

Physical-Virtual Agents for Healthcare Simulation

Salam Daher
University of Central Florida
Orlando, FL
salam@knights.ucf.edu

Jason Hochreiter
University of Central Florida
Orlando, FL
jhochrei@cs.ucf.edu

Nahal Norouzi
University of Central Florida
Orlando, FL
nahal.norouzi@Knights.ucf.edu

Laura Gonzalez
University of Central Florida
Orlando, FL
Laura.Gonzalez@ucf.edu

Gerd Bruder
University of Central Florida
Orlando, FL
Gerd.Bruder@ucf.edu

Greg Welch
University of Central Florida
Orlando, FL
welch@ucf.edu

ABSTRACT

Conventional Intelligent Virtual Agents (IVAs) focus primarily on the visual and auditory channels for both the agent and the interacting human: the agent displays a visual appearance and speech as output, while processing the human's verbal and non-verbal behavior as input. However, some interactions, particularly those between a patient and healthcare provider, inherently include tactile components. We introduce an Intelligent Physical-Virtual Agent (IPVA) head that occupies an appropriate physical volume; can be touched; and via human-in-the-loop control can change appearance, listen, speak, and react physiologically in response to human behavior. Compared to a traditional IVA, it provides a physical affordance, allowing for more realistic and compelling human-agent interactions. In a user study focusing on neurological assessment of a simulated patient showing stroke symptoms, we compared the IPVA head with a high-fidelity touch-aware mannequin that has a static appearance. Various measures of the human subjects indicated greater attention, affinity for, and presence with the IPVA patient, all factors that can improve healthcare training.

CCS CONCEPTS

• **Computing methodologies** → **Mixed / augmented reality**; • **Human-centered computing** → **Displays and imagers**; • **Applied computing** → **Psychology**;

KEYWORDS

Physical-virtual agents, patient simulator, neurological assessment

ACM Reference Format:

Salam Daher, Jason Hochreiter, Nahal Norouzi, Laura Gonzalez, Gerd Bruder, and Greg Welch. 2018. Physical-Virtual Agents for Healthcare Simulation. In *Proceedings of IVA 2018*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3267851.3267876>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IVA 2018, November 5-8, 2018, Sydney, NSW, Australia

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6013-5/18/11...\$15.00

<https://doi.org/10.1145/3267851.3267876>

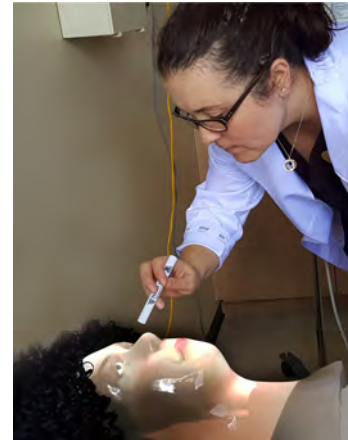


Figure 1: Nursing student interacting with the Intelligent Physical-Virtual Agent during a neurological assessment.

1 INTRODUCTION

Intelligent Virtual Agents (IVAs) are commonly used in training, simulation, and education across different fields and applications, such as healthcare, military and police training, serious games, entertainment, interview training, and educator training [38]. Most previous research and development efforts focused on the visual and auditory channels of communication between IVAs and users, which are arguably the most important modalities for most real-world applications. However, adding touch input and tactile feedback capabilities to an IVA has shown much potential to additionally improve the user's perception of the IVA [23, 26, 27]. Facilitating a high sense of *being* with a real person can make training with IVAs more effective, engaging, and relevant [10, 34].

In this paper, we introduce an Intelligent Physical-Virtual Agent (IPVA) in the shape of a life-size physical head supporting dynamic imagery (see Figure 1). An interactive computer-generated virtual agent is projected onto the physical shell of a head. The IPVA is capable of displaying a wide variety of symptoms related to its intended use as a simulated patient in healthcare applications. In particular, for a considered simulated stroke scenario we developed and integrated appropriate visual and behavioral content. In a user study, participants performed a simulated patient stroke assessment using our IPVA, which we compared to assessment of a high-fidelity mannequin in a baseline condition. Both simulators were able to respond verbally to participants, drawing from the same finite set of responses. Though both simulators were aware of participant touch,

the mannequin's reactions were limited to verbal responses, while the IPVA was additionally capable of visual feedback (Table 1).

In this paper, we address the following research questions:

- Q1 Will users experience greater *social presence* with an IPVA than when interacting with a mannequin?
- Q2 Will users have a more positive *user experience* with an IPVA than when interacting with a mannequin?
- Q3 Will users rate the *communication and interaction abilities* higher for an IPVA than a mannequin?
- Q4 Will users focus their *attention* more often on an IPVA than a mannequin?

2 BACKGROUND

In this section, we provide an overview of related work regarding the related concepts of presence and physicality, the use of IVAs in healthcare simulation, and the notion of physical-virtual agents.

2.1 Sense of Presence

An important concept is the user's sense of *being* with a real person, often measured by presence, co-presence, and social presence. *Presence* can be defined as "the sense of non-mediation": one who is oblivious to the existence of a technological medium can perceive presence via that medium [32]. The sense of another person's presence, called *co-presence*, exists when people mutually perceive one another [14]; it can also be defined as the degree to which one believes himself or herself to be in the presence of and interacting with other veritable human beings [4, 5]. *Social presence* is sometimes defined as one's sense of being socially connected with another person [9]. Harms and Biocca explain co-presence as one of several dimensions of social presence, confirming the validity of their measures with questionnaires [15].

Chuah et al. defined an IVA's *physicality* to include the physical size, volume, and position it occupies and its ability to interact with its surrounding environment [8]. Increasing an IVA's physicality has been observed to increase social presence [25], communication, and pro-active and re-active behavior [21]. Using agents with a physical component, Lok et al. showed that the physicality of an agent can have benefits for social presence and the training of communication skills in the scope of medical team training [31].

