Developing Virtual Patients with VR/AR for a Natural User Interface in Medical Teaching

Exploring Enhancements to Presence and Empathetic Communication

Marjorie A. Zielke, Djakhangir Zakhidov, Gary Hardee, Leonard Evans, Sean Lenox, Nick Orr, Dylan Fino, Gautham Mathialagan
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, TX, 75080
{ margez, dxz021000, ghardee, lxe160530, sxl105520, nao140030, drf150030, gxm141830 }@utdallas.edu

Abstract— **Professionalism and communication skills are important aspects of medical training, and virtual patient applications can offer cost effective, easily accessible platforms for communication practice which complement flexible, student-driven medical school curriculum design. Further, numerous virtual and augmented reality platforms have been introduced recently. This paper explores potential advantages and disadvantages Virtual and Augmented Reality (VR/AR) technologies offer to the development of a virtual patient application specifically for communication practice– the Emotive Virtual Patients – with a natural user interface. VR/AR technologies may offer highly interactive, immersive virtual patient experiences that tie to our research goals, improve presence and create a more fertile environment to practice empathy, however they may also present platform-specific challenges. A potential virtual patient design framework is discussed, and the unique benefits and limitations of VR/AR devices are analyzed. We put our research in the context of other virtual patient research, and hypothesize what benefits in terms of presence and natural user interfaces VR/AR may provide.**

Keywords— virtual patients; virtual reality; augmented reality; natural interface; natural language processing; HoloLens

I. Introduction

Learning good communication skills and developing the ability to show compassion for patients have long been considered important aspects of a physician's education [1]. Patients cite humaneness as the most highly rated aspect of care and express strong preferences for good communication and partnership [2]. Understanding a patient's non-verbal behaviors can provide important context to verbal information, such as the patient's emotional status, as well as open up opportunities to explore patient concerns or to express empathy [3]. The detection of patients' non-verbal clues can be defined, trained, evaluated and adapted for clinical specialties [3]. Additionally, attributes, values and skills that reinforce compassion and collaboration in healthcare delivery can be taught, modeled, learned and assessed when integrated into education at all levels and continuously reinforced [4].

This paper explores whether Virtual and Augmented Reality (VR/AR) technologies, can provide a more natural interface, increase presence and more highly facilitate the practice of empathy when compared to current medical education simulation forms. [5]. VR/AR technologies may afford the potential to take interactive, customizable 3D virtual patients to an enhanced level of interaction and realism, which will allow medical students to practice with more life-like communications scenarios anytime and anywhere. Many current virtual patient applications present linear, text-based conversations and focus more on biological education aspects presented through text interfaces or are monitor based. Robotic mannequins can be ineffective at simulating emotional or culturally accurate conversations [5]. Standardized patients, individuals who portray patients with medical conditions, can be effective, but accessing real humans for practice opportunities can be challenging and expensive.

Designing more realistic virtual patients requires a framework that comprises: a natural user interface aided by natural language processing; appropriate visual and auditory fidelity to represent an emotive virtual patient; the understanding of emotion through both sensing affective state of the user and translating this input into emotive/behavioral states for the patient; and a way to access the training modules asynchronously and without distractions typically associated with computer devices – tedious calibration and setup times, tethered wires, limitations of monitors and other computer peripherals. We aim to develop a natural, emotive virtual patient model that allows students to practice communication techniques in a manner that fits a self-driven curriculum environment

We acknowledge technology alone cannot result in a successful outcome, therefore the way it is implemented into the course curriculum by complementing other learning activities and social class engagement is important. However, we inquire as to whether VR/AR may increase user presence, defined as "a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part of all of the individual's perception fails to accurately acknowledge the role of the technology in the experience" [6].

Recently, there has been an influx of VR/AR technology platforms. VR head-mounted display (HMD) devices range from very inexpensive stereoscopic display devices, such as Google Cardboard, to computer and console systems such as the Oculus Rift, the HTC Vive and PlayStation VR. Some VR devices, like Samsung's Gear VR and Google's Daydream, use the processing capabilities of mobile smartphones. Google also launched an early attempt at a mobile AR with its Google Glass device. Microsoft HoloLens is an example of the evolution of AR head worn displays. Many others are being developed.

Designing realistic medical interviews that take advantage of the growing selection of VR/AR interfaces requires a framework that comprises: 1) high-fidelity virtual patients with enhanced, realistic, physical representations, i.e. visual, auditory, and behavioral; 2) a cognitive model to govern the virtual patient's emotions, logic and cultural representation; 3) a narrative system for customizable content presentation through patient stories; 4) educational scenarios with non-linear complexity; and 5) a natural user interface that replaces text with intrinsic human interactions – speech and head and body gestures.

