The Impact of a Mixed Reality Display Configuration on User Behavior with a Virtual Human

Kyle Johnsen[1], Diane Beck[2], and Benjamin Lok[3]

[1] Faculty of Engineering, University of Georgia
[2] College of Pharmacy, University of Florida
[3] Department of Computer Information Science and Engineering, University of Florida
kjohnsen@uga.edu, beck@cop.ufl.edu, lok@cise.ufl.edu

Abstract. Understanding the human-computer interface factors that influence users' behavior with virtual humans will enable more effective human-virtual human encounters. This paper presents experimental evidence that using a mixed reality display configuration can result in significantly different behavior with a virtual human along important social dimensions. The social dimensions we focused on were engagement, empathy, pleasantness, and naturalness. To understand how these social constructs could be influenced by display configuration, we video recorded the verbal and non-verbal response behavior to stimuli from a virtual human under two fundamentally different display configurations. One configuration presented the virtual human at life-size and was embedded into the environment, and the other presented the virtual human using a typical desktop configuration. We took multiple independent measures of participant response behavior using a video coding instrument. Analysis of these measures demonstrates that display configuration was a statistically significant multivariate factor along all dimensions.

Keywords: virtual humans, embodied agents, human-centered computing, medicine, displays.

1 Introduction

A primary aspect of a mixed reality environment is the integration of the real and virtual worlds, i.e. the virtual portion of the world appears as an extension of the real world. This means that the virtual environment appears at life-size (correct proportions given the content and surrounding real environment), and is embedded (the display placed such that it makes sense in the real environment). Mixed reality may be powerful for embedding a virtual human (VH) agent into the real world (e.g. as a virtual patient, soldier, or tutor); however, there is little empirical evidence comparing alternatives (e.g. using a standard desktop PC environment). In this paper, we discuss our experiment and results in evaluating if a mixed reality environment is justified for a VH application: provider-patient communication skills training with VH agents.

For the evaluation, we focused on an important aspect of provider-patient communication training: training providers to react appropriately to challenges from lpatients (e.g. difficult questions). The hypothesis was that a mixed reality training environment

J. Allbeck et al. (Eds.): IVA 2010, LNAI 6356, pp. 42–48, 2010.

would elicit different behavior along critical social dimensions from trainees than would a typical desktop-based training environment.Totest this hypothesis, we compared two configurations (Described in more depth in Section 4.2).

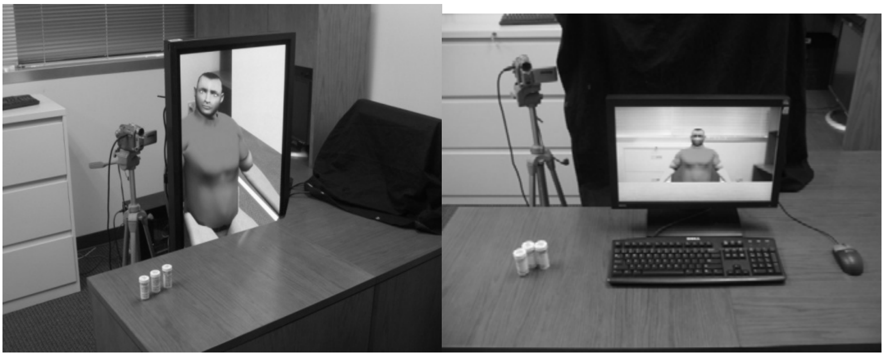

Fig. 1. The same VH displayed on a plasma television (Left) and a monitor (Right). A video camera records the experience for behavioral coding.

- Configuration PTV- A plasma TV (42” diagonal, oriented vertically) displayed the VH at life-size scale relative to the user. The display was placed in a chair across a desk from the user without a keyboard or mouse.
- Configuration MON- A typical LCD monitor (22” diagonal) displayed the VH at a smaller than life-size scale relative to the user. The display was placed on a desk in front of the user with a keyboard and mouse in a typical configuration.

2 Related and Previous Work

To generate a life-size VH relative to the user, mixed reality experiences may incoporate large-screen displays (projectors or large-screen televisions) [1, 2]. These large-screen displays show full-body life-size scale VHs. The use of a large-screen display may be justified from a cognitive perspective; researchers in media psychology have shown that, for *passive* media (e.g. television, movies), there is a strong positive correlation between imagery size and emotional response. People have a more powerful reaction to imagery on large screen displays [3]. Large screens also create a higher level of arousal and can amplify the effect of arousing imagery [4]. Furthermore, large screen displays motivate people to evaluate images of other people more favorably [5]. It seems reasonable then, that larger VHs would be more engaging to the user.