2.2 Healthcare Patient Simulators

In healthcare education and research, simulation is used to provide safe and consistent learning and practice environments. Typically, healthcare educators use standardized patients, mannequins, and computer-based simulation [28]. *Standardized patients* are individuals trained to simulate a set of symptoms to portray a real patient [33]. Standardized patients cannot change their physiology or certain appearance aspects at will (e.g., temperature, blood pressure, facial droop), which can make certain medical conditions hard to simulate (e.g., stroke, sepsis). *Mannequins* are robotic life-sized human-like patient simulators [33] that physically occupy volume and can change physiology to simulate a wide range of medical conditions, but they have a static appearance (e.g., facial expressions, skin color, inability to move). When a full body is not required, task trainers—models that represent a part or region of the human body, such as an arm or an abdomen—can be effective training

tools. *Computer-based simulation*, typically presented on interactive flat displays with graphical and text output, can include virtual humans (e.g., patients, nurses, doctors), virtual reality task trainers, and immersive virtual reality simulation [6]. Virtual patient simulations are consistently associated with higher learning outcomes compared to other educational methods [17], and virtual patient assessment shows high acceptance among medical students [13]. For example, learners can improve their pain observation skills by training with virtual patients [37], and it can be easier for learners to detect pain in a virtual human than in a high-fidelity, facially expressive humanoid robot patient simulator [36].

In this paper, we focus on the simulation of a stroke patient. Early recognition and treatment of stroke can limit brain damage and improve patient outcomes [40]. A variety of complicated symptoms can indicate a stroke, including smile asymmetry, irregular pupils, facial droop, and garbled speech [1]. Task trainers, standardized patients, and mannequins are capable of portraying some of these symptoms, and they are typically used for stroke assessment training. However, due to these simulators' inability to exhibit certain symptoms, supplementary information must often be provided to learners (e.g., using cue cards, images, videos, or verbal instructor responses) [12]. Computer-based VHs are able to simulate visual stroke characteristics through the use of dynamic imagery, providing more directly observable representations of stroke symptoms.

2.3 Physical-Virtual Agents

A special type of agent with a high sense of physicality is enabled by the technology of *Shader Lamps* [41]. Shader Lamps based virtual patients combine a front- or back-projected human-shaped surface with a projector to provide computer-generated visual feedback to observers, thus allowing medical students to conduct ophthalmic exams in an interactive training experience [41]. For instance, NERVE simulates a life-size virtual patient with cranial nerve injury [20, 30], giving medical students standardized experiences to interviewing, examining, and diagnosing virtual patients with cranial nerve disorders [17]. When medical students interacted with such virtual patients with cranial nerve injury of different gender and skin, female patients were correctly diagnosed more frequently than their male counterpart [42]. Physical-virtual agents can support automated touch sensing with integrated graphical response through rear-mounted IR lights and cameras [18, 19].

In this paper, we developed a Shader Lamps based physical-virtual agent for a stroke assessment scenario, where it is imperative for users to be able to touch the agent. For example, a patient experiencing a stroke may be unable to perceive touch on certain parts of their face. In related work, researchers built a virtual stroke patient that links a parametric model of aetiology to verbal and non-verbal behavior which can be manipulated in real time to allow learners to visually explore connections between different stroke aetiologies and social behavior [7]. However, to our knowledge, this virtual patient does not have a physical component and has not been compared to other healthcare simulators, such as mannequins.

3 EXPERIMENT

In this section, we describe the development of an IPVA head as a Shader Lamps based healthcare simulator, the development of a

medical training scenario with a simulated patient showing signs of a stroke, and a human-subject study that we performed to evaluate and compare this IPVA to a high-fidelity mannequin. In a between-subject study design, nursing students assessed the two simulators, and we evaluated their sense of social presence, mood, and other attributes using subjective responses and head tracking. To focus solely on differences due to audiovisual patient reactions, including to touch, we relied on equivalent human-in-the-loop control for both simulators.

3.1 Participants

Overall, 59 undergraduate nursing students (51 females, 8 males) from health assessment participated in this human-subject study. All 59 participants had been exposed to physical mannequins, 57 participants had been exposed to standardized patients (real actors), and 32 participants had been exposed to virtual simulations such as Shadow Health [44] or Second Life [43].

The study was performed as part of a nursing class where it is common practice to split students into pairs or triplets depending on the number of students and available simulators. Typically, two students interact with a patient, and the third person observes the interaction from inside the room. In this study, 44 students interacted with the simulator and 15 observed the simulation due to limited space around the patient. There is evidence that learning outcomes are not strongly affected based on whether students interacted or observed the interaction [29].

3.2 Scenario

We developed a training scenario in which a patient, called *Vera Real*, who is a woman in her 40s, shows up to the emergency department with one-sided upper and lower hemiplegia (paralysis of one side of the body). Upon examination, nurses would discover that the patient has visual loss and is complaining of a headache. The patient's speech is slurred with a one-sided facial droop. The patient's character is polite and neutral in general (neither positive nor negative). Vera is capable of showing different one-sided facial expressions when asked to do so (e.g., smile, frown, raise eyebrow). In neurological assessments, it is important to determine whether a patient can perceive touch; in particular, as Vera is experiencing a stroke, she is unable to feel and respond to touch on the affected half of her face. As part of the medical scenario, Vera starts alert and responsive, but as the examination continues, her condition deteriorates and she becomes forgetful, confused, and scared. Nursing students are expected to perform a neurological exam, recognize the stroke symptoms, and call the doctor for further instructions.