VR/AR technologies could allow more realistic interactions with virtual patients by providing both more lifelike medical interviews as well as the ability to study how well students verbally and non-verbally demonstrate understanding and empathy. Empathy refers to "an affective response that is more consistent with another person's situation than one's own situation." [7]. Empathy is crucial for the success of social interactions, and has been linked to increased patient trust and patient satisfaction [7, 8].

A central research question is whether VR/AR interfaces, can enhance 3D virtual patients by heightening presence and improving representation of empathy to provide highly realistic communications-based educational experiences for medical school students. We also are exploring whether VR/AR offers advantages for the development of a natural user interface.

What follows is an analysis of both the potential and the limitations of VR/AR for use in healthcare communications education with a focus on a specific use case scenario: conducting a medical interview with a potential stroke patient and his caregiver.

II. Our early work – the UT TIME Portal

The UT Dallas Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab (The Center) has conducted research on the educational efficacy of several virtual healthcare simulations it has designed and developed. In 2014 and 2015, 74 pre-med students enrolled in a UT System program called TIME (Transformation in Medical Education) and participated in research on The Center's web and monitor-based adaptive and emergent research learning platform -- the UT TIME Portal. Students practiced demonstrating rapport, empathy and active listening during a medical interview with virtual patient Walter and his wife Susan as shown in Figure 1. Three trials of the platform were conducted at the University of Texas Southwestern Medical Center (UTSW). Findings from quantitative and qualitative data suggest that the UT TIME Portal platform complements classroom activities, helps improve knowledge regarding basic patient interviewing skills and positively influences attitudes about professionalism in social media use [9, 10]. These experiments, offered through what now might be considered a traditional delivery medium of web-based delivery through a laptop or desktop computer, provided a foundation for further research on the design of adaptive and emergent learning systems and suggested ways to customize the student's experience to complement other learning module activity.

The UT TIME Portal used in the research described above was designed for pre-med students who are just beginning their medical education. The simulation provides a medical interview scenario with branching conversations, but the interface is text- and monitor-based, and as such less immersive and natural. Based on feedback from post-survey responses and qualitative textual input, students found the text-based interface easy to use and helpful in learning basic interviewing skills, but they described the user experience as less realistic than real-life patient interactions. Many students wanted to be able to find out more information about the patient, such as past medical and social history, and the ability to ask their own questions. Several students wished that virtual patient could point to the source of pain: "I would have asked him to point to where the pain was the worse in the back of his head." The majority of students wanted better graphics for representing virtual patients. Some students suggested non-verbal techniques like using body language and eye contact as means to switch attention back to the patient from the overly helpful caregiver. Similarly, occasional eye contact could be used to make the caregiver not feel ignored. Also, patient nodding while caregiver is talking could be taken as a sign that information provided by the caregiver is important. Building on student feedback from this research, the Center's development team is working on a more natural user interface. As part of this work we are exploring whether our platform could benefit from VR/AR technologies to achieve its goals.

III. Specific Challenges of the UT TIME Portal and Opportunities for Enhancement

The continuation of the UT TIME Portal project, recently funded by Southwestern Medical Foundation will continue the specific-use case – a medical interview of a potential stroke victim– and will continue to integrate the framework and design elements to match the evolving curriculum at UTSW. The team will aim to develop two prototype Emotive Virtual Patients (EVPs) specifically focusing on an interface for natural user interactions – including speech and gestures; high-fidelity behavioral animations; and a customizable system for content presentation through patient stories. Feedback from subject matter experts suggested adding scalable complexity of scenarios; recognizing cues in the patient; capacity for ambiguity in conversations; methods of bringing attention back

Figure 1. Virtual Reality technology immerses learners completely in a synthetic environment. This mockup represents an interaction with two virtual characters, Walter and Susan, using Oculus - a VR head-mounted display device.

to the virtual patient; the ability to grade both the effectiveness of interview skills and the pleasantness/professionalism of the user. The anticipated outcomes are that medical students will improve and enhance their verbal and nonverbal communication skills. Success could be measured in the way EVPs will accurately represent emotions and unique patient cultures and integrate with the overall curriculum and specific communications. Finally, EVP sessions will be self-driven practice, an "anytime, anywhere" competency-based model that is flexible and has immediate feedback for adaptive learning by students. As defined and discussed further in the next section, we aim to create a natural user interface for this specific use case.