However, the reaction of a user to a VH is complex. There is the phenomenon of the uncanny valley [6], where, as an artificial entity approaches human form, there is a dramatic drop in acceptance. Also, some have found evidence that a mismatch between aesthetic realism and behavioural realism reduces copresence [7]. These theories motivate the current work, towards understanding how user behavior with a

VH is influenced by changing how the VH is displayed while controlling VH appearance and behavior.

Evaluating the effectiveness of visual display alternatives is a common task for new training applications. For applications involving spatial tasks (e.g. navigating a virtual environment, manipulating a spatial dataset), there exists an abundance of literature [8-11]. However, it is not clear how to generalize evaluations of displays for spatial tasks to the social tasks of VH experiences. In VH experiences, locomotion and object manipulation are less emphasized relative to verbal and non-verbal communication and the emotional aspects of that communication.

Closely tied to visual display, the level of physical immersion afforded by a VE system may amplify aspects of social communication with VHs. In one study, an immersive head-mounted display amplified the effect of user anxiety when speaking to an audience of VHs [12]. In other work, immersed CAVE display users took a leadership role over non-immersed small monitor display users in a collaborative virtual environment task [13]. Finally, a previous study on the system discussed in this work compared two virtual reality displays, an HMD and a fish-tank projection display [14]. Participants who interacted with a VH through an HMD were significantly more likely to self-rate their use of empathy higher. However, significant behavioral differences were not found between the two groups. This result directly led to the current work, comparing mixed reality configurations to easier-to-implement desktop configurations, that we hypothesize is a larger factor in user behavior.

3 Communication Skills Training Platform

The InterPersonal Simulator is a software and hardware platform designed to support interpersonal scenario training [1]. The InterPersonal Simulator shares many similarities to modern video game engines such as sound, rendering and animation support, scripting, and built in simulations (e.g. gazing and breathing). In addition, the InterPersonal Simulator supports user-VH conversations through speech and gesture recognition and a natural language script building system.

The open source, object-oriented rendering engine (OGRE) was used for rendering and animation and the FMOD sound system supported audio mixing and playback. The VH models were animated through a combination of vertex and skeletal animation. Recorded audio was used for the VH's voice.

A marker-based optical tracking system (2-Camera NaturalPoint Optitrack) was employed to track user head motion. Tracking head-motion enables a more immersive presentation of a virtual environment on a single-screen display by changing the viewing perspective to coincide with the user's head location, called fish-tank VR [15]. Both displays used in the study employed fish-tank head tracking for a immersion.

While normally the VH in the InterPersonal Simulator is driven solely by an autonomous agent, for the study VH responses were selected by a Wizard-of-Oz (WOz) operator controlling the VH via a hidden terminal and assisted by an autonomous agent system. The WOz could only initiate scripted responses and animations, and as such was not a true WOz (i.e. the wizard could not generate an arbitrary response from the VH). *Participants were told they were interacting with an autonomous agent through speech recognition.* Additionally, they were shown a tutorial video that taught them how to recognize when the patient did not recognize their speech accurately.

4 User Study

For the study, a VH experience was developed using the model of the peptic ulcer disease case [16]. The VH simulates a 35-year-old Caucasian male patient, Vic, who has come into an ambulatory clinic complaining about increasing pain in his abdominal region for the past month. A nurse and dietician had already seen Vic. Vic is anxiously waiting to be seen by the physician. Before the physician sees Vic, the student-pharmacist is asked to interview Vic.

Population. The VH patient was integrated as part of a clinical assessment exercise at the University of Florida's College of Pharmacy's doctorate program for working professionals. As part of a clinical assessment exercise the student the student is given a task to complete with a patient, typically a medical and medication history interview. For the experiment, thirty-nine students (12 men, 27 women) from the pharmacy program were recruited. The average age of the participants was 41.2 years old (min=26, max=65, σ=8.65). Also, the population was culturally diverse (11 Asian/Pacific Islander, 11 African-American, 2 Hispanic, and 15 Caucasian).

Procedure. Participants first filled out a background survey, conducted speech recognition training, and watched tutorial video. Then, the participant performed the interview. Participants were instructed to take less than fifteen minutes (the time given for standard clinical practice assessments) for the interview. The experimenter acting as the WOz was hidden in a far corner of the room behind a large desk, and was only able to hear the participant, simulating a human-quality speech recognition system.