3.3 Apparatus Development

For this experiment, we developed and adapted two setups representing this healthcare training scenario: one with a physical-virtual head and one with a physical mannequin head. Both conditions used the same commercial Laerdal full-body SimMom mannequin and its physiology capabilities (i.e. breathing, heart sounds, and pulse) [45], whereas they differed in the type of head used (Figure 2). Typically in mannequin simulation, a person controls the patient's verbal responses from a control room by speaking into a microphone, and the sound comes out of the mannequin. The



Figure 2: The two study conditions: IPVA (top) and Mannequin (bottom).

patient's physiology is often controlled by this same operator using a computer interface. For this study, we created a finite set of garbled verbal responses common to both conditions. The verbal responses were limited to specific information regarding the patient (e.g. name, age), her condition (e.g. "I have a headache,") and basic responses (e.g. "yes," "no"); they did not include backchannels. Using a graphical interface, the simulator controller triggered patient responses that come from speakers located below the mannequin. The same software was used in both experimental conditions to control the verbal responses of the patient. The IPVA's verbal responses were lip-synced and contained facial expressions commonly used in speech, such as occasional eyebrow raising and blinking. Table 1 shows a comparison between the properties and capabilities of the simulated patient head for the two study conditions.

Physical-Virtual Agent Head. We created an IPVA head consisting of a wooden rig that supports a semi-transparent plastic head-shaped shell and a projector (Figure 3 right). We designed an interactive 3D graphical face that matches the shape and size of the head shell. An AAXA P300 pico projector (resolution 1920x1080) projects imagery of the animated face onto the shell from below.

Table 1: Properties and capabilities of the simulated patient head in the IPVA and Mannequin conditions.

Property/Capability	IPVA Head	Mannequin Head
Realistic physical shape	Yes	Yes
Human-in-the-loop control	Yes	Yes
Additional operator responses	Available	Required
Touch-aware	Yes	Yes
Verbal responses	Yes	Yes
to touch	Yes	Yes
to questions	Yes	Yes
Facial appearance	Dynamic	Static
Facial expressions	Yes	No
Lip syncing	Yes	No
Eye/pupil movement	Yes	No
Visual responses	Yes	No
to touch	Yes	No
to questions	Yes	No
to light and motion	Yes	No

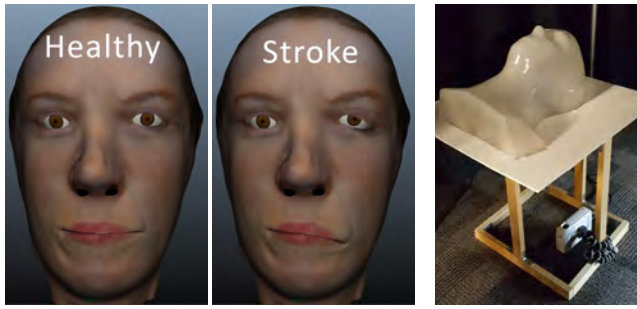


Figure 3: Left: virtual patients projected onto the IPVA. Right: wooden rig with projector and head-shaped shell.

The shell was scanned using the photogrammetry software Autodesk ReCap. The geometry from the scan is dense with a topology that is difficult to texture and animate. In Maya, we created a 3D mesh of the head that matched the form of the shell and designed a topology appropriate for facial animation. The head is made out of one mesh for the face (2657 vertices, 2541 faces) and two meshes for the eyeballs (72 vertices, 71 faces each). Each eyeball is connected to a joint to control its movement. The vertices of the face mesh are weighted to the head and jaw joints to allow for jaw opening. Combinations of blendshapes (e.g. eyelid and eyebrow movement, lip movement, blink, one-sided facial droop, smile) are used to create facial expressions and visemes. We imported the 3D model to Unity and designed a graphical user interface to trigger the 129 pre-recorded audio responses, which were lip-synced to the patient's lips using Rogo Digital Lipsync. Verbal responses were played on speakers below the head. The simulator controller observed the assessment from the control room and triggered the patient's verbal and facial responses with a graphical interface, whether prompted by participant speech or touch. The healthy and stroke patient models we designed for the IPVA are shown in Figure 3.

Physical Mannequin Head. The control condition consisted only of the SimMom mannequin, including the full body and the head. As in the experimental condition, the simulator controller observed the participants and triggered appropriate verbal responses from the patient whether prompted by participant speech or touch interactions. The mannequin was not capable of portraying certain visual symptoms, such as facial asymmetry; to obtain information related to such symptoms, participants would verbally state their question and receive a verbal response from the simulator controller.

3.4 Study Design and Procedure

In a between-subject design, 26 participants interacted and 10 observed the patient in the control condition (Mannequin), and 18 interacted and 5 observed the patient in the experimental condition (IPVA). The Institutional Review Board approved this human-subject study. The study was conducted over 4 class sessions, with one condition operational per session due to setup time; participants were assigned to conditions based on their date of attendance. First, all participants were asked to fill out a demographics pre-questionnaire. All participants were already familiar with the capabilities of the mannequin, and those who interacted with the IPVA watched a two-minute video to familiarize them with its capabilities. The video featured a healthy agent with a non-slurred

voice in an interaction with a healthcare provider that highlighted the capabilities of the IPVA, including verbal behavior, non-verbal behavior (such as facial expressions, eyeball movement following an object, pupillary reaction to light), and response to touch. Participants were given the patient's history and asked to perform a physical assessment of the patient. The simulation started when the controller said "begin simulation" and had a time limit of 15 minutes. After the simulation, participants were asked to answer a post-questionnaire with qualitative and quantitative sections.

For both conditions, the controller observed the simulation from a separate control room and used the speakers in the simulation room to inform participants about the start of the simulation and any patient actions (or inactions) that the simulator was (or was not) capable of representing (e.g., "begin simulation," "patient cannot move this arm," "one-sided smile"). When the participant touched the patient and asked if she could feel the touch, she would respond accordingly depending on the side (stroke vs. non-stroke). Participants had to actively inquire to receive certain cues; for example, in the Mannequin condition, the participant had to ask the patient to smile in order to receive the answer "asymmetric smile." In the IPVA condition, if the participant asked the patient to smile, they received their response directly as visual feedback by observing the patient; if the participant asked the patient to move her arm (part of the mannequin's body) they received a response from the controller (e.g., "patient cannot move right arm").

