaken, comparing participants who ran through encounters with the virtual patient portrayed on a large monitor versus those who ran through encounters with the virtual patient portrayed on a holographic display (Table II). Though there were few of the latter runs, results come close to indicating a greater number of conversational turns (p=0.075), shorter turn time (p=0.066), and greater number of matches (p=0.004) for the participants viewing the patient on a holographic display compared to participants viewing the patient on a monitor.

IV. Discussion

This study is one of many to look into virtual patients as a means for students to practice patient care. In both development and testing, no major technical limitations prevented the successful completion of sessions in all of the students, suggesting the reliability of the simulation implementation and workflow. Although there was not a match with the effectiveness of intent matching reported elsewhere [16], [17], the performance was satisfactory, and indicative of predicted differences found when comparing development versus testing.

TABLE I. SESSION CHARACTERISTICS, DEVELOPMENT STATUS
(IN DEVELOPMENT VS. TESTING)

Measure	Development Sessions				
	#	*%*	*Average*	*Min*	*Max*
Participants	45				
Sessions	101				
Matches		71%			
Session time (min)			29.1	2	41.2
Turn time (sec)			30.3	10	77.8
# Turns			18.3	5	116

Measure	Testing Sessions				
	#	*%*	*Average*	*Min*	*Max*
Participants	12				
Sessions	14				
Matches		86%			
Session time (min)			43.6	2.4	17
Turn time (sec)			23.0	13.5	46.8
# Turns			20.2	6	59

TABLE II. SESSION CHARACTERISTICS, DISPLAY TYPE
(MONITOR VS. HOLOGRAPHIC PROJECTION)

Measure	Monitor Sessions				
	#	*%*	*Average*	*Min*	*Max*
Participants	52				
Sessions	109				
Matches		72%			
Session time (min)			21.0	2	41.2
Turn time (sec)			29.9	10	77.8
# Turns			17.9	5	116

Measure	Holographic Sessions				
	#	*%*	*Average*	*Min*	*Max*
Participants	5				
Sessions	6				
Matches		92%			
Session time (min)			13.9	2.4	17
Turn time (sec)			20.5	13.5	46.8
# Turns			29.3	6	59

The preliminary data from the holographic display are interesting, hinting that the kind of immersion afforded by a three-dimensional presentation might better engage participants in dialog than traditional presentation [18], [19]—though it must be stressed there were few participants in the holographic condition. The authors plan to engage additional participants in studies using newly designed scenarios to further understand costs and benefits of engagement through natural language and different displays.

REFERENCES

[1] J. E. McLaughlin, M. J. Dean, R. J. Mumper, R. A. Blouin, and M. T. Roth, "A roadmap for educational research in pharmacy," *Am J Pharm Educ*, vol. 77, issue 10, no. 218, Dec. 2013, doi: 10.5688/ajpe7710218.

[2] M. Adamus and M. Jaworski, "Psychological knowledge and skills in clinical practice among selected medical professions," *Eur J Soc Behav Sci*, vol. XI, no. 5, pp. 1602-1611, Nov. 2014, doi: 10.15405/ejsbs.143.

[3] D. Nestel and T. Tierney, "Role-play for medical students learning about communication: Guidelines for maximising benefits," *BMC Med Educ*, vol. 7, no. 3, Mar. 2007, doi10.1186/1472-6920-7-3.

[4] W. L. Johnson, J. W. Rickel, and J. C. Lester, "Animated pedagogical agents: Face-to-face interaction in interactive learning environments," *Int J Artif Intell in Educ*, vol. 11, pp. 47-78, 2000.

[5] R. Hubal, P. Kizakevich, and R. Furberg, "Synthetic characters in health-related applications," in *Advanc Comput Intell Paradigm Healthc-2. Stud Comput Intell*, S. Vaidya, L. C. Jain, and H. Yoshida, Eds., Berlin, Heidelberg: Springer, 2007, vol. 65, pp. 5-26, doi: 10.1007/978-3-540-72375-2_2.

[6] W. L. Johnson and J. C. Lester, "Face-to-face interaction with pedagogical agents, twenty years later," *Int J Artif Intell in Educ*, vol. 26, pp. 25-36, 2016.

