12-2015

Virtual humans and Photorealism: The effect of photorealism of interactive virtual humans in clinical virtual environment on affective responses

Himanshu Chaturvedi
Clemson University, hchatur@clemson.edu

VIRTUAL HUMANS AND PHOTOREALISM: THE EFFECT OF PHOTOREALISM OF INTERACTIVE VIRTUAL HUMANS IN CLINICAL VIRTUAL ENVIRONMENT ON AFFECTIVE RESPONSES

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Computer Science

---

by
Himanshu Chaturvedi
December 2015

Accepted by:
Dr. Sabarish Babu, Committee Chair
Dr. Larry Hodges
Dr. Sophie Joerg

ABSTRACT

The ability of realistic vs stylized representations of virtual characters to elicit emotions in users has been an open question for researchers and artists alike. We designed and performed a between subjects experiment using a medical virtual reality simulation to study the differences in the emotions aroused in participants while interacting with realistic and stylized virtual characters. The experiment included three conditions each of which presented a different representation of the virtual character namely; photo-realistic, non-photorealistic cartoon-shaded and non-photorealistic charcoal-sketch. The simulation used for the experiment, called the Rapid Response Training System was developed to train nurses to identify symptoms of rapid deterioration in patients. The emotional impact of interacting with the simulation on the participants was measured via both subjective and objective metrics. Quantitative objective measures were gathered using skin Electrodermal Activity (EDA) sensors, and quantitative subjective measures included Differential Emotion Survey (DES IV), Positive and Negative Affect Schedule (PANAS), and the co-presence or social presence questionnaire. The emotional state of the participants was analyzed across four distinct time steps during which the medical condition of the virtual patient deteriorated, and was contrasted to a baseline affective state. The data from the EDA sensors indicated that the mean level of arousal was highest in the charcoal-sketch condition, lowest in the realistic condition, with responses in the cartoon-shaded condition was in the middle. Mean arousal responses also seemed to be consistent in both the cartoon-shaded and charcoal-sketch conditions across all time steps, while the mean arousal response of participants in

the realistic condition showed a significant drop from time step 1 through time step 2, corresponding to the deterioration of the virtual patient. Mean scores of participants in the DES survey seems to suggest that participants in the realistic condition elicited a higher emotional response than participants in both non-realistic conditions. Within the non-realistic conditions, participants in the cartoon-shaded condition seemed to elicit a higher emotional response than those in the charcoal-sketch condition.

DEDICATION

I would like to dedicate this to my parents who have always believed in me.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. Sabarish Babu, who has immensely supported me throughout this project. My amazing experience of working in the Virtual Environments group and doing great research would not have materialized without his guidance. I would like to thank Dr. Hodges and Dr. Joerg for their feedback on my research.

This project would have been incomplete without the help of Nate Newsome who developed the Sketch shader which we used in our experiments. Along with him, I would like to thank all the members of the Virtual Environments group for helping me in so many ways. I would also like to thank Dr. Knijnenburg who has provided keen insight on how this research can be extended by using advanced statistical methods.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

We have recently witnessed a rise in the development of human-like virtual agents, and a strong push to enhance their appearance and believability. Virtual humans are extensively being deployed to simulate social interactions with humans. Medical educators, in particular, have become increasingly more interested in using virtual humans for training systems that deal with interpersonal interactions that may be dangerous or difficult to reproduce in reality. Virtual humans have been implemented for many scenarios including: routine patient surveillance (Cairco et al., 2012), negotiation simulation (Dehghani, Carnevale, & Gratch, 2014), interview emulation (DeVault et al., 2014), interpersonal and leadership skills training (Campbell et al., 2011), and uncomfortable situations such as prostate exams (Robb et al., 2013).

The visual fidelity of virtual humans is important to consider in both research and entertainment applications. Significant amount of time can be invested in 3D modeling, animation, and system design to create realistic human-like digital characters. Moreover, advanced computational resources dedicated to real-time rendering, shading and animation are be needed to render high fidelity virtual characters. Stylized, or "cartoon", rendering of virtual humans has been widely used, as it takes less time and resources to create. However, it has been unclear to what extent these rendering styles influence empathetic response in simulated interpersonal experiences.

It has been shown that human-agent interactions can evoke similar social responses to that of human-human interactions (DeVault et al., 2014; Robb et al., 2013).