3.5 Measures

Before the simulation, a pre-questionnaire containing demographics questions was asked. During the simulation, participants were video recorded and tracked using a Kinect. After the simulation, participants answered a questionnaire with qualitative and quantitative questions. Observers completed the same questionnaires but indicated that they did not actually interact with the patient.

We measured realism by asking specific questions about the avatar, such as facial expressions and animations, and we measured social presence using a modified questionnaire from Harms and Biocca [15] shown in Table 2. Two questions from the original Harms and Biocca's perceived behavioral interdependence (Bhv) questionnaires were omitted as they did not fit in this patient-provider case. In addition, participants were asked to complete the affective attraction (AffAtt) questionnaire [16], shown in Table 3. As a gauge of their mood during the simulation, participants were asked "How did your interaction with the patient (Vera) make you feel?" with respect to feeling anxious, excited, tense, alert, in control, and having a desire to leave the situation, each as a Likert scale from "not at all" (0) to "extremely strong" (10). Finally, participants were also asked miscellaneous questions about the simulator's touch/response interaction and communication abilities, shown in Table 4. The touch response interactions were achieved by the controller using a GUI to trigger pre-recorded responses.

4 RESULTS

We used non-parametric statistical tests (Mann-Whitney U) to analyze the Likert-scale ordinal data from the questionnaires comparing the Mannequin condition with the IPVA condition. In some scientific disciplines it is common practice to treat Likert-type scales

Table 2: Social presence questions (scale from 1 to 7).

Co-Presence (CoP)	
1	I noticed the patient (Vera).
2	The patient (Vera) noticed me.
3	The patient (Vera)'s presence was obvious to me.
4	My presence was obvious to the patient (Vera).
5	The patient (Vera) caught my attention.
6	I caught the patient (Vera)'s attention.
Attentional Allocation (Att)	
1	I was easily distracted from the patient (Vera) when other things were going on outside this room.
2	The patient (Vera) was easily distracted from me when other things were going on outside this room.
3	I remained focused on the patient (Vera) throughout our interaction.
4	The patient (Vera) remained focused on me throughout our interaction.
5	The patient (Vera) did not receive my full attention.
6	I did not receive the patient (Vera)'s full attention.
Perceived Message Understanding (Msg)	
1	My thoughts were clear to the patient (Vera).
2	The patient (Vera)'s thoughts were clear to me.
3	It was easy to understand the patient (Vera).
4	The patient (Vera) found it easy to understand me.
5	Understanding the patient (Vera) was difficult.
6	The patient (Vera) had difficulty understanding me.
Perceived Affective Understanding (Aff)	
1	I could tell how the patient (Vera) felt.
2	The patient (Vera) could tell how I felt.
3	The patient (Vera)'s emotions were not clear to me.
4	My emotions were not clear to the patient (Vera).
5	I could describe the patient (Vera)'s feelings accurately.
6	The patient (Vera) could describe my feelings accurately.
Perceived Emotional Interdependence (Emo)	
1	I was sometimes influenced by the patient (Vera)'s moods.
2	The patient (Vera) was sometimes influenced by my moods.
3	The patient (Vera)'s feelings influenced the mood of our interaction.
4	My feelings influenced the mood of our interaction.
5	The patient (Vera)'s attitudes influenced how I felt.
6	My attitudes influenced how the patient (Vera) felt.
Perceived Behavioral Interdependence (Bhv)	
1	My behavior was often in direct response to the patient (Vera)'s behavior.
2	The behavior of the patient (Vera) was often in direct response to my behavior.
3	The patient (Vera)'s behavior was closely tied to my behavior.
4	My behavior was closely tied to the patient (Vera)'s behavior.

Table 3: Affective attraction questions (scale from 1 to 7).

Affective Attraction (AffAttr)	
1	How unpleasant/pleasant do you feel about the patient (Vera)?
2	How cold/warm do you feel about the patient (Vera)?
3	How negative/positive do you feel about the patient (Vera)?
4	How unfriendly/friendly do you feel toward the patient (Vera)?
5	How distant/close do you feel to the patient (Vera)?

Table 4: Miscellaneous questions (scale from 1 to 7).

Miscellaneous	
1	Rate the simulator's touch/response interaction.
2	From this interaction, rate the simulated patient's ability to communicate with you.
3	From this interaction, rate your ability to communicate with the simulated patient.

as interval-level measurements [3]. We avoid the discussion on whether parametric statistics can be a valid method for the analysis of non-parametric data [22, 24] by using non-parametric tests.

4.1 Subjective Questionnaires

Social Presence. Results show a significantly higher or a trend for higher social presence for the IPVA group in multiple dimensions (Figure 4), supporting our research question Q1.

Participants were more likely to have a higher perceived message understanding (Msg) in the IPVA group ($M = 5.342$, $SD = 0.861$) than in the Mannequin group ($M = 4.514$, $SD = 1.094$). The difference is statistically significant ($W = 117.0$, $p = 0.006$).

Participants were more likely to have a higher perceived behavioral interdependence (Bhv) in the IPVA group ($M = 4.150$, $SD = 1.375$) than in the Mannequin group ($M = 3.250$, $SD = 1.103$). The difference is statistically significant ($W = 143.0$, $p = 0.032$).

There is a trend suggesting participants could be more likely to have a higher co-presence (CoP) in the IPVA group ($M = 5.192$, $SD = 1.090$) than in the Mannequin group ($M = 4.420$, $SD = 1.334$). The difference is very close to being significant ($W = 150.0$, $p = 0.052$). Similarly, the attentional allocation (Att) shows a higher trend for the IPVA group. These results could become significant with a larger sample.

The remaining dimensions of social presence, perceived affective understanding (Aff) and perceived emotional interdependence (Emo), did not show a statistically significant difference ($p > 0.05$) between the groups.