IV. Natural User Interface

A natural user interface (NUI) can be defined as an interface that allows the user to directly interact with a system without a physical controller [11]. The NUI offers direct system control through natural human actions [12]. Control over a NUI does not need to be learned like control over artificial control devices such as a mouse or keyboard [11]. For example, an NUI may be controlled by making a gesture or speaking a key phrase, both of which are innate human actions. Inversely, a new user must learn how to use a remote control before watching television. Another key aspect of the NUI is invisibility [13]. The interface should not be visible to the user while they are interacting with the system. The invisibility criterion ensures that users have direct control over the system without intermediate interfaces [13]. By leveraging intrinsic natural human action, NUIs provide users with efficient and accessible channels of direct interaction with a system or machine. The Turing test, developed in 1950, measured a computer's ability to communicate and exhibit intelligence equivalent to, or indistinguishable from humans [14]. We are essentially extending these same ideas through our research, but now focusing on the advantages that VR/AR brings to a NUI.

V. Challenges and Progress in developing Virtual Patients

Despite these component advances, researchers and developers of virtual patients often encounter similar challenges and barriers when faced with developing realistic, easy-to-use interfaces. Accordingly, these challenges must be kept in mind when exploring potential advantages of developing virtual patients for VR/AR platforms. Virtual patients are complicated systems that require considerable effort to develop. Complex virtual patient architectures may consist of many different modular components for user input, speech recognition, patient response, animation, audio, and lip-syncing that must work in tandem to present a realistic virtual human, potentially making the development of such a system costly and difficult. Virtual patient systems may also include branching narratives which further increase the complexity of the system and the effort required to develop it and because of this, robust branching narratives have historically been uncommon in virtual patients [15]. While complex systems offer more realistic interaction, their sophisticated architectures make authoring and adding content difficult, meaning in some instances that only developers with a deep understanding of the architecture can add new content [15].

These complexity issues have posed a barrier to entry and progression in the field, but recent development of toolkits and resources, such as the University of Southern California Institute for Creative Technology's (ICT) SimCoach [16] and the unified medical taxonomy [17], have begun to remedy this issue and make the development process less costly and time-intensive. Speech recognition is continuously improving – as a result, recent virtual patient platforms such as ICT's SimSensei Kiosk boast deep NLP capabilities and can parse open-ended vocal input from the user, automatically categorizing their speech and responding in real-time to simulate a conversational flow [18].

In the development of SimCoach and the SimSensei Kiosk, ICT utilizes a Virtual Human Toolkit (VHTK) that contains a framework that suggests what foundational components may be necessary to achieve our research goals. ICT's VHTK includes: **Multisense** –tracks and analyzes users' facial expressions, body posture, acoustic features, linguistic patterns and higher-level behavior descriptors (e.g. attention, fidgeting); the **NPCEditor** -- a statistical text classification algorithm that selects the character's responses based on the user's utterances; **the Nonverbal Behavior Generator (NVBG)** a rule-based system that analyzes character text and functional markup to propose nonverbal behaviors; **SmartBody** -- a character animation library that provides synchronized locomotion, steering, object manipulation, lip syncing, gazing and nonverbal behavior in real-time; and **vhtoolkitUnity** which includes a tight integration with SmartBody, a messaging protocol, debug and authoring tools, and a graphical timeline editor for creating cut-scenes [19].

Natural language processing (NLP) is an important factor in the development of a NUI. Much has been accomplished in the field of NLP, resulting in a recent proliferation of intelligent assistants such as Amazon Echo, Siri, and Google Home. Today's NLP research landscape consists of hundreds of companies and continues to grow [20]. Speech recognition is continuously improving and has been integrated into many widely used applications – smartphones now allow dictation of text messages through NLP, and Windows 10 comes with the Cortana assistant which accepts natural language input [21]. – In this landscape of continual research and improvement of NLP and NUI systems, it is plausible and realistic to look toward applying NLP/NUI to VR/AR virtual patient systems.

Despite these advances in extensive AI and language processing research, this area remains one of the greatest challenges faced when developing virtual patients. Accurate natural language processing and understanding is still very difficult to achieve. Inaccurate language processing can lead to a frustrating, immersion-breaking experience that fails to simulate the flow of a clinical interview, and low-performing language systems are common, with virtual patient prototypes demonstrating 60-75% accuracy in their language processing [17]. Virtual patients may be extensively programmed to recognize words and phrases and respond accordingly, but true language understanding, conversational awareness, and natural conversation flow remain largely out of reach [15]. In order to simulate a natural conversational flow, many current virtual patient systems use "pause-based" language processing which interprets and responds to the user's voice input when they stop talking for a certain length of time. This poses problems, as many phrases and statements may contain internal pauses that do not necessarily signal the end of the statement. More sophisticated language processing systems designed to deal with this problem may also struggle when a user does not pause between thoughts. Pause-based systems also suffer from latency as they measure and detect pauses before responding, slowing the flow of the conversation [22]. Some virtual patient systems circumvent the issue of natural language processing with lower-tech interfaces such as multiple choice menus. These structured encounters are more reliable and do report positive learning outcomes but lack the robust educational/training potential of natural language input [17].