Independent Variable. Participants were divided into two groups, PTV (using the plasma TV configuration), and MON (using the monitor configuration). The PTV condition (See Figure 1 left) was designed to create the mixed reality illusion of a person seated across a desk from the user, deemphasizing the display. By orienting the plasma TV vertically, the upper body of the VH (torso, arms, and head) was closely framed. The plasma TV was then placed in a standard desk-chair and placed behind a desk. A picture was taken of the area directly behind the plasma TV, and used as a texture for the background of the VE. This configuration is mixed reality in that it leveraged and extended the real environment (chair, desk, background).

The MON condition (See Figure 1 right) was designed to appear similar to a typical computer interface. The monitor was placed on the desk, with the VH displayed as sitting behind a virtual desk. Further, a keyboard and mouse were placed in front of the monitor (although these were non-functional). The same picture was used for the VE background texture. This configuration is not mixed reality because it was not designed to leverage the real environment.

The plasma TV (native 720x480) and monitor (native 1280x1024) had different native resolutions (although both were driven at 1024 x 768). The difference was not expected to influence user behavior, as the features of the virtual human were clearly visible in both displays. In addition, no participants commented on the resolution of the displays.

Measures. The dependent variable in this study was the behavior of students in response to two important moments during the interaction:

- **Moment 1 (M1):** The first moment occurred when the pharmacist entered the room, and introduced herself as the pharmacist. The patient (rudely) demanded, "Why aren't I speaking to the doctor?" This type of moment is a common occurrence for a pharmacist working in a clinical setting. Pharmacists are trained to respond pleasantly, to explain the purpose of the interview, and to comfort the patient.
- **Moment 2 (M2):** Around seven minutes into the interview, the patient said to the participant: "my dad died of cancer" and asked, "could this be cancer?" This moment was designed to evoke an empathetic response from the participant.

Participants' interactions with the VH were recorded using a video camera. The video camera was positioned next to the display and recorded the participant's face. The participant's face was recorded because facial expressions encode up to 60% of nonverbal information [17].

Five video evaluators independently rated all videos. Evaluators were blind to the condition of each participant (video was from the front and did not record the display condition) A video evaluation form was designed to rate the response of the participant to each critical moment. The instrument assessed four interrelated constructs, determined by the authors as likely to be influenced by display configuration and important to patient interactions:

- Engagement – how much the response indicated interest and involvement in the patient's problems, i.e. not disinterested
- Empathy – how much the response indicated that the participant attempted to understand the feelings of the patient, i.e. not inconsiderate
- Pleasantness – how friendly the participants' behavior was, i.e. not rude
- Naturalness – how much the participant's behavior was realistic, i.e. not robotic

5 Results and Discussion

The average correlation among video evaluators was medium (0.614), and a factor analysis of the average data showed a single significant (Eigen value > 1.0) factor for each moment. This means that each item in the rating scale was highly interrelated. The behavioral data was analyzed using a multivariate ANOVA for each critical moment survey. There was a significant multivariate effect of display type for both critical moments (Critical Moment 1 Wilks' λ=.59 p=.02, Critical Moment 2 λ=.612 p=.01). As seen in Figure 2, participants in the PTV condition were more engaged, empathetic, pleasant, and natural.

The video evaluation results suggest that the participants in the mixed-reality condition demonstrated more appropriate responses towards the VH patient than participants in the desktop condition. Participants in the mixed reality condition were rated

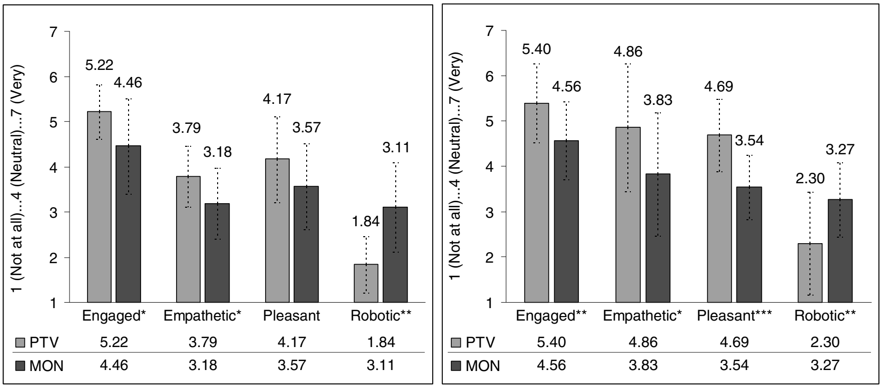

Fig. 2. Results for (Left) Moment 1 ("Why aren't I speaking to the doctor?") and (Right) Moment 2 (Could this be cancer?"). The dashed lines are +/- one standard deviation * (p<.05) ** (p<.01) *** (p<.001).