Affective Attraction. Participants felt a higher affective attraction (AffAtt) towards the patient in the IPVA group ($M = 4.780$, $SD = 1.180$) than in the Mannequin group ($M = 4.157$, $SD = 1.379$) (Figure 4). The difference is statistically significant ($W = 146.0$, $p = 0.043$).

Mood Rating. Results for the mood questionnaires are shown in Figure 5. Participants felt more "in control of the situation" in the IPVA group ($M = 5.650$, $SD = 2.323$) than in the Mannequin group ($M = 4.261$, $SD = 2.200$). The difference is statistically significant ($W = 142.0$, $p = 0.031$), supporting our research question Q2.

There is a trend suggesting participants in the IPVA group could more likely feel more excited and more alert than those in the Mannequin group. The difference is very close to being significant (excited: $W = 154.5$, $p = 0.064$; alert: $W = 155.0$, $p = 0.065$) and could actually be significant with a larger sample.

The rest of the mood rating measures did not show a statistically significant difference ($p > 0.05$) for feeling anxious, tense, or a desire to leave the situation.

Touch Response and Communication. Results for the touch response and communication questions are shown in Figure 6. Together, they support our research question Q3.

Participants were more likely to rate the IPVA's touch/response higher in the IPVA group ($M = 5.054$, $SD = 1.261$) than in the Mannequin group ($M = 3.294$, $SD = 1.750$). The difference is statistically significant ($W = 169.5$, $p < 0.001$). This is in spite of the fact that the touch response was achieved identically for both conditions.

Participants were more likely to rate the simulated patient's ability to communicate with them higher in the IPVA group ($M = 5.565$, $SD = 0.945$) than in the Mannequin group ($M = 4.559$, $SD = 1.599$). The difference is statistically significant ($W = 239.0$, $p = 0.011$).

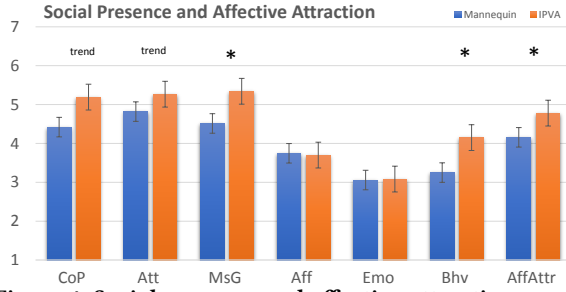


Figure 4: Social presence and affective attraction results.

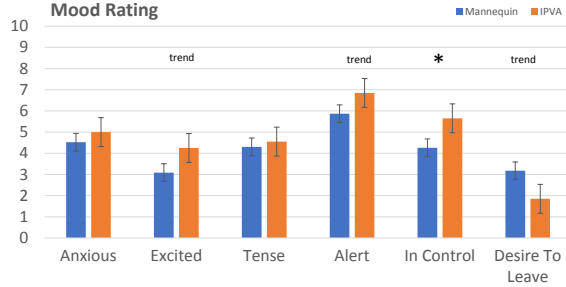


Figure 5: Mood rating results.

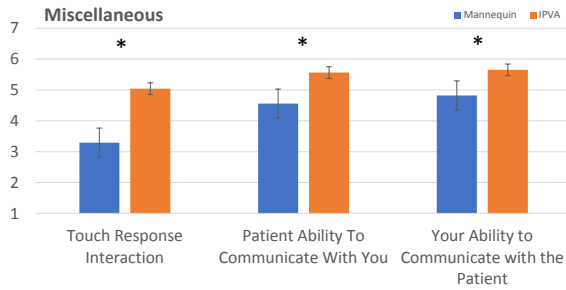


Figure 6: Miscellaneous results.

Participants were more likely to rate their ability to communicate with the simulated patient higher in the IPVA group ($M = 5.652$, $SD = 0.647$) than in the Mannequin group ($M = 4.824$, $SD = 1.527$). The difference is statistically significant ($W = 258.0$, $p = 0.025$).

Realism. Participants were asked to rate different aspects of the IPVA's realism on a 1 to 6 scale, where 1 represents the most "inexpressive" and 6 represents the most "expressive." Participants also had the option of "N/A" for questions that were not applicable for the Mannequin condition. Realism questionnaire results were all statistically significant ($p < 0.01$) and are shown in Figure 7.

4.2 Head Tracking

Human gaze behavior has been extensively studied as an objective measure of a person's focus of attention [11, 39]. Head orientation provides a less obtrusive but less accurate measurement of attention; several researchers have investigated estimating head pose and orientation from video recordings, including in the context of social interactions during group meetings [2, 35, 46]. As such, we tracked the participants' head positions and orientations using a Microsoft Kinect sensor to understand their head movement behavior and visual attention. From this data, we measured the amount of time participants spent facing Vera to compare how their attention

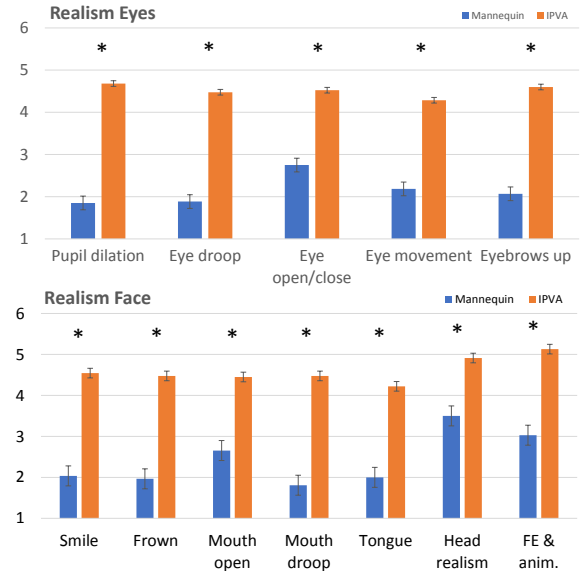


Figure 7: Realism results.

varied between the two conditions. We only tracked the heads of interacting participants, not observers.