[7] R. C. Hubal, P. N. Kizakevich, C. I. Guinn, K. D. Merino, and S. L. West, "The virtual standardized patient–Simulated patient-practitioner dialogue for patient interview training," *Stud Health Technol Inform*, vol. 70, pp. 133-138, 2000, doi: 10.3233/978-1-60750-914-1-133.

[8] P. Kenny, T. D. Parsons, J. Gratch, A. Leuski, and A. A. Rizzo, "Virtual patients for clinical therapist skills training," *Intell Virtual Agents*, C. Pelachaud, J. C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, Eds., Berlin, Heidelberg: Springer, 2007, vol. 4722, pp. 197-210, doi: 10.1007/978-3-540-74997-4_19

[9] T. B. Talbot, K. Sagae, B. John, and A. A. Rizzo, "Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations," *Int J Gaming Comput-Mediat Simul*, vol. 43, no. 3, pp. 1-19, 2012, doi: 10.4018/jgcms.2012070101.

[10] K. Maicher, D. Danforth, A. Price, L. Zimmerman, B. Wilcox, B. Liston, H. Cronau, L. Belknap, C. Ledford, D. Way, D. Post, A. Macerollo, and M. Rizer, "Developing a conversational virtual standardized patient to enable students to practice history-taking skills," *Simul Healthc*, vol. 12, no. 2, pp. 124-131, 2017, doi: 10.1097/SIH.0000000000000195.

[11] R. Hubal, J. Folsom-Kovarik, A. Woods, R. Jones, and J. Carbone, "Patterns of life in the foreground and background: Practical approaches to enhancing simulation-based interaction skills training," in *Proc Behav Represent Model Simul*, 2015, pp. 75-83.

[12] J. Abich, L. Reinerman-Jones, G. Matthews, G. Welch, S. Lackey, C. Hughes, and A. Nagendran, "Good enough yet? A preliminary evaluation of human-surrogate interaction," *Virtual, Augment Mix Real. Des Dev Virtual Augment Env*, R. Shumaker and S. Lackey, Eds., Cham: Springer, 2014, doi: 10.1007/978-3-319-07458-0_23.

[13] D. M. Nicholson DM, D. D. Schmorrow, and J. Cohn, Eds, *The PSI Handb Virtual Env Train Educ: VE Compon Train Technol: Vol. 2. Technol, Psychol Health*. Praeger, 2009.

[14] M. J. Roy, D. L. Sticha, P. L. Kraus, and D. E. Olsen, "Simulation and virtual reality in medical education and therapy: A protocol," *Cyberpsychol Behav*, vol. 9, no. 2, pp. 245-247, 2006, doi: 10.1089/cpb.2006.9.245.

[15] C. Guinn and R. Hubal, "Extracting emotional information from the text of spoken dialog," in *Proc Workshop Assess Adapt User Attitudes Affect: Why, When and How?*, 2003, doi: 10.1.1.119.5549.

[16] F. A. A. Laleye, A. Blanié, A. Brouquet, D. Behnamou, and G. de Chalendar, "Semantic similarity to improve question understanding in a virtual patient," 2019, *arXiv:1912.07421* [cs.CL].

[17] L. Campillos Llanos, C. Thomas, E. Bilinski, P. Zweigenbaum, and S. Rosset, "Designing a virtual patient dialogue system based on terminology-rich resources: Challenges and evaluation," *Nat Lang Eng*, vol. 26, no. 2, pp. 183-220, 2020, doi: 10.1017/S1351324919000329.

[18] M. A. Zielke, D. Zakhidov, G. Hardee, L. Evans, S. Lenox, N. Orr, D. Fino, and G. Mathialagan, "Developing virtual patients with VR/AR for a natural user interface in medical teaching," in *Proc Int Conf Serious Games Appl Health*, 2017, pp. 1-8.

[19] L. Heinrichs, P. Dev, and D. Davies, "Virtual environments and virtual patients in healthcare," *Healthc Simul Educ: Evid, Theory Pract*, D. Nestel, M. Kelly, B. Jolly, and M. Watson, Eds., Wiley, 2018, pp. 69-79, doi: 10.1002/9781119061656.ch10.