Mixed Reality Humans: Evaluating Behavior, Usability, and Acceptability

Aaron Kotranza, Benjamin Lok, *Member*, *IEEE*, Adeline Deladisma, Carla M. Pugh, and D. Scott Lind

Abstract—This paper presents mixed reality humans (MRHs), a new type of embodied agent enabling touch-driven communication. Affording touch between human and agent allows MRHs to simulate interpersonal scenarios in which touch plays a crucial role. Two studies provide an initial evaluation of user behavior with an MRH patient and the *usability* and *acceptability* of an MRH patient for practice and evaluation of medical students' clinical skills. In Study I $(n = 8)$, it was observed that students treated MRHs as social actors more than students in prior interactions with virtual human patients $(n = 27)$, and used interpersonal touch to comfort and reassure the MRH patient similarly to prior interactions with human patients $(n = 76)$. In the within-subjects Study II $(n = 11)$, medical students performed a clinical breast exam on each of the MRH and human patient. Participants performed equivalent exams with the MRH and human patients, demonstrating the *usability* of MRHs to evaluate students' physical exam skills. The *acceptability* of the MRH patient for practicing exam skills was high as students rated the experience as believable and educationally beneficial. Acceptability was improved from Study I to Study II due to an increase in the MRH's visual realism, demonstrating that visual realism is critical for simulation of specific interpersonal scenarios.

Index Terms—Intelligent agents, virtual reality, human-centered computing, user interfaces, evaluation/methodology, medicine.

1 Introduction

Mixed reality humans (MRHs) are a new type of embodied agent that affords touch-driven communication. By affording touch-driven communication between a human and an embodied agent, MRHs are able to simulate interpersonal scenarios in which touch is a critical component of effective communication. This paper presents the implementation of an MRH patient (Fig. 1) and explores the potential of MRHs for practicing of interpersonal and physical exam skills in medical education. Results of two user studies are presented: Study I makes observations of medical students' behavior with MRH, virtual human (VH), and human patients. Study II evaluates the usability and acceptability of MRHs for practice and evaluation of medical students' clinical examination skills.

1.1 Motivation

Interpersonal simulation allows users to practice their interpersonal skills by communicating with a VH. In current interpersonal simulators, the communication between a user and the VH takes the form of a conversation consisting of bidirectional speech and simple gestures.

- *A. Kotranza and B. Lok are with the Department of Computer and Information Science and Engineering, University of Florida, E301 CSE Building, PO Box 116120, Gainesville, FL 32611. E-mail: {akotranz, lok}@cise.ufl.edu.*
- *A. Deladisma and D.S. Lind are with the Department of Surgery, Medical College of Georgia, BB 4514, Augusta, GA 30912. E-mail: {adeladisma, dlind}@mcg.edu.*
- *C.M. Pugh is with the Feinberg School of Medicine, Northwestern University, 201 E Huron St, Chicago, IL 60611. E-mail: cmpugh@nmh.org.*

Manuscript received 25 June 2008; revised 4 Sept. 2008; accepted 21 Oct. 2008; published online 20 Nov. 2008.
Recommended for acceptance by M.C. Lin.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCGSI-2008-06-0081.
Digital Object Identifier no. 10.1109/TVCG.2008.195.

Use of VHs and interpersonal simulation is rapidly expanding to educate users in military [1], [2], law enforcement [3], cultural competency [4], [5], and medical [6], [7], [8] domains.

In these and other application domains, many scenarios exist in which interpersonal touch plays a crucial role (e.g., medical physical examinations). As current interpersonal simulations lack interpersonal touch, they provide an incomplete simulation of human-human communication and can only be applied to a limited set of interpersonal scenarios. By enabling touch-driven communication, interpersonal simulation can be applied to a wide range of scenarios in which touch is a critical component, and can more completely simulate human-human communication.

1.2 Innovations of Mixed Reality Humans

The innovations of MRHs are to use tangible interfaces and image processing techniques to afford touch-driven communication between a person and a virtual person.

The tangible interface represents a portion of the virtual person's body and is registered to the virtual person's visual representation. The tangible interface provides passive haptic feedback and actively detects a user's touch. The user's touch is interpreted by a simulation which allows the MRH to respond with appropriate speech and gestures (Fig. 2).

Image processing techniques are used to track passive physical objects, such as the MRH's clothes, and to incorporate a visually faithful avatar of the user into the virtual world. Image processing techniques and the tangible interface allow the user's touch to drive communication with the MRH.

1.3 Driving Application: Medical Education

Our goal is to apply MRHs to training of health profession students' interpersonal and physical exam skills. This is

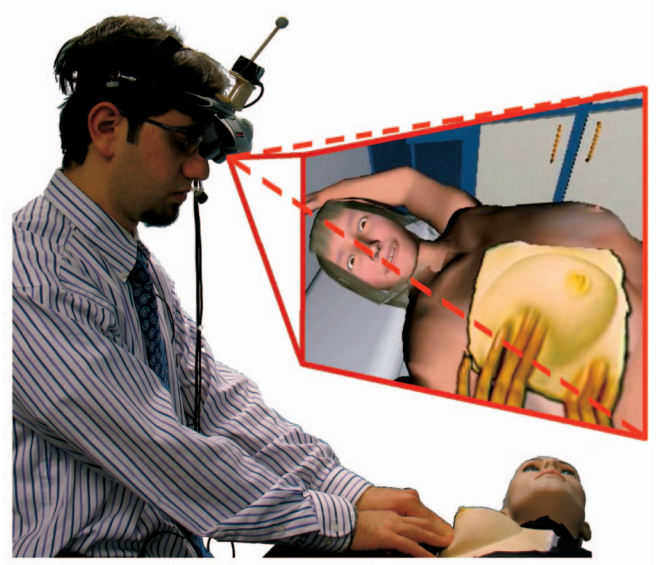

Fig. 1. A medical student performs a CBE on an MRH patient. The student touches a tangible interface (an instrumented mannequin) and views the patient's visual representation through an HMD.

motivated by strong evidence that interpersonal skills do not simply improve with experience, but must be specifically targeted and practiced in order to improve [9]. MRH patients may be able to both target students' communication skills and train physical examination skills.

Current methods of training physical examination include standardized human patients (SPs, actors trained to represent an illness) and physical simulators (e.g., Fig. 3). SPs are a limited resource and are unable to simulate the symptoms of many illnesses (e.g., a lazy eye, cancer). Additionally, it is difficult and costly to recruit SPs for intimate exams such as breast, pelvic, and prostate exams. Physical simulators accurately simulate a wide range of illnesses. However, these simulators lack the social interaction required to train interpersonal skills.

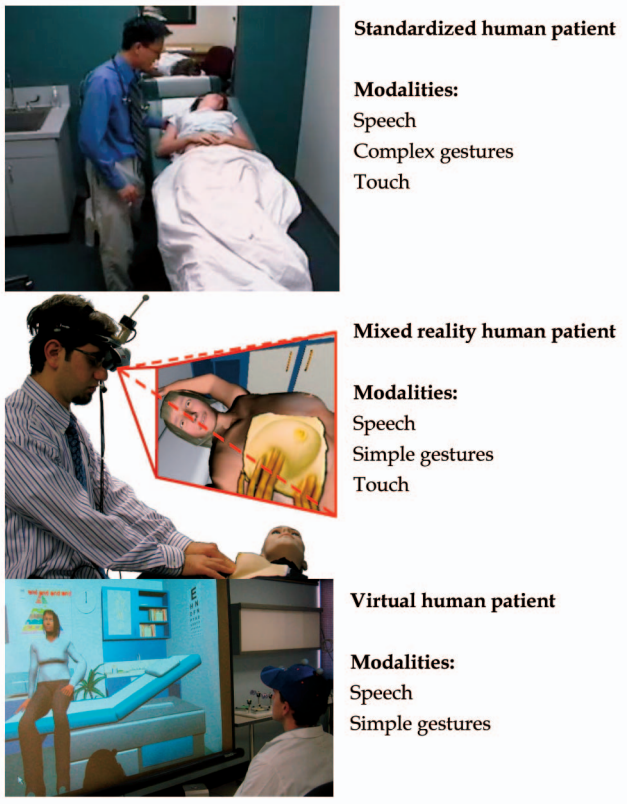

Fig. 2. Bottom to top: Medical interpersonal simulations in increasing order of communication modalities afforded.

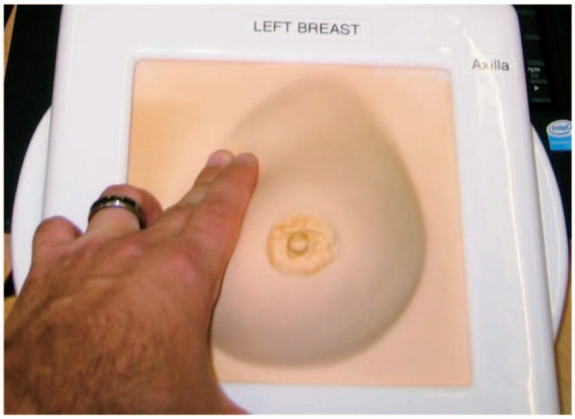

Fig. 3. A physical breast simulator is harnessed as a natural interface to the MRH breast exam patient.