We collected 3D reference points in the simulation environment corresponding to Vera's body and head. Using these points, we constructed cubic volumes representing various regions of interest: Vera's head, Vera's body, and everywhere else. Intersections of head orientation vectors and these regions represent participant attention throughout the simulation. Using the Kruskal-Wallis test, we observed significant differences indicating that participants spent a greater portion of simulation time focusing on the head of the IPVA than on the head of the mannequin ($H = 6.0208$, $p = 0.014$). There were no significant differences between the two conditions regarding participant attention to the mannequin's body ($H = 0.3333$, $p = 0.56$) or to everywhere else ($H = 2.5208$, $p = 0.11$).

4.3 Qualitative Feedback

Participants (both interactors and observers) were asked to answer the following question by free writing: "Based on your assessment session, what do you think about this simulated patient (specifically the head)?"

Mannequin. Many of the participants (17 of 36) who assessed the mannequin felt it was difficult to perform a neurological assessment due its static appearance. Six of the qualitative responses mentioned the absence of facial expressions, and 9 mentioned the lack of eye and pupil movement. In terms of realism, 4 participants commented on the mannequin's general inability to "reenact neurological signs." Other participant responses indicated that mannequin's static appearance hindered the "realistic aspect of the scenario," that the lack of realism "[took me] out of my element of treating a real patient," and that the limited interaction capabilities prevented them from "physically [seeing] the abnormal." Three participants complained about the delayed responses. Seven of the participants had positive experiences with the mannequin, with several finding the patient and its responses to be realistic. One felt it was "as realistic as it could be," and another suggested it was "fine for this test."

IPVA. Of the participants in the IPVA condition, 9 out of 23 “liked it,” and 3 indicated that it was “easy” or “helpful for neuro,” with 2 highlighting the facial expressions and pupils. Eight commended the realism, and 5 appreciated the head’s interactivity and ability to answer questions (e.g. “more interactive and reactive to prompts,” “more lifelike,” “it could sense touch,” “much better interaction with a face that can do things”). Eleven specifically mentioned that it was more realistic or more helpful than traditional mannequins. Three noted it was “slightly creepy” or “a little scary.” Five complained that the responses were “a bit laggy, slowing down assessment time.” Three did not like that the head was separate from the body, expressing a preference to see a full-body version. Three commented negatively on realism: the “quality of [the] face is poor. Should reflect [an] actual person,” the “tongue was a bit awkward,” and it is “not very lifelike but an improvement from the mannequins.”

5 DISCUSSION

Life-size touch-aware interactive virtual patients that occupy volume are desirable and can be effective healthcare training simulators, such as for assessing patients with stroke. Unsurprisingly, the IPVA was significantly more realistic than the mannequin; we measured realism to have a benchmark compared to the mannequin. Nursing students who interacted with the IPVA showed significantly higher *Msg*, rating the communication abilities of the IPVA higher than the mannequin in terms of both of the patient’s ability to communicate with them and their ability to communicate with the patient. The *Bhv* was also significantly higher for the IPVA. We think these results are due to the IPVA’s ability to indicate understanding and respond to participant behavior; for example, the IPVA can respond verbally and visually to participant actions, including verbal requests and physical touches, whereas the mannequin was only capable of providing verbal responses.

The trend towards higher *CoP* for the IPVA group could be due to the simulator’s greater ability to show environmental awareness and react to participant behavior compared to the mannequin. For example, the IPVA can react to a participant’s touch, exhibit pupillary movement in response to light, and display facial expressions in response to participant behavior. While the *Att* was not significantly different between groups, it was close enough to support head tracking results. The fact that the IPVA did not rotate towards the participants could have affected the results of the *Att*. With a larger sample, the *CoP* and *Att* could show significance. Interestingly, touch interaction was rated higher for the IPVA than the mannequin even though the responses were controlled equivalently by an observer pressing buttons. It seems that the addition of visuals during the lip-sync of the verbal responses might have made it appear that the IPVA was more responsive to touch. We were not surprised that there was no difference in the *Emo* and *AffAtt* between conditions as the patient was designed to be neutral. The *AffAtt* questionnaire measured how participants felt toward the patient. The IPVA was more likely to be considered “pleasant, warm, positive, friendly, close” than the mannequin.

By design, the patient deteriorated cognitively during the simulation, starting from being alert and oriented to becoming confused, forgetful, and disoriented, which was demonstrated verbally and was equivalent in both conditions. While the patient’s outcome

was out of the control of participants, those who assessed the IPVA felt more “in control of the situation” than those who assessed the mannequin. The IPVA provided participants with more direct control over visual and tactile assessment through integrated graphics, while assessment of the mannequin was indirect, requiring participants to explicitly request information and wait for a verbal response from the simulation operator. The increased perception of being in control might have increased participants’ social presence in the *Bhv* dimension, which measures the degree to which patient actions were directly affected by participant actions.

There is a trend suggesting participants in the IPVA condition felt more alert and more excited. The excitement could be attributed to the novelty of the simulator, while the alertness could be attributed to the fact that the IPVA eyes could blink and look at them.

Limitations. We designed this study with a consistent finite set of verbal responses between the Mannequin and IPVA conditions. Performing patient eye movement in response to medical tests in real time while observing participants was challenging; automating this capability of the patient is recommended. The IPVA’s eyes and pupils could move, but the head could not physically rotate to face the participants.

6 CONCLUSION

In this paper, we described the development and evaluation of an Intelligent Physical-Virtual Agent (IPVA) head for neurological simulation. Participants were split into a group that assessed a high-fidelity healthcare mannequin and a group that assessed the IPVA with the mannequin body. In both groups, the simulator occupied space, allowing participants to touch the patient. The addition of interactive realistic visuals on the IPVA head resulted in higher social presence and *AffAtt* compared to the mannequin, supporting our research question Q1. Participants felt more in control of the situation when interacting with the IPVA, supporting our research question Q2, and they rated the IPVA’s touch response and communication abilities higher than those of the mannequin, supporting our research question Q3. Head tracking results indicated that participants in the IPVA condition spent a greater percentage of assessment time looking at the patient’s head than those in the Mannequin condition, supporting our research question Q4. Adding a touch sensation to IVAs could enhance the user’s perception towards them, which could lead to better training outcomes.