VI. SYSTEMS FOR EMOTIONS BETWEEN VIRTUAL PATIENTS AND USERS

Much effort has been put toward understanding users' emotions [23] when interacting with computers and virtual agents. "Human beings are a largely untapped source of in-the-loop knowledge and guidance for computational learning agents, including robots" [24], and studying the ways humans teach each other and computers is now being pursued with vigor. However, the challenge of developing fully realistic, non-mechanical high fidelity EVPs with realistic and trustworthy personalities [25, 26], that can communicate naturally with the user still remains. Continuing this research appears to be worthwhile. Affective responses to the learning experience often result in lasting attitudinal changes for learners. A recent study from MIT Media Lab augmented the behavior of a "Social Robot Tutor" for children with an affective policy, and the results confirmed that children's affective states responded to the nonverbal behavior exhibited by the social robot [25]. EVP designs vary in their goals and implementation, and The Center's prior research indicates that the most successful ones are evidence-based and attached to a medical curriculum. For example, the two EVPs from the UT Time Portal, Walter and Susan from Figure 1 above, have shown significant impact on learning outcomes both through quantitative and qualitative measures [9]. Students empathized with and commented about a virtual character's feelings, and the majority of students displayed strong emotional responses to a gameplay episode in which they accidentally revealed private information about Susan and Walter [9].

VII. WHY VR/AR MAY BE A SOLUTION TO MORE NATURAL, EMOTIVE VIRTUAL PATIENTS

Empathetic communication is a key element of virtual patient systems. "The physician is supposed to satisfy the patient in a holistic manner or in other words, win the trust. It is this trust primarily, that needs to be assessed" [27]. Thus in addition to portraying accurate medical symptoms and conditions, the virtual patients should be affectively responsive to the user. Believable virtual patients with a natural user interface will undoubtedly help medical students improve their verbal and non-verbal communication skills. We therefore want to evaluate whether VR/AR gets us closer to our emotive virtual patient objective.

As shown in Figure 2, according to Milgram, Virtual Reality occupies one extreme of the "virtuality continuum" and is an environment in which the participant observer is totally immersed in, and able to interact with, a completely synthetic world [28]. Augmented Reality sits closest to the opposite end of the virtuality continuum, and refers to cases in which an otherwise real environment is "augmented" by means of virtual (computer graphic) objects. Mixed Reality is a term used to describe the varying types of devices that fall along this continuum [28].

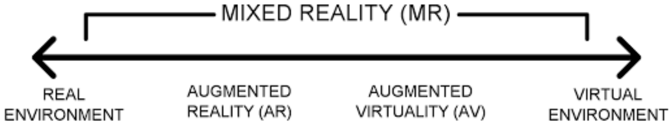

Figure 2. Virtuality Continuum [28].

VR/AR headsets offer nuanced alternatives to monitor-based interfaces. While VR (Gear VR, Oculus, Vive) headsets provide higher fidelity renditions and a fully embodied interactive experience, they can lack in ergonomics, potentially be disorienting to the user, and may have unresolved safety concerns. AR can be a powerful and successful alternative that removes some of the above-mentioned issues but can have its own shortcomings such as limited fidelity of holograms, limited field of view, and limited built-in CPU/HPU. Augmented Virtuality combines the VR experience inside the headset together with a physical space that perfectly matches the virtual environment's dimensions [28]. Furthermore, augmented reality platforms such as Microsoft Hololens's 3D audio helps achieve immersion by projecting accurate spatial sounds based on size and shape of the user's skull [29]. It is also possible to simulate an AR-like mixed reality experience inside a VR headset by feeding a live recording of the user's surroundings via head-mounted cameras [30]. Regardless of the approach chosen, enough attention must be paid to the cyber-psychologic experience of the user being driven by game mechanics such as plot, storyline, and sound design; intrinsic and extrinsic motivations; a system of rewards; and adaptive assessment instruments that may lead to positive behavior change outcomes.

VIII. Advantages and Disadvantages of VR and AR for Virtual Patients

Both VR and AR systems have their own unique "immersion" capability, or the capability to present a convincing virtual environment to the user [31]. Where virtual reality (VR) technologies can offer an immersive alternate reality, augmented reality (AR) technologies can supplement a student's existing reality to provide an in situ experience such that the physical world and all accompanying real and virtual objects are perceived to exist in the same locale. Virtual objects such as 3D holograms and web pages can be positioned within the physical environment and used to interact with a virtual patient that responds to the student's actions --further providing a sense that the virtual patient is present and exists in the real world.