MRHs overcome the drawbacks to SP and physical simulation approaches. An MRH patient combines 1) the touching of a simulator to perform a physical exam and 2) the interpersonal communication of conversing with a patient. This combination provides an experience more similar to clinical context than practicing interpersonal and physical exam skills separately.

1.4 Implementing and Evaluating an MRH Patient

This paper presents the implementation of an MRH patient who requires a clinical breast examination (CBE). Building toward the goal of incorporating MRH patients into a medical education curriculum for training clinical exam skills, two user studies provide an initial evaluation of 1) medical students' behavior when interacting with the MRH breast exam patient and 2) the usability and acceptability of using MRH patients for practice and evaluation of students' communication and physical exam skills.

Study I $(n = 8)$, presented at the IEEE Virtual Reality 2008 [10], observed medical students' behavior when talking to and performing a physical examination of the MRH breast exam patient. Participants' behavior was compared to that of medical students interacting with both VH and SPs in previous studies (Study VH: $n = 27$, Study SP: $n = 76$). More students expressed empathy toward the MRH patient than toward the VH patient, and students exhibited increased social engagement with the MRH patient. Students also used comforting and reassuring touches with the MRH patient, and used these touches with a frequency similar to students interacting with an SP.

The evaluation of Study II focuses on the usability and acceptability of the MRH patient for practicing and evaluating medical student's physical exam skills. Evaluation of usability and acceptability of a new user interface or

system for training is a critical step before the system can be handed over to users [11]. The usability and acceptability of the MRH patient must be evaluated before the MRH can be incorporated into a medical school curriculum. The usability of a system is the extent to which it can be used for its intended purpose (*can the MRH patient be used for practicing and evaluating student exam skills?*). The acceptability of a system is the extent to which users are willing to use the system for its intended purpose (*do medical students perceive an educational benefit from interacting with the MRH patient?*).

To evaluate the usability and acceptability of MRH patients for practicing physical examination and communication skills, the within-subjects **Study II** $(n = 11)$ compared students' performance in CBEs of each of an MRH patient and an SP. Objective ratings of students' performance with the SP and MRH were significantly correlated, indicating that students performed equivalent breast exams on the SP and MRH patient. This demonstrates the usability of MRH patients for practicing physical examination skills and for allowing medical educators to evaluate student's examination skills. That students performed equivalent exams suggests that students treated the evaluation of the breast health of the MRH patient and SP as being of similar importance.

To evaluate the acceptability of the MRH patient, the believability and perceived benefit of the MRH and SP experiences were compared. Students reported similar believability and educational benefit of the MRH and SP interactions, showing that students accepted the MRH experience as a worthwhile practice tool. Study II does not investigate the validity of the MRH patient for training clinical skills, nor does it investigate training transfer; we leave this to future work.

From Study I to Study II, the visual realism of the MRH patient was increased. Participants' subjective ratings of the two MRH patients showed that the increase in visual fidelity enhanced the believability and perceived educational benefit of the MRH patient. This demonstrates that domain-specific visual realism of the MRH experience is critical to users' acceptability of MRHs.

Taken as a whole, the observations of Study I and Study II demonstrate that the touch-driven communication afforded by MRHs expands the range of interpersonal scenarios to which interpersonal simulation can be applied, and does so effectively: Users perceive these experiences as being similar to real-world counterparts, and treat interaction with an MRH similarly to interaction with a human.

2 Previous Work

2.1 Interpersonal Simulation

Interpersonal simulators allow a person to communicate with a VH through speech and gestures. In an interpersonal simulator, the VH is the focal point for interaction. This distinguishes interpersonal simulators from applications which incorporate VHs as interfaces to an underlying system.

The goal of interpersonal simulators is to train users' communication skills. Interpersonal simulators have been employed to train users' communication skills in military leadership [1] and negotiation [2] scenarios. In the medical education domain, interpersonal simulators aid in training students' clinical examination skills [6] and communication in medical interviews [7].

2.1.1 People Treat Virtual and Real Interpersonal Scenarios Similarly

Studies of interactions with VHs have demonstrated that people treat virtual and real people similarly. Pertaub et al. [12] found that people with a fear of public speaking experienced a similar anxiety when speaking in front of an audience of VHs. Bailenson et al. have shown that people afford VHs personal space [13]. Zanbaka et al. have shown that VHs are able to persuade people [14].

Johnsen et al. [15] compared the Virtual Objective Structured Clinical Exam (VOSCE), an interpersonal simulation of a medical interview, to its real counterpart, the OSCE. Students' performance in the VOSCE predicted their performance in the OSCE. This indicates that, for this scenario, the interaction skills used with the VH translate to the interaction skills used with a real human.

Raij et al. [8] directly compared medical students' behavior in two interpersonal scenarios: a medical interview of a VH patient and a medical interview of an SP. It was found that the content of the virtual and real interactions was similar, but participants' behavior was different. Participants appeared to be less engaged and were less sincere when conversing with the VH.

2.2 Effective Communication Requires Touch

Interpersonal touch is a critical component of communication in interpersonal scenarios. Without affording touch, VHs provide an incomplete simulation of an interpersonal scenario. In clinical and professional situations, interpersonal touch has been shown to increase information flow and causes people to evaluate communication partners more favorably [16]. Interpersonal touch is the most commonly used method of comforting [17] and allows social norms to be enforced (e.g., handshakes).

In medicine, touch improves doctor-patient communication by increasing patient verbalization, disclosure of information, and attitudes toward the medical professional [17]. Patients obtain increased reassurance through doctors' nonverbal cues, including touch, than through verbal communication alone [18]. Physical contact plays a critical role in forming a close doctor-patient relationship. Interpersonal touch conveys the idea that the caregiver is aiding the patient and is an important component of expressing empathy [19].

2.3 Virtual Experiences are Enhanced by Touch

Adding haptics to virtual environments has been shown to increase task performance and believability of the experience. Adding passive haptics to the UNC Pit environment increased users' sense of presence [20]. Remote users collaborating on a virtual task reported an increased sense of togetherness and improved task performance when active-haptic feedback was provided [21].

Bailenson and Yee proposed the concept of virtual interpersonal touch [22], allowing touch between either a human and VH or two remotely located humans. In an initial evaluation, participants were able to brush virtual dirt from a VH using an active haptic PHANTOM Omni.

2.4 Social Interaction Is Absent in Medical Simulation

Physical simulator (e.g., Fig. 3) approaches to teaching hands-on medical skills place an emphasis on teaching technique over communicating with the patient. Because these physical simulators do not afford bidirectional communication between doctor and patient, the experience is not similar to the clinical context of patient interaction, leaving many graduating students unprepared to enter clinical practice [23].

With the focus on teaching technique, many physical simulators aim only to provide only a realistic look and feel by, e.g., simulating soft tissue with physical models and allowing users to interact using real surgical tools. These simulators may provide feedback of student performance, but universally lack the social interaction necessary for students to place the examination in the clinical context of patient interaction.

Concerns over the learning of technique separately from clinical context have been addressed by Kneebone et al. [23]. To simulate a more realistic patient experience, a physical simulator is used to augment an SP. Students converse with the SP while performing the exam on the simulator placed near the SP. However, this approach limits the availability of the simulation and the ability to simulate abnormal conditions.

An MRH patient harnesses physical simulators as natural interfaces to a virtual person, incorporating the advantages of physical simulators (e.g., high fidelity soft-tissue simulation) while providing touch, verbal, and gestural communication similar to interaction with a human patient.

3 Mixed Reality Humans

To afford touch-driven communication between human and virtual person, MRHs merge virtual and real spaces: the virtual person's visual representation is colocated with a tangible interface that represents a portion of the virtual person's body. Users touch the tangible interface while viewing the virtual person through a display device such as a head-mounted display (HMD) (Fig. 1).

This section presents the general design of an MRH and the specific implementation of the MRH breast exam patients used in Study I and Study II.

3.1 Virtual Person (Visual Representation)

The virtual person is a full-body virtual character with facial animation and gestural capabilities including lip-synched speech, eye blinking, breathing, pointing, idle behaviors (e.g., swinging legs, looking around), the ability to maintain eye contact with the user, as well as scenario-specific animations.

Breast exam patient. The MRH breast exam patient of Study I and Study II is able to point to locations in her left breast where she has found a mass and where she is experiencing pain. The patient of Study II has animations to transition between four positions used in a breast exam: sitting with arms relaxed, sitting with hands pressed on hips and chest pushed out, sitting with arms raised above head, and lying down with arms raised above her head.