Future work involves allowing for more head and neck rotations for the patient, whether virtual or physical. Also, many participants noted that the head in the IPVA condition was separate from the rest of the patient’s body and expressed interest in seeing the same technology extended to a full body. We plan to apply these same principles to a full body physical-virtual simulator. We are also interested in incorporating low-latency touch input through automated touch sensing. Finally, we plan to investigate the learning outcomes of assessment of the IPVA.

ACKNOWLEDGMENTS

This work is supported primarily by NSF Award no. 1564065 CHS: Medium: Physical-Virtual Patient Bed for Healthcare Training and Assessment, Program Director Dr. Ephraim P. Glinert. We acknowledge the RADM IITSEC committee, the Link Foundation, and the

UCF Modeling and Simulation graduate program for their support of co-author Salam Daher via research fellowships. We also acknowledge Florida Hospital for their support of Prof. Welch via their Endowed Chair in Healthcare Simulation.

REFERENCES

- [1] Act FAST | Stroke.org. 2018. <http://www.stroke.org/understand-stroke/recognizing-stroke/act-fast>. (May 2018). Accessed: 05/09/2018.
- [2] Ben Benfold and Ian D. Reid. 2009. Guiding visual surveillance by tracking human attention. In *BMVC*, Vol. 2. 7.
- [3] Norman Blaikie. 2003. *Analyzing quantitative data: From description to explanation*. Sage.
- [4] Jim Blascovich. 2002. Social influence within immersive virtual environments. In *The social life of avatars*. Springer, 127–145.
- [5] Jim Blascovich, Jack Loomis, Andrew C. Beall, Kimberly R. Swinth, Crystal L. Hoyt, and Jeremy N. Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry* 13, 2 (2002), 103–124.
- [6] Elodie Bonnetain, Jean-Michel Boucheix, Maël Hamet, and Marc Freysz. 2010. Benefits of computer screen-based simulation in learning cardiac arrest procedures. *Medical education* 44, 7 (2010), 716–722.
- [7] Harry Brenton, Peter Woodward, Marco Gillies, Jonathan Birns, Diane Ames, and Fernando Bello. 2015. Linking aetiology with social communication in a virtual stroke patient. In *International Conference on Intelligent Virtual Agents*. Springer, 270–274.
- [8] Joon Hao Chuah, Andrew Robb, Casey White, Adam Wendling, Samsun Lam-potang, Regis Kopper, and Benjamin Lok. 2013. Exploring agent physicality and social presence for medical team training. *Presence: Teleoperators and Virtual Environments* 22, 2 (2013), 141–170.
- [9] Salam Daher, Kangsoo Kim, Myungho Lee, Ryan Schubert, Gerd Bruder, Jeremy Bailenson, and Greg Welch. 2017. Effects of Social Priming on Social Presence with Intelligent Virtual Agents. In *International Conference on Intelligent Virtual Agents*. Springer, 87–100.
- [10] Gianluca De Leo, Leigh A. Diggs, Elena Radici, and Thomas W. Mastaglio. 2014. Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment. *Simulation in Healthcare* 9, 1 (2014), 1–6.
- [11] Nathan J. Emery. 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews* 24, 6 (2000), 581–604.
- [12] Mark J Garside, Matthew P Rudd, and Christopher I Price. 2012. Stroke and TIA assessment training: A new simulation-based approach to teaching acute stroke assessment. *Simulation in Healthcare* 7, 2 (2012), 117–122.
- [13] Neil Gesundheit, Pauline Brutlag, Patricia Youngblood, William T. Gunning, Nabil Zary, and Uno Fors. 2009. The use of virtual patients to assess the clinical skills and reasoning of medical students: initial insights on student acceptance. *Medical Teacher* 31, 8 (2009), 739 – 742.
- [14] Erving Goffman. 1969. *Behavior in public places; notes on the social organization of gatherings*. New York, Free Press of Glencoe 1969.
- [15] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. (2004).
- [16] Kenneth C. Herbst, Lowell Gaertner, and Chester A. Insko. 2003. My head says yes but my heart says no: Cognitive and affective attraction as a function of similarity to the ideal self. *Journal of personality and social psychology* 84, 6 (2003), 1206–1219.
- [17] Atsusi Hirumi, Andrea Kleinsmith, Kyle Johnsen, Stacey Kubovec, Michael Eakins, Kenneth Bogert, Diego J. Rivera-Gutierrez, Ramsamooj Javier Reyes, Benjamin Lok, and Juan Cendan. 2016. Advancing virtual patient simulations through design research and interPLAY: part I: design and development. *Educational Technology Research and Development* 64, 4 (2016), 763–785.
- [18] Jason Hochreiter, Salam Daher, Arjun Nagendran, Laura Gonzalez, and Greg Welch. 2015. Touch sensing on non-parametric rear-projection surfaces: A physical-virtual head for hands-on healthcare training. In *Virtual Reality (VR), 2015 IEEE*. IEEE, 69–74.
- [19] Jason Hochreiter, Salam Daher, Arjun Nagendran, Laura Gonzalez, and Greg Welch. 2016. Optical touch sensing on nonparametric rear-projection surfaces for interactive physical-virtual experiences. *PRESENCE: Teleoperators and Virtual Environments* 25, 1 (2016), 33–46.
- [20] Teresa R. Johnson, Rebecca Lyons, Regis Kopper, Kyle J. Johnsen, Benjamin C. Lok, and Juan C. Cendan. 2014. Virtual patient simulations and optimal social learning context: a replication of an aptitude-treatment interaction effect. *Medical Teacher* 36, 6 (2014), 486 – 494.