In head mounted AR systems, the convergence of real and augmented environments is accomplished by allowing the wearer to see through the visor while AR techniques involving multiple sensors, spatial mapping, computer vision, and holographic positioning systems are implemented. Retaining the ability to see the natural environment allows the physical world itself to be leveraged in development as it does not need to be recreated, only supplemented. This could allow for a greater focus on the creation of a graphically detailed virtual patient and for more attention to be given to interface, in which real or virtual elements including the virtual patient could play a role. Conversely, in VR systems one cannot see the real world through the visor, which poses potential safety hazards. VR systems provide an opportunity to create an entire alternate environment for the student, which could be used by a designer to define a smaller or larger scope for what environmental elements should be realistically present in the project, as well as to implement and control the broader visual motif. A custom environment could offer a more consistent level of control over the student's educational experience while potentially providing a greater level of immersion.

AR can also be safer for the wearer in any existing environment as they can maintain their ability to visually navigate their natural surroundings. Additionally, though head mounted AR device options are currently limited, some of the emerging technologies are "untethered" and do not require physical wired connections to external equipment. The elimination of external wiring further diminishes potential safety issues and could allow a student to focus more intently

Figure 3. Augmented Reality technology can overlay synthetic objects onto a real-world environment. This mockup represents the possibility for projecting a virtual patient in a hospital room using a device such as the Microsoft HoloLens.

on the educational experience of interacting with a virtual patient. Untethered devices also reduce the time required to setup and start using a device, which could allow more flexibility regarding where and when a student may choose to learn or engage with a virtual patient depending on the device's solution for removing the wires. Some devices are untethered because they have an internal computer, while other emerging solutions for VR devices involve wirelessly connecting to an external computer providing a greater graphical capability. A gaming laptop for example, could allow the system to remain portable while potentially still delivering a high fidelity visual experience.

When presented with a completely virtualized reality, some users can experience disorientation or a version of motion sickness known as "simulator sickness." There are competing theories regarding the cause of the sickness, such as the sensory conflict theory and the postural instability theory [32]. Most head mounted AR solutions do not encapsulate the user or obscure access to timely and expected visual feedback of the real world, nor do they inhibit the wearer's ability to maintain postural control, which inherently negates the likelihood of encountering simulator sickness specifically. However, strategies and best practices have been suggested for developers when creating software for VR devices to mitigate simulator sickness by properly handling issues like latency, flicker, and acceleration, which have been identified as potentially contributing factors to simulator sickness [33].

While interaction with the natural environment using VR devices may be completely removed due to the encapsulating nature of the technology, AR devices allow for seamless interaction with both virtual and physical worlds. Some current AR systems are natively capable of using various predefined gestures as a natural interface to trigger interactions, as well as providing speech to text functionality that can further connect the wearer to the augmented realm through natural means. Virtual reality systems can support additions that parallel these functionalities and may have more processing power available for implementation as well, depending on the choice of external hardware. Sensors such as a microphone or a Leap Motion controller can be attached externally to some systems and software can be extended as needed to support the desired functionality. In this way, a VR device is further able to function as a natural user interface between the student, a virtual patient, and the virtualized environment.

Another feature aiding in the enhancement of some AR/VR systems as a natural user interface, as well as the creation of a sense of presence, is 3D audio. Research indicates that 3D spatialized audio is associated with higher reported "presence," or psychological sensation of physically being in a virtual space or with holographic imagery [31, 34]. Due to the Head-Related Transfer Function (HRTF) effect, everyone hears sounds differently [28]. This effect occurs because sound waves interact with other parts of human anatomy along the way into the ear canals. To account for HRTF, tiny microphones are positioned inside the ear canals of a model or a human listener and samples are taken from many angles inside an anechoic chamber. Research indicates that HRTF audio increases users' reported presence [35]. However, most 3D audio implementations for wearable AR/VR solutions draw from existing HRTF libraries that may only be populated with measurements taken every 5 or 15 degrees [36]. The result is that even after the required calibration of a device, the selected HRTF set may result in audible discontinuities that diminish the wearer's ability to locate virtual objects in 3D space based on sound.

In addition to the above-described potential VR/AR configurations, the HoloLens, represented in Figure 3 above, is a potentially applicable AR solution in the implementation of a natural interface between the virtual patient, student, and the environment. An interesting feature of HoloLens specifically, is that the calibration process for spatial audio is invisible to the user while an algorithm uses the distance between the user's eyes as an indication of what the distance between their ear canals may be [29]. Personalized 3D audio combined with head tracking can allow a student to physically roam the mixed reality environment and gaze naturally in any direction while remaining able to locate and interact with a virtual patient using audible and visual cues. The position of the student in the space and the direction of their gaze not only changes what they hear, but creates a natural interface by providing input to the virtual patient who can then respond accordingly with feedback that is meaningful to the student or changes the course of the educational experience.