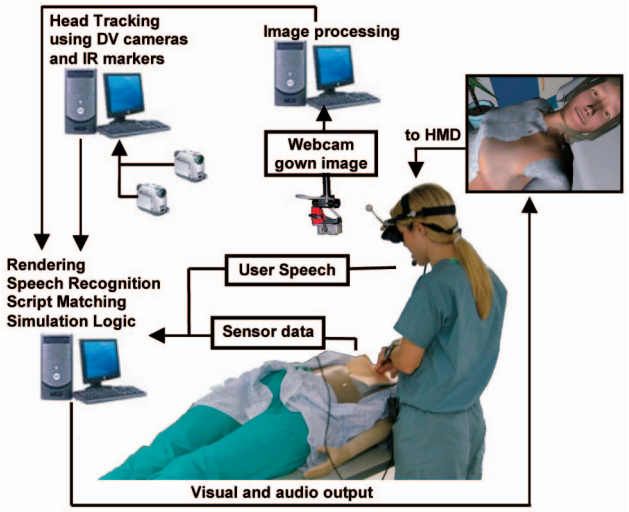

Fig. 4. System design: users interact with an MRH patient using natural speech, gestures, and touch.

3.2 Tangible Interface (Physical Representation)

The tangible interface of an MRH represents a portion of the MRH's body. The surface of the tangible interface is augmented with force sensors to detect the user's touch. A webcam overlooks the tangible interface, allowing tracking of passive components of the tangible interface (such as the MRH's clothes), and incorporation of the user's hands into the virtual scene.

Breast exam patient. The tangible interface of the MRH breast exam patient is an instrumented mannequin incorporating a physical breast simulator (Fig. 3) in place of the mannequin's left breast. The physical breast simulator incorporates 12 force sensors to detect palpation of the breast. For Study II, an additional 24 force sensors were added to the mannequin's abdomen, upper arm, left armpit, and clavicle in order to support palpation ancillary to the breast exam and to detect additional interpersonal touches such as a comforting touch on the shoulder.

The MRH patient of Study I wears a virtual hospital gown, the physical correlate of which is worn by the mannequin. The opening and closing of the gown is tracked by a webcam (Fig. 4), using background subtraction (details presented in [10]), causing the MRH's virtual gown to open and close. The gown is important to teach the process of a breast exam; to maximize patient comfort, only one side of the gown should be opened at a time during the examination. By tracking the gown, the opening and closing of the gown become inputs to the interaction.

3.3 Merging Real and Virtual Spaces

The virtual person is able to move freely, while the tangible interface is static. However, the scenario of a physical examination involves touching of the patient in a specific pose. This is the pose to which the tangible interface is registered. For Study I and Study II, registration of the virtual person and the tangible interface was performed manually with accuracy of approximately 5 cm.

3.4 Designing for Communication

The system layout is shown in Fig. 4. The user wears an HMD and wireless microphone. For Study I and Study II, the

HMD used was an eMagin z800 with 800 × 600 resolution and 40-degree diagonal field of view. This HMD was chosen for its lightweight (<8 oz), small form-factor, and unobtrusiveness relative to most high-resolution HMDs. Head pose (6 DOF) is tracked using an in-house infrared-marker tracking system and an Intersense InteriaCube 2. Users are able to interact with the MRH through a combination of verbal, gestural, and haptic communication techniques.

Verbal. The user communicates verbally with the MRH patient using natural speech. A wireless microphone transmits the user's speech to a PC which performs speech recognition using a commercial product, Dragon Naturally Speaking Pro 9. Recognized speech is matched to a database of question-answer pairs using a keyword-based approach. The database for a scenario consists of 100-300 question-answer pairs. The many syntactical ways of expressing a question are handled by the keyword-based approach and a list of common synonyms. The MRH responds to matched user speech with speech prerecorded by a human patient.

Gestural. The MRH has gaze behavior including looking at the user when the user is talking. The user's head gaze is also computed. To communicate with the user, the MRH changes facial expressions and has a set of key-frame-based animations for gestures that are specific to the scenario (e.g., the breast exam patient points to a painful part of her breast).

Haptic. The tangible interface of the MRH provides passive haptic feedback and allows the user's touch to be an input to drive the communication.

3.4.1 Touch-Driven Communication

A feedback loop allows the user's touch to drive communication with the MRH. The feedback loop proceeds as

1. the user's touch is detected by force sensors,
2. outputs of the force sensors are processed by a microcontroller and sent to a PC over a serial link,
3. the simulation module matches the current set of sensor values to an XML database of touch-gestures (e.g., *light palpation of left breast, resting hand on shoulder*), and
4. the simulation module instructs the rendering module to cause the MRH to respond to the touch using speech and gestures.

In this way, the user's touch drives the communication and the flow of the interaction. For example, in Study I, the MRH patient reported that she had pain in her left breast only after the user palpated her breast, eliciting verbal and gestural expressions of pain. In this case, the user's touch steered the conversation to the topic of the patient's breast pain.

4 Study I: Design

A pilot study was conducted with the goal of observing medical students' perceptions and behavior during a breast history and CBE of an MRH patient.

Students' behavior with the MRH patient is compared to participant behavior in previous studies of interactions with VH and SPs. This comparison provides an initial evaluation of whether adding touch-driven communication can impact a user's behavior when interacting with a virtual person.

4.1 Population

Eight $(n = 8)$ second-year physician-assistant students at the Medical College of Georgia conducted 10-min-long interactions consisting of a breast history and CBE of the MRH breast exam patient. Only one participant had previous experience performing a CBE on a human patient, although all had conducted medical histories of human patients (an average of 2.4 interviews).

4.2 Procedure

Demographic data was collected in a background survey. To view and speak to the MRH, participants wore an HMD and a wireless microphone. Participants completed a brief (5 minutes) speech training session. Prior to the interaction, the MRH's tangible interface was briefly visible to participants, but its affordances were not explained, and participants were not able to touch the tangible interface before the interview began.

The interaction lasted for approximately 10 minutes. Participants first took a breast history which is a medical history focusing on the patient's breast health. A medical history takes the form of a question and answer session. The breast history is followed by a CBE, which consists of visually inspecting the patient's breasts and palpating the patient's breasts to check for masses. After finishing the CBE, a postexperience survey was administered, and feedback was given in an informal debriefing.

4.3 Scenario

The MRH patient of Study I is a 56-year-old female who has recently found a lump in her left breast. Her sister is a breast cancer survivor. Due to her family history, she is very fearful that she may have cancer, and expresses this to the participant. The patient is also experiencing tenderness in an area of her left breast, which she does not express until the palpation portion of the exam.

Critical moments. Critical moments are points in the interaction where the patient challenges the participant either to express empathy or comfort the patient. Because MRHs support touch-driven communication, we are able to incorporate touch into these critical moments, both by having touch trigger critical moments and by eliciting comforting touches as responses to critical moments. The critical moments of Study I were 1) when a tender portion of her left breast is palpated, the patient expresses pain, exclaiming "*ouch!*" or "*that hurts*" and displaying pain with her facial expression and 2) after the participant finds the mass, the patient issues the challenge: "*does it feel like cancer?*" Participants' responses to critical moments are unrehearsed, providing genuine indications of a participant's engagement in the experience and feelings toward the patient.

4.3.1 Interaction Flow

Participants spent the first 5 minutes conversing with the MRH patient in order to take a breast history. The goals of this conversation are to elicit the patient's current complaint, past medical history and social history, as well as her family medical history.

A 5-minutes CBE was then performed. Participants opened one side of the gown, palpated the breast, closed

TABLE 1
Maastricht Assessment of Simulated Patients [24]

The patient appears authentic
The patient is challenging the student
The patient simulates physical complaints unrealistically**
The patient answers questions naturally
The patient withholds information unnecessarily**
The patient's appearance fits the role
The patient stimulates the student to ask questions
I can judge from the reactions of the patient whether she listens to me
The patient communicates how she felt during the interaction

Questions are presented as five-point Likert items. Items marked by a (**) are reverse coded.

the gown, and then repeated the process for the other breast. During the examination, the patient expressed pain when the participant palpated a tender area of her left breast. A set time after it was detected that the participant had palpated the mass, the patient challenged the participant with the question "*does it feel like cancer?*" This challenge was also elicited if students palpated the mass and asked "*is this the mass that you found?*" When finished with palpation, participants typically recommended that the patient receive a mammogram, and concluded the interview.

4.4 Measures

A prestudy survey assessed participants' subjective level of anxiety when performing medical histories and physical exams of human patients. The poststudy survey asked participants to report perceived believability of the MRH patient, and perceived educational benefit of the experience.

4.4.1 Believability

The Maastricht Assessment of the Simulated Patient (MaSP) [24] is a questionnaire validated for assessing the quality (realism and *believability*) of simulated human patients (Table 1). The MaSP provides a measure of how believable the MRH is in the role of a patient. Participants also gave an overall rating (1-10) of how realistic the MRH was as a patient.

4.4.2 Educational Benefit

The perceived educational benefit of interacting with the MRH was assessed on global (big picture) and local (specific benefits) measures.

Local. The following statements were rated on a five-point Likert scale: "*Interacting with the patient . . .*:

1. Increased my confidence in taking a breast history.
2. Increased my confidence in performing a CBE.
3. Decreased my anxiety over taking a breast history.
4. Decreased my anxiety over performing a CBE.

Global. Additionally, two global measures of educational benefit were also assessed on a five-point Likert scale:

1. I found the experience worthwhile.
2. I would use the system as a practice tool.

4.4.3 Behavioral

Video of participants' interactions were evaluated by the authors' collaborating medical education experts. To maintain objectivity, participants' behavior was evaluated using quantitative measures. Evaluation of behavior focused on 1) did the participant express empathy at the critical moment when the patient verbalized that she was in pain, 2) would the participant's response to the patient's verbalization of pain be acceptable when interacting with a human patient, and 3) how often did the participant use interpersonal touch to comfort or reassure the patient.