Generating an emotional response from participants is important when working with learning systems, as it has a strong influence on memory retention (Dunsworth & Atkinson, 2007). If a system is able to elicit emotions from a participant, they may be more engaged and more likely to associate emotions with task interpretation. Research has shown that when the virtual character's responses were empathetic towards the participant's emotions, there was a significant decrease in stress and a positive effect on the participants' perception of the task (Prendinger, Mori, & Ishizuka, 2005). These emotional responses can also be used to measure co-presence in virtual environments to see if participants are forming a bond with the character. When working with medical training systems, it is especially important to accurately mimic human behaviors and characteristics in order to elicit empathetic responses from participants. Research shows a significant impact of the elicited emotional responses on participants' learning outcomes on a task with virtual humans (Robb et al., 2013). Furthermore, evidence suggests that the emotional state of medical practitioners can have an impact on their professional abilities (Roter, Frankel, Hall, & Sluyter, 2006).

Avoiding the Uncanny Valley phenomenon has been a challenge in the computer animation and robotics fields for some time (Mori, 1970; Qu, Brinkman, Ling, Wiggers, & Heynderickx, 2014). The theory suggests that empathetic response increases with the human-likeness of the virtual human to a certain point and then drops to a point of revulsion before increasing again. The entertainment industry has provided a great example for this phenomenon. Levy (Levy, 2004) discusses how audiences found the highly realistic human characters in *Polar Express* to be creepy and off-putting, while most found the stylized cartoon characters in Pixar's *The Incredibles* more appealing. While there have been significant strides in understanding the phenomenon, it is our intention to examine the impact of visual fidelity of virtual human realism on this phenomenon when measured using quantitative and qualitative methods examining the emotional responses of users.

For a previous study, we developed a medical virtual reality training system to help nurses identify the signs of rapid patient deterioration (Wu et al., 2014). While developing the system, we worked closely with medical experts to create accurate virtual humans based on actual patient data. In this work, we used this system to study how different rendering techniques (realistic and non-realistic) emotionally impact the participants. Our empirical evaluation adds to the current understanding of the effects of visual fidelity of virtual humans on human emotional responses.

CHAPTER TWO

RELATED WORK

Presence of emotional effects in human-virtual agent interactions has already been exhibited in previous studies (Pan, Banakou, & Slater, 2011; Robb et al., 2013). De Melo et al. (De Melo & Gratch, 2009a), observed that people attempted methods of emotional negotiation with virtual humans. It has also been established that users do not disregard their social standards and stereotypes when interacting with virtual agents (Obaid et al., 2012). Researchers have studied whether synthetic emotions expressed by virtual humans elicit emotion in a human conversation partner (Qu et al., 2014). They found that negative compared to positive synthetic emotions expressed by a virtual human can elicit a more negative emotional state in a human conversation partner. Effects of virtual agents on users while performing tasks were examined (Zanbaka, Ulinski, Goolkasian, & Hodges, 2007) and it was found that users were socially inhibited while performing complex tasks, and social facilitation did not occur while doing simple tasks in the presence of virtual humans.

Attempts have been made to understand the characteristics of virtual agents which have the greatest effect on users' emotions. Bouchard et al. (Bouchard et al., 2013) studied the empathetic responses of users interacting with virtual avatars of known versus unknown individuals expressing pain. They found that even though the users were empathetic towards both the virtual humans, they were more empathetic towards the avatar of the known individual. Tsai et al. (Tsai, Bowring, Marsella, Wood, & Tambe, 2012) showed that a still image of a virtual agent could invoke empathy in the user when comparing a happy expression with that of a neutral one. Garau et al. (Garau, Slater, Pertaub, & Razzaque, 2005) studied the effect of responsiveness of virtual humans on users. They found that users experienced a higher sense of personal contact with visually responsive virtual humans as opposed to unresponsive ones. The results of the experiment conducted by Ehrlich et al. (Ehrlich, Schiano, & Sheridan, 2000) suggested that while communicating facial affect, if there is a trade-off between image realism and motion realism, preference should be given to motion realism. Bailenson et al.(Bailenson, Yee, Merget, & Schroeder, 2006) studied the effect of behavioral realism and form realism in collaborative virtual environments. They found that verbal and non-verbal self-disclosure was lowest in the highest behavioral and form realism condition.