Microsoft's HoloLens is one example platform that allows us to explore the potential of AR. The device is an untethered all-in-one AR solution that is passively cooled, has an onboard battery, and has processor and memory capabilities more similar in performance to current mobile technologies than to the performance of the gaming machines typically paired with VR devices. While this would likely result in a less graphically intensive application, there are some core features unique to the HoloLens allowing it to be leveraged as a natural user interface that achieves a level of engagement suitable for a virtual patient project.

One such feature is its Holographic Processing Unit (HPU). The HPU is a dedicated custom processor that integrates data supplied by the device's sensors, most notably the accelerometers for motion detection, and the camera system for depth detection, spatial mapping, gesture detection, and hologram positioning. Many AR technologies are in development or in an emerging state such that specific metrics are not readily available. However, it has been reported that Microsoft indicated the HPU is 200 times faster than a pure software (CPU) equivalent [37]. The result of the inclusion of an HPU as a dedicated "sensor processor" is the virtual removal of all holographic "jitter" conventionally experienced in computer vision implementations on current software based AR systems. The fluidity and consistent positioning of a holographic virtual patient would again contribute to the feeling that the patient is present in the room.

A caveat, however, is that the field of view (FOV), is narrower compared to some other current VR and AR solutions.

While there are indications that Microsoft Research may be exploring approaches to mitigating FOV shortcomings in some AR/VR devices using lower resolution peripheral modifications [38], there is no information currently suggesting this technique will be used in the HoloLens specifically. This does pose a potential problem when considering the HoloLens for use as a natural interface, as it may be difficult to position a virtual patient in a student's vicinity while maintaining proper visibility and a distance that is appropriate for a realistic doctor-patient interaction.

In addition to limited FOV, other caveats to HoloLens adoption currently exist and may require consideration. One potential drawback to the HoloLens as a development platform is that the app must be built as a Universal Windows Platform (UWP) application [39]. It should be considered that the process of first building in Unity and then outputting to Visual Studio before deployment as a UWP app adds some complexity to the development process. Furthermore, there are rules to follow if the application is destined for the Windows Store [40]. Currently, not all features and APIs of a UWP app are available for HoloLens [41]. In addition, the HoloLens is geared to the development community. In January of 2017 the cost of the HoloLens is $3000 for developers [42]. Finally, reported short battery life [43] could introduce additional potential system limitations.

Native gesture and native NLP support on the HoloLens need further research and evaluation for customizability and applicability to a virtual patient project specifically. However, support for this functionality can be added depending on project goals and the availability of external supplemental libraries. Having onboard speakers, a microphone, and a camera for gesture detection avoids additional setup time for the user and eliminates the need to attach ad hoc sensors to the device, allowing the student to maintain focus on the educational experience.

The state of available and emerging AR technologies is such that a choice currently has to be made between devices that specialize in hologram tracking and spatial positioning or devices that offer a greater field of view coupled with hardware selected based on the graphical and performance needs of the project.

IX. The HoloLens As a Potential Solution to EVP Challenges

Augmented reality platforms such as Microsoft's HoloLens may present a viable solution for the many challenges of creating believable EVPs. Due to the built-in gestures and native NLP support, the all-in-one untethered device approach, the advanced 3D spatial sound and the presence of a holographic processor unit (HPU), devices such as the HoloLens might make it possible to further advance the EVP interaction experience. While the EVPs can arguably be better rendered in the current VR headsets, an AR device allows them to seamlessly integrate with the user's natural environment. The personalized 3D audio combined with head tracking will allow a student to physically roam the mixed reality environment and gaze naturally in any direction while remaining able to locate and interact with a virtual patient using audible and visual cues. The HRTF calibration, invisible to the user, ensures that sounds are projected more accurately via the HoloLens's speakers than with the typical headphone solutions on VR headsets.

Devices such as the HoloLens may also offer novel solutions for the challenge of directing user's attention back to the virtual patient or as needed. Spatially accurate sound cues together with visual cues and indicators may be used to direct the viewer's attention while gaze tracking and computer vision may be used to better track the user's attention. The untethered, all-in-one, asynchronously accessible solution such as offered in HoloLens may integrate well with medical curriculum and aid in the self-driven "anytime, anywhere" practice of medical students. AR platforms such as the HoloLens could result in a more natural and realistic conversation experience with a virtual patient because of holistic technology.

X. Summary

VR/AR technologies may advance the field of NUI research and afford the potential to make interactive, customizable 3D virtual patients effective for practicing communication skills in medical education. A VR system affords potentially greater processing power for language processing and rendering high fidelity EVPs, while an AR system such as HoloLens provides an immersive holographic, holistic experience and supports built in NLP and gesture support with no external computer. The innumerable challenges of developing intelligent machines that appear to think and behave like humans, as framed by Turing back in 1950, still persist. We frame measures of a NUI, empathy and presence as a way to calibrate progress in this regard. We will continue to work on our central research question – whether VR/AR can enhance interactions with 3D virtual patients and, ultimately, complement or replace, current medical simulation methodologies for practicing communication with a more natural interface. However, this early research suggests that VR/AR technology has potential to move us substantially forward toward that goal.