5 Study I: Results and Analysis

5.1 Statistical Analysis

Multivariate analysis of variance was used to test for significant differences at $\alpha < 0.05$. Throughout this section, statistics are presented in the form $\mu \pm \sigma$, where μ is a mean and σ is a standard deviation.

5.2 Believability and Educational Benefit

The average believability (MaSP) rating was 2.8 ± 0.8 out of 5.0. Only for the item "patient (does not) withhold information unnecessarily" did a majority (5/8) of participants rate the MRH patient positively. Of note, no participants felt that they were able to tell if the MRH patient was listening to them. The overall rating of the MRH patient was 4.5 ± 2.1 out of 10, lower than ratings reported with VH patients [15]. We believe the low reported believability is due to the addition of touch to the experience. Adding touch to the experience makes the experience more similar to users' prior human-human interactions, increasing users' expectation for the MRH to act like a human.

On global measures of educational benefit, 25 percent of the participants agreed that the experience as worthwhile, while 50 percent rated this as neutral. Thirty-eight percent of the participants reported that they would use the system as a practice tool. On local measures of educational benefit, 38 percent of the participants reported that the experience increased their confidence in both taking a breast history and a CBE. Only 25 percent of the participants reported that the experience decreased their anxiety in performing a breast history and a CBE.

Despite the low subjective ratings of the experience, participants displayed behavior with the MRH patient similar to what is expected with a human patient.

5.3 Behavior during Empathetic Moments

When participants touched a predefined (but unknown to participants) section of the MRH patient's left breast, the patient grimaced in pain and exclaimed "*ouch!*" or said "*that hurts*." Out of the seven participants who elicited this expression of pain, five responded empathetically (e.g., "*I'm sorry; I know that hurts and I'll try to be gentler*"). The other two participants responded with confirmatory statements ("*oh, I see you are tender there*"). One hundred percent of the participants responded to the MRH patient's expression of pain.

Both types of reactions are appropriate for an interaction with a human patient. This demonstrates that *participants*

treated the MRH patient how they are taught to (and how they are expected to) treat a human patient.

5.3.1 Comparison to Behavior with Virtual Humans

Study I participants' use of empathy was compared to the use of empathy by medical students in a previous interaction with a VH patient. In the previous Study VH, 27 $(n = 27)$ medical students took a breast history of a VH patient who had found a mass in her breast. Content of the medical histories of the Study VH patient and the MRH patient of Study I were the same. The Study VH patient used a different empathetic challenge: "*could this be cancer?*" to express fear.

In Study VH, only one-third of participants (9/27) responded empathetically to the patient's expression of fear. Additionally, more than one-quarter (7/27) participants completely ignored the patient's expression of fear. Ignoring the patient is not an appropriate response to a human patient's expression of fear. Participants in Study VH did not treat the VH patient similarly to how they are taught to treat a human patient.

Significantly more participants used an appropriate response with the MRH (Fisher's exact test: MRH 7/7; VH 9/27; $p < 0.005$). This suggests that more participants interpreted the MRH as a social actor than the VH. The difference in the number of participants who responded with empathy to the MRH and VH patients was nearly significant (Fisher's exact test: MRH 5/7; VH 7/27; $p = 0.09$). Although the two empathetic moments were not identical, this trend suggests that a larger study might find that medical students use empathy significantly more with MRH than VH patients.

5.4 Use of Interpersonal Touch

All participants were observed using interpersonal touch to comfort and reassure the patient. This was unexpected, as students were not told that touching the patient would affect the interaction.

5.4.1 Assumption of Tangible Interface Affordances

Participants appeared to assume that the tangible interface of the MRH had many or all of the affordances of a human body. Participants asked the MRH patient to, e.g., sit up, put her arms behind her head, and lie on her side. The MRH was unable to do any of these, which caused many participants to physically move the tangible interface in an attempt to, e.g., get the MRH to raise her arms. This observation demonstrates that 1) the affordances and limitations of the tangible interface must be made clearer and 2) providing a simple mannequin allowed students to fulfill their desire (and training) to use interpersonal touch with the MRH patient.

5.4.2 Comparison to a Human-Human Interaction

The number of comforting or reassuring touches used by participants with the MRH was similar to the number of touches used by medical students to comfort and reassure an SP in the previous Study SP. In Study SP, third-year medical students $(n = 76)$ conducted a medical history and physical exam of an SP with abdominal pain. Study I and Study SP participants' uses of interpersonal touch other than that required for the physical exam (e.g., a comforting touch on the shoulder) were compared through video review. The number (and frequency) of touches was similar (Study I: 1.4 ± 0.9; Study SP: 1.8 ± 1.8). Testing for equivalence yields a result consistent with the hypothesis that the number of touches in the MRH and SP interactions is equivalent (95 percent confidence interval (CI) $[-0.2, 1.1]$ overlaps zone of indifference (ZI) of $[-1, 1]$); however, a larger population is required to demonstrate statistical equivalence. *This observation shows that medical students will use interpersonal touch to comfort an MRH patient and will use these touches about as frequently as they would with a human patient.*

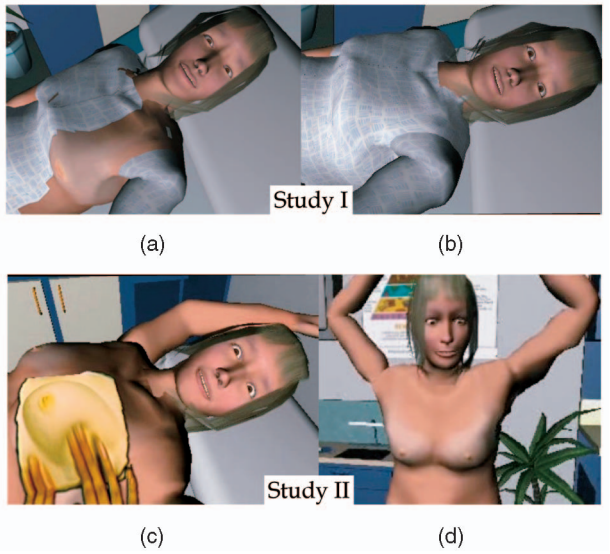

Fig. 5. (a) and (b) The appearance of the MRH breast exam patient in Study I. For Study II, (c) the appearance of the virtual breasts was improved to support visual inspection, and (d) during palpation, the physical breast and user's hands were overlaid to enhance the visual realism.

5.5 Feedback

Participants gave feedback in an informal debriefing. All participants commented that they wanted to be able to see their hands: e.g., "*should I be able to see my hands?*," "*I really wanted to be able to see my hands.*" Participants additionally commented that the visual and behavioral fidelity of the virtual breast tissue was not realistic. The MRH patient's virtual breast was a low-polygon, low-detail model (Fig. 5a) that did not deform when palpated by the participant. General consensus was that, although some participants also had issues with speech recognition, visual fidelity was the most important component of believability and acceptability of the MRH patient.

5.6 Discussion

Three observations are made with respect to participant behavior: 1) more participants empathized with the MRH than with the VH, 2) participants demonstrated increased social engagement with the MRH, and 3) participants used comforting and reassuring touch with the MRH similarly to with an SP.

5.6.1 Use of Empathy Is Increased with MRHs

A majority (71 percent) of Study I participants empathized with the patient's expression of pain ("*ouch!*" or "*that*

hurts") during palpation. Only one-third of participants in Study VH emphasized with the VH patient's expression of fear ("*could this be cancer?*"). The trend of increased empathy with the MRH is nearly significant (Fisher's exact test, $p = 0.09$). We acknowledge that the difference could be due to the fact that the two empathetic moments involved different expressions. The expression of pain may be more immediate, causing more participants to react to it. However, we suggest that this immediacy is due to participants being within the MRH's intimate physical space and engaging in interpersonal touch during the expression of pain. This scenario can only arise with an MRH, as a VH does not provide this physical interpersonal touch. More participants empathized with the MRH than with the VH. By affording touch, the MRH presented a more engaging experience. This leads to the observation that participants appeared to be more engaged socially with the MRH than with the VH.

5.6.2 Social Engagement Is Increased with MRHs

The majority of Study VH participants did *not* respond appropriately to the VH patient's expression of fear, while all Study I participants responded appropriately to the MRH patient's expression of pain. The difference in appropriate responders is significant (Fisher's exact test, $p < 0.005$). We believe that this is due to increased social engagement with MRHs.

Medical students are trained to respond appropriately to patient questions. That two-thirds of Study VH participants did not respond appropriately suggests that participants were not sufficient engaged in the experience to respond as they had been trained to do. Additionally, the ignoring of the VH patient's expression of fear by 26 percent of Study VH participants suggests that these participants did not afford the VH basic social respect they are trained to display—participants did not treat the VH as a social actor.

All participants in Study I responded appropriately to the MRH's expression of pain—participants appropriately treated the MRH as a social actor. This suggests that *participants treated the MRH more similarly to a human patient because they were more engaged in the experience*.