- [21] Kangsoo Kim, Arjun Nagendran, Jeremy N. Bailenson, Andrew Raij, Gerd Bruder, Myungho Lee, Ryan Schubert, Xin Yan, and Gregory F. Welch. 2017. A Large-Scale Study of Surrogate Physicality and Gesturing on Human-Surrogate Interactions in a Public Space. *Frontiers in Robotics and AI* 4 (2017), 1–20.
- [22] Thomas R. Knapp. 1990. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research* 39, 2 (1990), 121–123.
- [23] Aaron Kotranza, Benjamin Lok, Adeline Deladisma, Carla M. Pugh, and D. Scott Lind. 2009. Mixed reality humans: Evaluating behavior, usability, and acceptability. *IEEE Transactions on Visualization and Computer Graphics* 15, 3 (2009), 369–382.
- [24] William M. Kuzon Jr, Melanie G. Urbanchek, and Steven McCabe. 1996. The seven deadly sins of statistical analysis. *Annals of plastic surgery* 37, 3 (1996), 265–272.
- [25] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *International Journal of Human-Computer Studies* 64, 10 (2006), 962–973.
- [26] Myungho Lee, Gerd Bruder, and Gregory F. Welch. 2017. Exploring the effect of vibrotactile feedback through the floor on social presence in an immersive virtual environment. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 105–111.
- [27] Myungho Lee, Kangsoo Kim, Salam Daher, Andrew Raij, Ryan Schubert, Jeremy Bailenson, and Greg Welch. 2016. The wobbly table: Increased social presence via subtle incidental movement of a real-virtual table. In *Virtual Reality (VR), 2016 IEEE*. IEEE, 11–17.
- [28] Adam I. Levine, Samuel DeMaria Jr, Andrew D. Schwartz, and Alan J. Sim. 2013. *The comprehensive textbook of healthcare simulation*. Springer Science & Business Media.
- [29] Tze Wei Liew, Su-Mae Tan, and Hishamuddin Ismail. 2017. Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website. *Human-centric Computing and Information Sciences* 7, 1 (2017), 42.
- [30] Chen Lin, Candelario Laserna, D. Scott Lind, Chandni Kalaria, Tia Aryal, Benjamin Lok, Kyle Johnsen, Aaron Kotranza, Juan Cendan, and J. Ned Pruitt. 2012. User Response to the Simulation of a Virtual Patient with Cranial Nerve Injury. *Bio-Algorithms and Medical Systems* 8, 1 (2012), 1.
- [31] Benjamin Lok, Joon Hao Chuah, Andrew Robb, Andrew Cordar, Samsun Lam-potang, Adam Wendling, and Casey White. 2014. Mixed-reality humans for team training. *IEEE Computer Graphics and Applications* 34, 3 (2014), 72–75.
- [32] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication* 3, 2 (1997), 0–0.
- [33] Joseph O. Lopreiato, Dayna Downing, Wendy Gammon, Lori Lioce, Barbara Sittner, Vickie Slot, Andrew Spain, Terminology & Concepts Working Group, et al. 2016. *Healthcare simulation dictionary*. Rockville, MD: Agency for Healthcare Research and Quality (2016).
- [34] Fabrizia Mantovani and Gianluca Castelnuovo. 2003. The sense of presence in virtual training: enhancing skills acquisition and transfer of knowledge through learning experience in virtual environments. (2003).
- [35] Benoît Massé, Silève Ba, and Radu Horaud. 2017. Tracking Gaze and Visual Focus of Attention of People Involved in Social Interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).
- [36] Maryam Moosaei, Sumit K. Das, Dan O. Popa, and Laurel D. Riek. 2017. Using Facially Expressive Robots to Calibrate Clinical Pain Perception. *ACM/IEEE International Conference on Human-Robot Interaction* (2017), 32 – 41.
- [37] Maryam Moosaei, Michael J. Gonzales, and Laurel D. Riek. 2014. Naturalistic pain synthesis for virtual patients. In *International Conference on Intelligent Virtual Agents*. Springer, 295–309.
- [38] Toni Bloodworth Pence, Lauren Cairco Dukes, Larry F. Hodges, Nancy K. Meehan, and Arlene Johnson. 2013. The effects of interaction and visual fidelity on learning outcomes for a virtual pediatric patient system. In *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 209–218.
- [39] David I. Perrett and Nathan J. Emery. 1994. Understanding the intentions of others from visual signals: Neurophysiological evidence. (1994).
- [40] Recognizing stroke early - Harvard Health. 2018. <https://www.health.harvard.edu/heart-health/recognizing-stroke-early>. (May 2018). Accessed: 05/09/2018.
- [41] Diego Rivera-Gutierrez, Greg Welch, Peter Lincoln, Mary Whitton, Juan Cendan, David A. Chesnutt, Henry Fuchs, and Benjamin Lok. 2012. Shader Lamps Virtual Patients: the physical manifestation of virtual patients. *Studies In Health Technology And Informatics* 173 (2012), 372 – 378.
- [42] Diego J. Rivera-Gutierrez, Regis Kopper, Andrea Kleinsmith, Juan Cendan, Glen Finney, and Benjamin Lok. 2014. Exploring gender biases with virtual patients for high stakes interpersonal skills training. In *International Conference on Intelligent Virtual Agents*. Springer, 385–396.
- [43] Second Life. 2018. <http://secondlife.com/>. (May 2018). Accessed: 05/10/2018.
- [44] Shadow Health. 2018. <https://shadowhealth.com/>. (May 2018). Accessed: 05/10/2018.
- [45] SimMom. 2018. <https://www.laerdal.com/us/products/simulation-training/obstetrics-pediatrics/simmom/>. (May 2018). Accessed: 05/09/2018.
- [46] Michael Voit and Rainer Stiefelhagen. 2008. Deducing the visual focus of attention from head pose estimation in dynamic multi-view meeting scenarios. In *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 173–180.