as significantly more engaged, empathetic, pleasant, and natural. In the desktop condition, participants appeared disconnected from the social interaction, and focused more on the general task (taking a medication history) than on the VH. A single dominant factor in the 4-item rating scale for each moment, suggesting that what observers were measuring was a single higher-level construct, such as appropriateness.

The raw magnitude of engagement is also worth noting. Participants in both the MON and PTV conditions were found to be engaged, but there was a main effect of display type in both M1 (p=.03) and M2 (p=.03), showing participants in the PTV condition were significantly more engaged. The reason for the high level of engagement in both conditions may have been the use of the WOz, and the isolation of the participant and the VH. During the interview, the participant was left alone in an office environment. With minimal interruptions occurring from technology, such as errors in speech recognition and outside influences, participants were engaged throughout the experience.

Overall, the results show that the visual display configuration is an important component to human-VH interactions, particularly those that may rely on or assess how the user responds to the VH. This work demonstrated that display configuration can have a strong influence on both cognition and behavior, and this result has important implications for designers of VH experiences. Designers should be aware of the limitations of small desktop display configuration. These types of display configurations are easily accessible and low-cost; however, interpersonal communication skills go beyond simple procedures. Interpersonal communication involves complex social and emotional behavior. Users may not treat VHs on a small monitor as they would a life-size VH. As a result, small monitor based VH experiences may be limited for the evaluation and training of interpersonal communication skills.

References

1. Johnsen, K., Dickerson, R., Raij, A., et al.: Experiences in using immersive virtual characters to education medical communication skills. IEEE Virtual Reality (2005)
2. Hill, R., Gratch, J., Marsella, S., Rickel, J., Swartout, W., Traum, D.: Virtual Humans in the Mission Rehearsal Exercise System. Künstliche Intelligenz 17, 5–10 (2003)
3. Lombard, M., Ditton, T.B.: The Role of Screen Size in Viewer Responses to Television Fare. Communication Reports 10, 95–106 (1997)
4. Reeves, B., Lang, A., Kim, E.Y., Tatar, D.: The Effects of Screen Size and Message Content on Attention and Arousal. Media Psychology 1, 49 (1999)
5. Lombard, M.: Direct Responses to People on the Screen: Television and Personal Space. Communication Research 22, 288–324 (1995)
6. Mori, M.: The Uncanny Valley. Energy 7, 33–35 (1970)
7. Bailenson, J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J.: The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. Presence: Teleoperators & Virtual Environments 14, 379–393 (2005)
8. Pausch, R., Proffitt, D., Williams, G.: Quantifying immersion in virtual reality. ACM SIGGRAPH, 13–18 (1997)
9. Bowman, D., Datey, A., Ryu, Y., Farooq, U., Vasnaik, O.: Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments. Human Factors and Ergonomics Society, 2134–2138 (2002)
10. Swan, J.E., Gabbard, I.I., Hix, J.L., Schulman, D., Kim, R.: A comparative study of user performance in a map-based virtual environment. In: IEEE Virtual Reality, pp. 259–266 (2003)
11. Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R.: With Similar Visual Angles, Larger Displays Improve Spatial Performance. In: ACM SIGCHI (2003)
12. Slater, M., Pertaub, D., Steed, A.: Public Speaking in Virtual Reality: Facing an Audience of Avatars. IEEE Computer Graphics and Applications 19, 6–9 (1999)
13. Slater, M., Sadagic, A., Usoh, M., Schroeder, R.: Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. Presence: Teleoperators & Virtual Environments 9, 37–51 (2000)
14. Johnsen, K., Lok, B.: An Evaluation of Immersive Displays for Virtual Human Experiences. In: IEEE Virtual Reality (2008)
15. Ware, C., Arthur, K., Booth, K.S.: Fish tank virtual reality. In: ACM SIGCHI, pp. 37–42 (1993)
16. Schwinghammer, T.L.: Pharmacotherapy Casebook: A Patient-Focused Approach. McGraw-Hill Medical (2005)
17. Mehrabian, E.W.: Silent Messages. Wadsworth, Belmont (1971)