5.6.3 Interpersonal Touch Is Used with MRHs

Participants were observed using touch to comfort and reassure the MRH patient as frequently as was observed in doctor-patient interactions between two humans. We have observed 220 previous medical interviews of VH patients. In these 220 interactions, no touch was afforded. *As soon as touch was afforded by the tangible interface of the MRH patient, users began employing interpersonal touch as they would with a human patient*. Providing a simple plastic mannequin allowed students to satisfy their need (and training) to touch their interaction partner.

5.6.4 Poststudy Reflections

The feedback given by participants focuses further development on improving the visual realism of the MRH. Additionally, observing students' use of comforting and reassuring touches on the shoulder and abdomen emphasizes the need to place additional sensors on the tangible interface. These additional sensors will allow comforting and reassuring touches to be incorporated into the feedback-loop and to cause the MRH patient to appear to be comforted.

Because of the small population size of Study I, results are framed as observations and are considered preliminary. The comparisons of Study I participant behavior to the larger VH and human patient studies is promising, but should also be considered preliminary, as the three studies were not conducted in conjunction. Though Study I suggests that users will treat the MRH much like a human, Study I did not examine usability or provide evidence of acceptability of MRHs for training. Study II seeks to determine the usability and acceptability of MRHs for the practice and evaluation of clinical exam skills.

6 Improving Visual Realism of the MRH

To improve the visual realism of the MRH patient, video of the tangible interface's physical breast and the user's hands are incorporated into the virtual scene (Fig. 5c). This was accomplished through color segmentation of the video stream of a webcam overlooking the tangible interface. The tangible interface is colored black except for the breast, allowing the breast and the user's hands to be extracted from the video stream with little noise. This is accomplished by performing a Gaussian blur to remove noise, and color segmentation to separate the foreground from the black background. The resulting video stream (640×480, 30 fps) is projected onto a mesh (created by scanning the tangible interface with a laser scanner) using a projected texture. The alignment of the projected texture and mesh was performed manually, but, to automate alignment, the webcam and mannequin positions can be tracked by the IR tracking system.

Additionally, the virtual breast of the MRH was improved to support visual inspection, and the MRH was given additional animations to assume the different poses required for visual inspection (e.g., Fig. 5d).

These improvements allow a more complete CBE to be performed and increase the visual and behavioral fidelity of the MRH, allowing users to see their hands during palpation and the virtual breast tissue to behave realistically by deforming to the user's palpation.

7 Study II: Design

Study II evaluates the usability and acceptability of the MRH patient for student practice of clinical exam skills as well as allowing medical educators to evaluate students' skills. Participants completed a breast history and physical exam (CBE) of each of an MRH patient and an SP. Participant self-reports and objective measures of performance are used to evaluate the MRH's usability and acceptability.

To demonstrate usability for practice of exam skills, students must *be able to perform* equivalent exams of the MRH and SP. To demonstrate usability for evaluation, students must treat the MRH and SP exams with equal importance, shown by *performing* similar exams. Acceptability of the MRH patient will be judged by students' reports of believability (a motivating factor for using such a training tool) and perceived educational benefit.

7.1 Population

Eleven $(n = 11)$ medical students in their second semester of medical school conducted a breast history and CBE on each of an MRH and SP. A counterbalanced design was used with $(n = 6)$ students examining first the SP and then the MRH patient (Group SP-MRH), and $(n = 5)$ students examining first the MRH patient and then the SP (Group MRH-SP). One participant in each group was unable to perform the second interview due to medical school schedule conflicts. For within-subjects comparisons, nine participants are used. None of the participants had previous experience performing a CBE on a human patient, although all had conducted medical interviews of an SP (an average of 2.0).

7.2 Procedure

Demographic data was collected in a preexperience survey. Participants then performed the first exam (SP or MRH) taking 10 minutes. After completing a postexperience survey, participants moved on to the second exam (MRH or SP). Before examining the MRH patient, students completed a brief (5 minutes) speech training session, and were fitted with a wireless microphone and an HMD. Participants were instructed to first take a medical history of the MRH patient, and, if they determined that a physical exam was required, to perform the palpation portion of the exam using the MRH's tangible interface. The tangible interface was shown to participants during the instruction, but they did not touch it prior to beginning the MRH interaction.

7.3 Scenario

7.3.1 MRH Scenario and Typical Interaction

The MRH patient of Study II is a 34-year-old female who has been experiencing a persistent pain in her lower left breast for the past three weeks. She has lost her mother to breast cancer within the past two years, and is fearful that the exam will find an indication that she too has cancer.

Participants spent the first 5 minutes conversing with the MRH patient to take a breast history. Participants then conducted a visual inspection of the patient's breasts, asking her to first remove her gown, and then to pose in two positions: with arms raised over her head (Fig. 5d) and with hands pressed to her hips and chest thrust out.

After the visual inspection, participants asked the patient to lie down, and proceeded with palpation of the patient's left breast. When the participant was first detected palpating the patient's breast, the patient expresses fear that the examination might find something bad (cancer): "*wait! I'm kind of scared about this. Is it going to be ok?*" This expression of fear and prompting for reassurance should be responded to empathetically by the participant. When finished with palpation, participants discussed their findings with the patient. This is an opportunity for the participant to reassure and comfort the patient. The majority of participants empathized or responded reassuringly to the patients' expression of rear. As the focus of this paper is on usability and acceptability, a detailed discussion of participant behavior is left to future work.

7.3.2 SP Scenario and Typical Interaction

The SP of Study II is a middle-aged female who has found a breast mass that "comes and goes" for the last 6 months. Recently, the mass has increased in size. She has a family history of cancer, but not breast cancer, and is currently taking oral contraceptives. Due to SP availability, three actresses played the SP (and the exact age of the SP varied); varying of the actress was balanced between SP-MRH and MRH-SP groups.

Participants took a 5-minute breast history of the SP. The breast history was similar in content to that of the MRH patient. The CBE could not be performed on the SP's breasts (SPs with abnormalities, i.e., breast masses, are rarely if ever available). Instead, participants performed palpation on a physical breast simulator (Fig. 3) as the SP watched. Similarly to the MRH, the SP commented during the palpation, expressing pain if a tender area was palpated and acknowledging the location of the mass if asked by the participant (e.g., "*is this the mass you found?*"). The breast simulator is the same type used in the MRH patient.

It has previously been shown that, when an intimate exam cannot be conducted on the SP, integrating simulators into an SP encounter provides a similar educational benefit [25]. After palpation, participants discussed findings with the SP and concluded the interview.

At an appropriate point in the interview, the SP expressed fear concerning the exam and her situation, asking "*could it be cancer?*" Again, discussion of participants' behavioral reactions to this critical moment is left to future work.

7.4 Measures

In a preexperience survey, participants reported their experience level (number of interactions) in taking breast histories and performing CBEs on SPs. Postexperience, acceptability was assessed by two measures: believability of the MRH patient was assessed with the MaSP (Table 1) and perceived educational benefit was queried using a metric similar to that of Study I. Usability was evaluated by comparing objective measures of participant performance in the MRH and SP breast exams:

1. finding masses,
2. time spent palpating for masses,
3. differential diagnosis (what is wrong with the patient), and
4. diagnostic workup (what to do next).

The use of 2 and 4 stems from [28]. Participants were also asked to report what aspects of the MRH experience should be improved to increase the value of the MRH patient as a practice tool.

7.4.1 Usability: Task Performance

To assess whether participants conducted equivalent exams on the MRH patient and SP, participants were graded on

1. whether they found all masses in the patient's breast,
2. how much time they spent palpating the breast (indicating thoroughness of exam and amount of effort put into the exam),
3. their differential diagnosis, and
4. what they would recommend as the next step in treating the patient.

Items 1, 3, and 4 were reported by the participant in a note written after each interview, and item 2 was assessed by the authors watching video of the participants' exams.

The MRH had two masses in her breast and the SP had one. If the participant performed a complete breast exam,

the participant would find all masses present regardless of the number. The masses in the MRH and SP breasts were judged as being of equal difficulty to find by the authors' collaborating medical educators. A correct differential diagnosis was breast cancer or benign tumor. The correct next step is for the patient to have a mammogram.

Evaluating task performance assesses the usability of MRH patients for practice and evaluation of clinical skills. Usability will be established if performance with the MRH and SP are significantly correlated (i.e., a student's exams were equivalent, independent of whether the student has good exam skills). This will demonstrate that students can practice the clinical exam, and educators can evaluate student's exams, using the MRH as a substitute for (or augmentation of) currently used SP interactions.

7.4.2 Acceptability: Educational Benefit and Believability

The perceived educational benefit of the MRH experience was assessed on global (big picture) and local (specific benefits) measures (Section 4.4.2). Local measures were expanded from Study I by adding the following: "*Interacting with the patient* ...:

5. Increased my preparedness for taking a BH.
6. Increased my preparedness for performing a CBE.

Decreasing anxiety and improving confidence and preparedness in performing CBEs are major goals of medical education [27].