Acknowledgment

The research team would like to thank colleagues from UTSW Medical Center, Southwestern Medical Foundation and the students from the UT TIME program for their support of this research.

References

[1] J. M. Corrigan, M. S. Donaldson, I. of Medicine, Committee on Quality of Health Care in America, and Committee on Technology for a Quieter America, *To err is human: Building a safer health system*, L. T. Kohn, Ed. Washington: National Academies Press, 2000.

[2] A. Schattner, "The silent dimension," in *Archives of Internal Medicine*, vol. 169, American Medical Association, 2009, pp. 1095–1099. [Online]. Available:

http://jamanetwork.com/journals/jamainternalmedicine/fullarticle/773524. Accessed: 2016.

[3] B. Molinuevo, R. M. Escorihuela, A. Fernández-Teruel, A. Tobeña, and R. Torrubia, "How we train undergraduate medical students in decoding patients' nonverbal clues," *Medical Teacher*, vol. 33, no. 10, pp. 804–807, Sep. 2011.

[4] "Recommendations from a Conference on Advancing Compassionate, Person- and Family-Centered Care Through Interprofessional Education for Collaborative Practice", Atlanta, 2014.

[5] M. A. Zielke, J. LeFlore, F. Dufour and G. Hardee, "Game-Based Virtual Patients – Educational Opportunities and Design Challenges", in *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2010*, Orlando, Florida, 2010.

[6] M. Lombard, "Resources for the study of presence.," 2000. [Online]. Available: http://www.presenceresearch.org.

[7] E. Mavromihelaki *et al*., "Cyberball3D+: A 3D Serious Game for fMRI Investigating Social Exclusion and Empathy," *2014 6th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, Valletta, 2014, pp. 1-8.

[8] A. Kleinsmith, D. Rivera-Gutierrez, G. Finney, J. Cendan, and B. Lok, "Understanding empathy training with virtual patients," *Computers in Human Behavior*, vol. 52, no. C, pp. 151–158, Jan. 2015. [Online]. Available: http://dl.acm.org/citation.cfm?id=2822639.

[9] M. A. Zielke, D. Zakhidov, G. Hardee, and S. Lenox, "Using Qualitative Data Analysis to Measure User Experience in a Serious Game for Premed Students", in *International Conference on Virtual, Augmented and Mixed Reality*, Springer International Publishing, 2016.

[10] M. A. Zielke, D. Zakhidov, D. Jacob and G. Hardee, "Beyond Fun and Games: Toward an Adaptive and Emergent Learning Platform for Pre-Med Students with the UT TIME Portal," *in IEEE SeGAH 2016*, Orlando, Florida, 2016.

[11] A. Raffii, T. Zuccarino, "Method and system enabling natural user interface gestures with user wearable glasses," U.S. Patent 9310891, Apr 12, 2016.

[12] C. Falcao, A. C. Lemos, and M. Soares, "Evaluation of natural user interface: A usability study based on the leap motion device," *Procedia Manufacturing*, vol. 3, pp. 5490–5495, 2015.

[13] "Natural User Interface," in NUI Group, 2011. [Online]. Available: http://wiki.nuigroup.com/Natural_User_Interface. Accessed: Nov. 11, 2016.

[14] A. TURING, "I.—COMPUTING MACHINERY AND INTELLIGENCE", *Mind*, vol., no. 236, pp. 433-460, 1950.

[15] T. Talbot, K. Sagae, B. John and A. Rizzo, "Sorting Out the Virtual Patient", *International Journal of Gaming and Computer-Mediated Simulations*, vol. 4, no. 3, pp. 1-19, 2012.

[16] "SimCoach," 2013. [Online]. Available: http://www.simcoach.org/. Accessed: Jan. 2017.

[17] Combs, C. Donald, John A. Sokolowski, and Catherine M. Banks, eds. *The Digital Patient: Advancing Healthcare, Research, and Education*. John Wiley & Sons, 2015.

[18] D. DeVault *et al.*, "SimSensei kiosk," International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068. [Online]. Available: http://dl.acm.org/citation.cfm?id=2617388.2617415.

[19] "Virtual human Toolkit," 2017. [Online]. Available: https://vhtoolkit.ict.usc.edu/. Accessed: Dec. 2016.

[20] D. Miller, "Intelligent assistant landscape shows slow growth but huge potential", *VentureBeat*, 2016. [Online]. Available:http://venturebeat.com/2016/02/14/intelligent-assistance-the-slow-growth-space-that-will-eventually-wow-us/. [Accessed: 22- Nov-2016].