Believability, as measured by the MaSP, is a component of acceptability—if students do not believe in the MRH in the role of a patient, the MRH cannot be accepted for simulating the doctor-patient encounter. If participants rate the MRH and SP as being similarly believable patients, and perceive similar educational benefit of the MRH and SP, this will establish that the MRH is acceptable in simulating a patient for the purpose of practice of clinical exams.

7.4.3 Feedback

Participants were asked which system component improvements would make them more likely to use the MRH experience for practicing CBEs. Three possible improvements were ranked on a three-point scale from *most important to improve* to *least important to improve*:

1. The appearance of the patient.
2. The patient's ability to respond correctly to speech.
3. The tactile realism of the exam.

This provides a ranking of the user-perceived need for improvement in 1) visual realism and visual (gestural) communication, 2) verbal communication, and 3) tactile realism and touch-based communication.

8 Study II: Results and Analysis

8.1 Statistical Analysis

Repeated measures ANOVA was used to test for order effects—none were found—and test for significant differences at the $\alpha < 0.05$ level. Equivalence [26] was tested using a comparison of 90 percent CI to a ZI of $[-0.5, 0.5]$. The ZI was chosen because survey questions had a difference of 1.0 between possible values. *Equivalence* of two conditions is reported if the CI lies entirely inside the ZI. *Similarity* of two conditions is reported if the CI overlaps the ZI (this is consistent with equivalence but requires a larger population to detect). *Nonequivalence* is reported if the CI lies entirely outside of the ZI. When a comparison of the percentage of participants agreeing with a survey item is of interest, Fisher's exact test is used to test for significance.

8.2 Usability: Task Performance

Participant performance in the CBEs of the MRH and SP was significantly correlated.

The participant's ability to find the masses in the SP's breast was significantly correlated with the participant's ability to find the masses in the MRH's breast $(r^2 = 0.63, p < 0.05)$. Five participants found all masses in both the MRH and SP breasts; three participants did not find the masses in either breast; and only one participant found the mass in the SP breast but failed to find the masses in the MRH breast.

The time spent palpating the patients' breasts was significantly correlated $(r^2 = 0.56, p < 0.05)$. The time spent in palpation is representative of the thoroughness of the exam and the effort the participant put into the exam. This correlation shows that if a participant conducted a thorough exam on the SP, he performed a similarly thorough exam on the MRH.

All participants arrived at the same differential diagnoses and diagnostic workups with the SP and MRH patient (i.e., if the participant arrived at the correct diagnosis for the SP, he also arrived at the correct diagnosis for the MRH).

These results show that participants performed equivalent exams on the SP and MRH patient. This demonstrates that 1) students are able to perform equivalent exams on MRHs and SPs, 2) the participants treated the physical exam of the MRH with the same importance which they treated the physical exam of the SP, spending similar time and effort to search for abnormalities in the patient's breast, and 3) because a participant's performance with the MRH was indicative of his performance with the SP, an interaction with the MRH patient allows medical educators to evaluate students' exam skills. This result *establishes the usability of MRH patients for practice and evaluation of students' clinical exam skills*.

8.3 Acceptability: Believability

Believability of the MRH and SP was measured with the MaSP (Table 1). A comparison of MaSP ratings with the MRH and SP is given in Table 2. *The MRH and SP were rated similarly on all items of the MaSP except for those pertaining to verbal communication.* The SP was judged to answer questions more naturally ($SP: 3.9 \pm 0.9$; $MRH: 3.1 \pm 0.9$; $F = 7.1$, $p < 0.05$; 4/10 thought the MRH answered naturally and 9/11 thought the SP answered naturally). The SP also did a better job at making the participant aware that the patient was listening (SP: 4.0 ± 0.7; MRH: 2.9 ± 0.8; $F = 10.8$, $p < 0.05$; 3/10 agreed that they could tell the MRH was listening; 9/11 could tell the SP was listening).

The overall rating (1-10) of the SP was greater than that of the MRH patient, but not significantly so (SP: 7.6 ± 1.6; MRH: 5.6 ± 1.6; $F = 4.5$; ns, $p > 0.07$). Though there appears to be a trend favoring the SP, the 95 percent CIs overlap, indicating noninferiority of the MRH rating—even so, it is possible that in a larger population the SP would be rated significantly higher. This is expected, as, logically, the

TABLE 2
Ratings of MRH and SP Believability

MaSP Item (5-point Likert item)	SP	MRH
Patient appears authentic	3.7 ± 1.3	3.3 ± 0.9
Patient challenges student	3.8 ± 0.8	3.4 ± 1.1
Patient simulates physical complaints unrealistically (higher score denotes disagreement)	3.9 ± 1.1	3.8 ± 0.4
Patient answers questions naturally**	3.9 ± 0.9	3.1 ± 0.9
Patient withholds information (high score denotes disagree)	3.9 ± 0.6	3.3 ± 0.7
Patient's appearance fits the role	4.3 ± 0.5	4.0 ± 0.7
Patient stimulates the student to ask questions	3.8 ± 0.8	3.2 ± 0.8
Student can tell that patient listens**	4.0 ± 0.7	2.9 ± 0.8
Patient communicates how she feels	3.4 ± 0.9	3.8 ± 0.8
Overall rating (1-10 scale)	7.6 ± 1.7	5.6 ± 1.6

Significant differences at $(\mathrm{p} < 0.05)$ *are denoted by* $(^{**})$.

SP should be superior—*she is human* and has greater capacity for verbal communication and rapport than does the MRH. However, this result demonstrates that an MRH interaction can approach the fidelity of a human interaction. Coupled with the MRH's ability to overcome drawbacks of SPs, this result suggests the MRH to be a valuable educational tool.

8.4 Acceptability: Perceived Educational Benefit

Participants rated the MRH and SP experiences as having similar overall educational benefit. The MRH and SP experiences were rated as being equivalently *worthwhile*: $\mathrm{MRH}(4.33 \pm 0.5)$, $\mathrm{SP}(4.33 \pm 0.7)$, $\mathrm{F} = 0.007$, $\mathrm{CI} = [-0.47, 0.47]$. All participants (10/10) agreed that the MRH experience was worthwhile and (10/11) participants agreed that the SP experience was worthwhile.

The MRH and SP experiences were both deemed to be good practice tools, with 10/10 participants agreeing that they would use the MRH system for practice and 10/11 participants agreeing that they would use the SP for practice. The CI of $[-0.8, 0.4]$ is consistent with the MRH and SP experiences being equivalent, but requires a larger population to determine statistical equivalence.

The MRH and SP experiences were rated similarly for most of the specific educational benefits. Participants felt that the MRH and SP experiences were equivalent for increasing their preparation to perform both a breast history (MRH: 8/10; SP: 10/11; $\mathrm{CI} = [-0.48, 0.26]$) and a CBE (MRH: 7/10; SP: 9/11; $\mathrm{CI} = [-0.49, 0.05]$). The MRH and SP were rated similarly for increasing confidence $(\mathrm{MRH} : 6/10; \mathrm{SP} : 7/11; \mathrm{F} = 1.11, ns)$ and decreasing anxiety $(\mathrm{MRH} : 5/10; \mathrm{SP}6/11; \mathrm{F} = 0.78, ns)$ in performing a CBE.

The SP was rated significantly higher at increasing the participant's confidence in taking a breast history $(\mathrm{SP} : 3.7 \pm 1.3; \mathrm{MRH} : 2.9 \pm 1.3; \mathrm{F} = 5.5, p < 0.05)$ and at decreasing the participant's anxiety over taking a breast history $(\mathrm{SP} : 3.7 \pm 1.3; \mathrm{MRH} : 2.9 \pm 1.3; \mathrm{F} = 5.5, p < 0.05)$. We believe that the higher fidelity verbal communication afforded by the SP is responsible for the SP being rated higher on these two items (explored in Section 8.6.2).

Overall, participants rated the believability of the MRH similarly to the SP, and found similar educational benefit of the MRH patient and SP. As SPs are the accepted gold standard for training clinical exams, this result *establishes the acceptability of MRHs for practicing clinical exam skills*.

8.5 Feedback

Participants ranked visual realism, verbal communication, and tactile realism in the order in which their improvement would most benefit usability of the MRH experience. This ranking was mapped to a three-point scale. A one-way ANOVA was run with the three categories as independent samples. A Tukey HSD test revealed that improving verbal communication was more important than improving visual realism ($p < 0.01$) or tactile realism ($p < 0.01$). There was no significant difference between the importance of improving visual and tactile realism.

8.6 Discussion

8.6.1 MRH and SP Provided Similar Experiences

The MRH patient provided a similar experience to the accepted standard for practicing and evaluating clinical exam skills, the SP.

Participants were able to perform equivalent clinical exams on the MRH patient and the SP, and rated the MRH and SP as being similarly believable and educationally beneficial. This demonstrates the usability and acceptability of an MRH patient for practice and evaluation of students' clinical exam skills.

The amount of time participants spent palpating the SP and MRH patients' breasts searching for abnormalities was significantly correlated. This indicates that not only did participants take the MRH experience seriously, but that participants placed similar value on evaluating the breast health of the MRH and SP. Participants treated the health of the MRH as equal importance as that of the SP. Supporting the interpersonal-touch result of Study I, it is again shown that people treat MRHs similar to humans.

It is exciting that the MRH experience is scored by users similarly to the accepted standard of the SP. One hundred percent of participants rated the MRH as worthwhile and reported that they would use the MRH for practicing their clinical exam skills. We have conducted 14 studies with over 300 health-profession students interviewing VH patients, and in none of those studies did 100 percent of the population rate the experience as a worthwhile educational tool. Experiences with MRHs seem to elicit excitement and interest from participants more often than VH experiences. We believe that the reason for the immediacy of MRH experiences is based in the ability to touch the MRH; this is an area that requires further exploration.

8.6.2 Differences of MRH and SP Experiences

One area in which the SP was rated more favorably than the MRH was verbal communication. This was expressed in three results: 1) The MRH's verbal communication abilities (appearing to listen to a question and then answering it naturally) were not rated as believable as the SP's verbal abilities. However, this is expected; to talk to the MRH, the participant's speech is filtered through

imperfect speech recognition and speech understanding modules. 2) Although 100 percent of participants reported they would use the MRH as a practice tool, participants identified verbal communication as the aspect that requires most improvement to make the MRH experience more desirable as a practice tool. 3) The perceived educational benefit of the MRH may also have been impacted by her verbal communication ability. The MRH was rated lower than the SP on two of the three items pertaining to the breast history, the portion of the experience relying heavily on verbal communication. We expect this to be related to participants not being able to build rapport with the MRH due to the lower quality verbal communication. However, the relationship is not demonstrated through any statistical correlations, and may be more complex.

It is interesting that the MRH was rated similarly to the SP in her ability to communicate *how she felt* during the interaction. This reflects on the MRH's nonverbal communication, such as grimacing when she experiences pain. The touch-driven communication of the MRH contributes to the ability of the MRH to express how she feels using both verbal and nonverbal communication. For example, she expresses fear as the participant first begins to touch her, and she expresses pain as a result of the participant's touch. It is this ability to react to touch which emphasizes the MRH's verbal shortcomings; *by removing the barrier to touching a virtual person, the user's expectations for human-like behavior increase*. The user's thought process may take the form of: "If the virtual person can react to my touch like a real person, why can't she answer my questions like a real person?"

8.6.3 Poststudy Reflections

Population size. The within-subjects design provided a total of 18 interactions with a simulated (SP or MRH) patient, but the population size is still limited. Differences and similarities in medical students' interactions with MRHs and SPs will be reexamined as part of a larger future study that examines the learning effects of interacting with MRHs.

Experience transfer. Although no effect of the order of students' interviews was found, it has previously been shown that experience transfers from real to virtual interactions [15]. Students' first interview in Study II likely affected some aspect of their second interview, but did not significantly affect performance or perceptions.

Limitations of the SP interaction. Students were unable to perform the visual inspection and palpation directly on the SP's breasts. Three SPs were needed to cover the time period of the study. The limited availability of SPs for intimate exams, the unavailability of SPs with breast abnormalities, and the time-availability constraints of SPs highlight the limitations of using SPs to teach intimate examinations and the CBE in particular. Combining SPs and simulators has been shown to be effective [25], [23], so our approach is expected to have produced results similar to if the CBE was performed directly on the SP. However, the use of a physical simulator for the SP's CBE makes the hands-on portions of the MRH and SP experiences more similar than if the CBE was performed directly on the SP's breasts.

Generalizability of results. Both Study I and Study II participants had a limited number of experiences (average of 2.2) with SPs and no experience with VH or MRH patients. The behavior and subjective perceptions of these participants may differ from that of more experienced health profession students (e.g., fourth-year medical students typically have more than 10 SP experiences).

9 Effects of Increased Visual Realism

To determine whether the improvement in the MRH patient's visual realism affected users' perceptions of the experience, a between-subjects comparison $(n = 18)$ of Study I and Study II was conducted using analysis of variance.

Study I and Study II populations were similar: 1) Study I and Study II populations had an equivalent amount of experience with SPs and CBEs and 2) the third-semester physician-assistant students of Study I and the second-semester medical students of Study II have received a similar amount of training in both clinical and interpersonal skills.

The main difference between Study I and Study II is that half of the Study II population conducted an SP interaction 20 minutes before the MRH interaction. Because no significant effects of order were found in Study II, we do not believe that the SP interaction affected the participants' perceptions and ratings of the MRH interaction of Study II. Although scenarios of Study I and Study II differed slightly in the MRH's complaint and family history, the most identifiable change and the only *improvement* to the experience was the increased visual realism. We hypothesize that any perceived improvement in the experience is primarily due to the increase in visual realism.

9.1 Believability

The visually improved MRH patient interviewed by Study II participants was rated as appearing more authentic than the MRH patient interviewed by Study I participants $(\text{Study II}: 3.4 \pm 0.84 > \text{Study I}: 2.4 \pm 1.2; \text{F} = 4.59, p < 0.05)$. Overall rating of the believability of the MRH as a patient (average MaSP score) was significantly higher for the visually improved MRH patient of Study II $(\text{Study II}: 3.5 \pm 0.24 > \text{Study I}: \ 2.8 \pm 0.73; \text{F} = 7.58, p < 0.05)$. The overall rating of the MRH patient of Study II was greater than that of the Study I patient $(\text{Study II}: 5.6 \pm 1.6 > \text{Study I}: 4.5 \pm 2.1; \ \text{F} = 1.6, ns)$. Although the difference is not significant, the ratings are statistically nonequivalent. Additionally, unlike the MRH of Study I, the overall rating of Study II's MRH patient fits within the range of values reported for VH patients [15]. Overall, it appears as though the believability of the MRH patient has increased with the increased visual realism.

9.2 Perceived Educational Benefit

Fisher's exact probability test was used to compare the number of participants in each of Study I and Study II that agreed with the measures stated in Section 4.4.2.

Study II participants reported greater perceived specific educational benefits (on all local measures) than did Study I participants. Although the differences were not significant, participants' ratings were found to be statistically nonequivalent.

Global measures of perceived educational benefit were significantly different between Study I and Study II. Significantly, more Study II participants reported the experience as being worthwhile (in Study II, 10/10 found the experience worthwhile versus 2/8 in Study I; $p < 0.005$).

More Study II participants reported that they would use the MRH system as a practice tool (Study II: 9/10 versus 3/8 in Study I; $p < 0.05$).

9.3 Importance of Domain-Specific Visual Realism

Previous work [13] has shown that in exploring a *general* virtual world, the visual realism of the VHs inhabiting the virtual world does not affect users' behavior or the believability of the VHs.

However, comparison of Study I and Study II shows that increasing the visual fidelity of the MRH patient significantly increased the believability and the perceived educational benefits of the experience. Believability and educational benefit are the two metrics used to evaluate the acceptability of the MRH experience. *Increasing the visual fidelity of the MRH patient increased the acceptability of the MRH experience as an educational tool*.

This demonstrates that the simulation of a CBE is an interpersonal scenario in which the visual realism of the simulation is critical. We expect that the importance of domain-specific visual realism will extend to other simulations of *specific* real-world scenarios.

10 Future Work

MRH patients will be created to simulate additional *intimate exams* such as pelvic and prostate exams. Intimate physical examination scenarios present an excellent platform for studying user anxiety with embodied agents. We have observed that performing intimate exams on human patients commonly causes students to become extremely anxious. The same type of anxiety response has been observed with the MRH breast exam patient. By making users aware of their anxiety (and other behavior) with MRHs, it may be possible to improve users' behavior with real people. We expect that through repeated interactions with MRHs, students will be able to decrease their anxiety in intimate exams with human patients.

11 Conclusions

MRHs are a new type of embodied agent affording touch-driven communication by combining a virtual person with a tangible interface. The additional interaction modality of touch allows MRHs to simulate interpersonal scenarios in which touch plays a critical role, such as medical physical examinations. The implementation of an MRH breast exam patient is presented.

Two user studies were conducted to provide an initial evaluation of user behavior with an MRH patient as well as the usability and acceptability of using MRH patients for practice and evaluation of health-profession students' clinical exam skills.

The observations of Study I participant behavior demonstrated that participants treated the MRH patient similarly to how they are taught to treat a human patient, on dimensions of empathy and use of comforting interpersonal touch. Comparisons of Study I participant behavior to that of participants interacting with a VH patient suggests that 1) medical students will treat an MRH patient more like a human patient than they will treat a VH patient, and that 2) participants treated the MRH more like a human patient because they were more engaged in the MRH experience than the VH experience. These observations suggest that the addition of touch afforded by the MRH increases the quality of the communication and the overall experience.

Study II evaluated the usability and acceptability of MRH patients for practicing and evaluating students' clinical examination skills. Usability was established by analyzing task performance on CBEs of each of an MRH patient and an SP. Participants performed equivalent exams on the MRH patient and SP, showing that 1) students can use MRH patients to practice exam skills, 2) medical educators can use results of students' exams of MRH patients to evaluate the students' skill, and 3) students treat the evaluation of an MRH patient and an SP's breast health as being of equal importance.

Acceptability was established through participant self-report of the believability of the MRH patient and the perceived educational benefits of the MRH experience. Believability and educational benefits of the MRH were rated similarly to the SP, the accepted *"gold standard"* for training clinical exams.