[21] "What is Cortana?", *Microsoft Support*, 2016. [Online]. Available: https://support.microsoft.com/en-us/help/17214/windows-10-what-is. [Accessed: 29- Nov- 2016].

[22] R. Manuvinakurike, M. Paetzel, C. Qu, D. Schlangen and D. DeVault, "Toward incremental dialogue act segmentation in fast-paced interactive dialogue systems", in 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2016.

[23] M. Valstar, M. Mehu, Bihan Jiang, M. Pantic and K. Scherer, "Meta-Analysis of the First Facial Expression Recognition Challenge", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 966-979, 2012.

[24] W. Knox, B. Glass, B. Love, W. Maddox and P. Stone, "How Humans Teach Agents", *International Journal of Social Robotics*, vol. 4, no. 4, pp. 409-421, 2012.

[25] G. Gordon, et al., "Affective Personalization of a Social Robot Tutor for Children's Second Language Skills", in *Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, 2016.

[26] N. Jaques, D. McDuff, Y. Kim and R. Picard, "Understanding and Predicting Bonding in Conversations Using Thin Slices of Facial Expressions and Body Language", in *International Conference on Intelligent Virtual Agents*, 2016, pp. 64-74.

[27] P. Gupta, "Assessment in Medical Education: Time to Move Ahead", in *Annals of the National Academy of Medical Sciences*, 2015, pp. 156-165.

[28] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays", in *IEICE TRANSACTIONS on Information and Systems*, 1994, pp. 1321-1329.

[29] M. Lalwani, "3D audio is the secret to HoloLens' convincing holograms," 2 November 2016. [Online]. Available: https://www.engadget.com/2016/11/02/microsoft-exclusive-hololens-spatial-sound/. [Accessed November 2016].

[30] W. Steptoe, "AR-RIFT," William Steptoe, 20 November 2013. [Online]. Available: http://willsteptoe.com/post/66968953089/ar-rift-part-1. [Accessed November 2016].

[31] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality: Abstract: Nature reviews Neuroscience," *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 332–339, Apr. 2005. [Online]. Available: http://www.nature.com/nrn/journal/v6/n4/abs/nrn1651.html.

[32] D. M. Johnson, "Introduction to and Review of Simulator Sickness Research," Defense Technical Information Center, 2005.

[33] Oculus VR, LLC, "Simulator Sickness," 2016. [Online]. Available: https://developer3.oculus.com/documentation/intro-vr/latest/concepts/bp_app_simulator_sickness/. [Accessed November 2016].

[34] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 5, no. 3, pp. 290–301, Jan. 1996. [Online]. Available: http://dl.acm.org/citation.cfm?id=2871028.2871031.

[35] R. S. Pellegrini, "Quality assessment of auditory virtual environments," Georgia Institute of Technology, 2001. [Online]. Available: https://smartech.gatech.edu/handle/1853/50631.

[36] Oculus VR, LLC, "3D Audio Spatialization," 2016. [Online]. Available: https://developer3.oculus.com/documentation/audiosdk/latest/concepts/audio-intro-spatialization/. [Accessed November 2016].

[37] P. Bright, "Microsoft sheds some light on its mysterious holographic processing unit," Condé Nast, 23 August 2016. [Online]. Available: http://arstechnica.com/information-technology/2016/08/microsoft-sheds-some-light-on-its-mysterious-holographic-processing-unit/. [Accessed November 2016].

[38] R. Xiao and H. Benko, "Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays," in *The 2016 CHI Conference*, 2016.

[39] "Exporting and building a unity visual studio solution,". [Online]. Available: https://developer.microsoft.com/en-us/windows/holographic/exporting_and_building_a_unity_visual_studio_solution. Accessed: Jan. 2017.

[40] "Windows Store Policies - Windows App Development," in *Windows Dev Center*. [Online]. Available: https://msdn.microsoft.com/en-us/library/windows/apps/dn764944.aspx. Accessed: Jan. 2017.

[41] "Current Limitations For Apps Using Apis From The Shell," in *Windows Dev Center*. [Online]. Available: https://developer.microsoft.com/en-us/windows/holographic/current_limitations_for_apps_using_apis_from_the_shell. Accessed: Jan. 2017.

[42] "Order Options," in *Microsoft HoloLens*. [Online]. Available: https://msdn.microsoft.com/en-us/library/windows/apps/dn764944.aspx. Accessed: Nov. 2016.

[43] B. Gilbert, "Microsoft's futuristic head-mounted computer has a major limitation," in *Business Insider*, Business Insider, 2016. [Online]. Available: http://www.businessinsider.com/hololens-battery-life-is-just-2-to-3-hours-2016-2. Accessed: Jan. 2017.