Through comparison of Study I and Study II participants' ratings of MRH acceptability, it was found that the increase in visual realism from Study I to Study II resulted in increased acceptability of the MRH patient. We believe that this highlights the need for visual realism in simulations of *specific* real-world interpersonal scenarios.

The results of Study I and Study II demonstrate the usability and acceptability of MRHs in medical education scenarios, and provide an example of how the touch-driven communication afforded by MRHs allows for the simulation of interpersonal scenarios in which touch is a critical component.

Acknowledgments

The authors wish to thank Kyle Johnsen, Andrew Raij, and Brent Rossen, who assisted in system development, and Toufic Imam, who assisted with user studies. This work was supported in part by US National Science Foundation (NSF) Grant IIS-0643557.

References

[1] R.H. Hill Jr., J. Gratch, S. Marsella, J. Rickel, W. Swartout, and D. Traum, "Virtual Humans in the Mission Rehearsal Exercise System," *Artificial Intelligence,* vol. 17, 2003.

[2] W. Swartout, J. Gratch, R.W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum, "Toward Virtual Humans," *AI Magazine,* vol. 27, no. 2, pp. 96-108.

[3] G. Frank, C. Guinn, R. Hubal, P. Pope, M. Stanford, and D. Lamm-Weisel, "Just-Talk: An Application of Responsive Virtual Human Technology," *Proc. Interservice/Industry Training, Simulation and Education Conf. (I/ITSEC '02),* Dec. 2002.

[4] S. Babu, T. Barnes, and L. Hodges, "Can Immersive Virtual Humans Teach Social Conversational Protocols?" *Proc. IEEE Int'l Conf. Virtual Reality (VR '07),* pp. 215-218, Mar. 2007.

[5] E. Deaton, C. Barba, T. Santarelli, L. Rosenzweig, V. Souders, C. Mc-Collum, J. Seip, W. Knerr, and J. Singer, "Virtual Environment Cultural Training for Operational Readiness (Vector)," *Virtual Reality,* vol. 8, no. 3, pp. 156-167, 2005.

[6] A. Manganas et al., "The Just VR Tool: An Innovative Approach to Training Personnel for Emergency Situations Using Virtual Reality Techniques," *J. Information Technology in Healthcare,* vol. 2, no. 6, pp. 399-412, 2004.

[7] K. Johnsen, R. Dickerson, A. Raij, C. Harrison, B. Lok, A. Stevens, and D.S. Lind, "Evolving an Immersive Medical Communication Skills Trainer," *Presence: Teleoperators and Virtual Environments,* vol. 15, no. 1, pp. 33-46, Feb. 2006.

[8] A. Raij, K. Johnsen, R. Dickerson, B. Lok, M. Cohen, M. Duerson, R. Pauly, A. Stevens, P. Wagner, and D. Scott Lind, "Comparing Interpersonal Interactions with a Virtual Human to Those with a Real Human," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 3, pp. 443-457, May/June 2007.

[9] F. Lang, R. McCord, L. Harvill, and D.S. Anderson, "Communication Assessment Using the Common Ground Instrument: Psychometric Properties," *Family Medicine,* vol. 36, no. 3, pp. 189-198, 2004.

[10] A. Kotranza and B. Lok, "Virtual Human + Tangible Interface = Mixed Reality Human. An Initial Exploration with a Virtual Breast Exam Patient," *Proc. IEEE Int'l Conf. Virtual Reality (VR '08),* pp. 99-106, 2008.

[11] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* fourth ed. Addison-Wesley, 2004.

[12] D.-P. Pertaub, M. Slater, and C. Barker, "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience," *Presence: Teleoperators and Virtual Environments,* vol. 11, no. 1, pp. 68-78, 2002.

[13] J.N. Bailenson, J. Blascovich, A.C. Beall, and J.M. Loomis, "Equilibrium Revisited: Mutual Gaze and Personal Space in Virtual Environments," *Presence: Teleoperators and Virtual Environments,* vol. 10, pp. 583-598, 2001.

[14] C. Zanbaka, P. Goolkasian, and L.F. Hodges, "Can a Virtual Cat Persuade You? The Role of Gender and Realism in Speaker Persuasiveness," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '06),* pp. 1153-1162, 2006.

[15] K. Johnsen, A. Raij, A. Stevens, D.S. Lind, and B. Lok, "The Validity of a Virtual Human Experience for Interpersonal Skills Education," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '07),* pp. 1049-1058, 2007.

[16] J.D. Fisher, M. Rytting, and R. Heslin, "Hands Touching Hands: Affective and Evaluative Effects of an Interpersonal Touch," *Sociometry,* vol. 39, no. 4, pp. 416-421, 1976.

[17] D.J. Dolin and M. Booth-Butterfield, "Reach Out and Touch Someone: Analysis of Nonverbal Comforting Responses," *Comm. Quarterly,* vol. 41, no. 4, pp. 383-393, 1993.

[18] P. Ellsworth, H.S. Friedman, D. Perlick, and M. Hoyt, "Effects of Direct Gaze on Subjects Motivated to Seek or Avoid Social Comparison," *J. Experimental Social Psychology,* vol. 14, pp. 69-87, 1978.

[19] J.G. Bruhn, "The Doctor's Touch: Tactile Communication in the Doctor-Patient Relationship," *Southern Medical J.,* vol. 71, no. 12, pp. 1473-1479, 1978.

[20] B. Insko, "Passive Haptics Significantly Enhances Virtual Environments," PhD dissertation, Dept. Computer Science, UNC-Chapel Hill, 2001.

[21] C. Basdogan, C. Ho, M.A. Srinivasan, and M. Slater, "An Experimental Study on the Role of Touch in Shared Virtual Environments," *ACM Trans. Computer-Human Interaction,* vol. 7, no. 4, pp. 443-460, Dec. 2000.

[22] J.N. Bailenson and N. Yee, "Virtual Interpersonal Touch: Haptic Interaction and Copresence in Collaborative Virtual Environments," *Int'l J. Multimedia Tools and Applications,* vol. 37, no. 1, pp. 5-14, 2007.

[23] R. Kneebone, J. Kidd, D. Nestel, S. Asvall, P. Paraskeva, and A. Darzi, "An Innovative Model for Teaching and Learning Clinical Procedures," *Medical Education,* vol. 36, pp. 628-634, 2002.

[24] L.A. Wind, J. van Dalen, A.M.M. Muijtjens, and J.-J. Rethans, "Assessing Simulated Patients in an Educational Setting: The MaSP (Maastricht Assessment of Simulated Patients)," *Medical Education,* vol. 38, no. 1, pp. 39-44, 2004.

[25] M. Duerson, J. Cendan, J. Woodard, and M. Hammoud, "Integrating Anatomical Simulators with Standardized Patients in a Clinical Skills Examination," *Poster Presentation at the Sixth Ann. Conf. of the Assoc. of Standardized Patient Educators,* 2007.

[26] S. Wellek, *Testing Statistical Hypotheses of Equivalence.* Chapman and Hall, Nov. 2002.

[27] D. Saslow et al., "Clinical Breast Examination: Practical Recommendations for Optimizing Performance and Reporting," *CA: Cancer J. Clinicians,* vol. 54, pp. 327-344, 2004.

[28] D.C. Hall et al., "Improved Detection of Human Breast Lesions Following Experimental Training," *Cancer,* vol. 46, pp. 408-414, 1980.

Aaron Kotranza received the BS degree in computer engineering from the University of Florida, Gainesville, in 2005, where he is currently a PhD candidate in the Department of Computer and Information Science and Engineering. His research interests include virtual humans, mixed reality, human-computer interaction, and computer graphics.

Benjamin Lok received the BS degree in computer science from the University of Tulsa in 1993, the MS degree from the University of North Carolina at Chapel Hill in 1997, and the PhD degrees in 2002. He is an assistant professor in the Department of Computer and Information Science and Engineering, University of Florida, Gainesville. He is an adjunct assistant professor in the Department of Surgery, Medical College of Georgia. His research areas include computer graphics, virtual environments, and human-computer interaction. He is a member of the IEEE.

Adeline Deladisma received the BS degree from Duke University in 1996, the MPH degree from Emory University in 2000, and the MD degree from Penn State University in 2004. She is currently a resident in the Department of Surgery, Medical College of Georgia, Augusta.

Carla M. Pugh is an assistant professor of surgery and the associate director of the Center for Advanced Surgical Education, Northwestern University, Chicago. She also holds an appointment in the School of Education at Northwestern. She has a broad interest in the use of technology for medical and surgical education and is especially interested in how medical professionals learn. In addition to her appointments at Northwestern, she also holds an appointment at the Telemedicine and Advanced Technology Research Center (TATRC) as a special assistant to the director. At TATRC, she manages the Advanced Distributed Learning Portfolio and the Medical Skills Proficiency area.

D. Scott Lind is the Jarrell distinguished professor and the chief of the Section of Surgical Oncology within the Department of Surgery, Medical College of Georgia, Augusta. He has a strong interest in medical education and a scientific approach to the study of educational problems. His educational research interests include the development and evaluation of new technologies, particularly simulation, in medical education.