Katsyri et al.(Kätsyri & Sams, 2008) studied the effects of dynamic information in identifying facial emotions. The results showed that although dynamics does not improve identification to emotions which are distinctive statically, it plays an important role in identification of subtle emotions. The effects of presence of facial expressions have been studied at a granular level as well. Tinwell et al.(Tinwell, Grimshaw, Nabi, & Williams, 2011) found that the lack of facial expressions in the upper part of the face intensify the uncanny appearance of the virtual character by restricting effective communication of emotion. De Melo et al.(De Melo & Gratch, 2009b) studied the effect of presence of wrinkles, blushing, sweating and tears on perception of emotions and found a significant positive effect. Courgeon et al.(Courgeon, Buisine, & Martin, 2009) concentrated their efforts on studying the impact of different renderings of expressive wrinkles on users' perception of facial expression of emotions. They found that realistic representation of wrinkles increased the expressiveness of the virtual human but it did not help with the recognition of emotions by the users.

In a previous study, we examined the effect of presence of animation on emotion contagion in a medical virtual reality simulation with interactive virtual patients (Wu et al., 2014). Our research showed that the presence of the animation had a significant effect on users' emotional responses, especially negative emotions corresponding to the simulated deterioration of the virtual patient, and it also enhanced the levels of social presence in user's interactions with the virtual patients. Guadagno et al.(Guadagno, Blascovich, Bailenson, & Mccall, 2007) studied the factors which affect the persuasive abilities of virtual humans. They found that virtual agents of same gender as the participant induced greater attitude change. Their results also support the theory that virtual humans with high behavioral realism are more influential.

Garau et al.(Garau et al., 2003) also studied the effect of behavioral realism using eye gaze control as the variable and visual realism photorealism of the avatar as another variable. Their results show that inferred gaze behavior of the avatar out-performed random gaze behavior. They also found that random gaze behavior adversely affected the non-photorealistic condition and hence suggested that aligning both visual and behavioral realism is important. In our experiment, we kept the eye gaze behavior consistent throughout the conditions in order to avoid any discrepancies.

More recently, Wellerdiek et al. (Wellerdiek et al., 2015) found that the body shape and pose of virtual characters can potentially affect the perception of physical strength and social power. They found that characters with a weaker looking body shape can be perceived as more powerful when presented in a high-power pose, as compared to a stronger looking body shape. In our experiment design, when modeling the visually realistic appearance of the virtual human, we strived to match the body shape and actions of the realistic appearing virtual human to exactly that of the stylized appearance conditions so that these unforeseen effects, such as perceived strength and power do not confound the measured emotional reactions of the users.

Research has also focused on enhancing the expressiveness of virtual characters so that the users respond to interpersonal simulations in the intended manner. Various models of virtual character's emotion, models of interpersonal behavior and methods of generating expression for creating virtual characters with higher fidelity have been discussed in (Vinayagamoorthy et al., 2006). Silverman (Silverman, 2004) discussed the need to make better use of human performance moderator functions (HPMFs) and presented a framework to integrate them in the simulation. HPMFs are relations derived from empirical data which can be used to predict human behavior due to internal and external influences such as fatigue, workload or temperature.

One of the basis of our research is the paper published by Gaggioli et al.(Gaggioli, Mantovani, Castelnuovo, Wiederhold, & Riva, 2003). They thoroughly discussed three major factors which should be considered while developing avatars for clinical psychology applications namely, visual realism of the avatars and the environment, the behavior of the avatars, and interaction with the avatars. They suggested that the level of photorealism of virtual humans should vary depending on the task performed by the user.

They also described the concept of *continuum of visual realism* which refers to a scale from high visual fidelity to low visual fidelity.

Effects of using realistic virtual environments have been studied before. Slater et al.(Slater, Khanna, Mortensen, & Yu, 2009) compared the effect of using recursive real-time raytracing vs ray casting. Recursive real-time ray tracing produces a more realistic illumination with the help of real-time shadows and reflections. They found that with real-time raytracing user's experienced higher anxiety as well as higher presence. Yu et al.(Yu, Mortensen, Khanna, Spanlang, & Slater, 2012) extended this work to study if the enhanced responses were due to dynamic lighting or due to higher overall visual quality. To rule out the factor of dynamic lighting they designed two conditions, both of which used dynamic lighting but one condition had enhanced overall illumination quality by using global illumination. Their results suggest that the inclusion of global illumination played an important role in creating a more convincing virtual environment. In our study, we have used consistent dynamic lighting across all conditions.

Quite a few studies have been done to understand the relation between various rendering techniques of virtual humans and the emotional response of the users. Mandryk et al. (Mandryk, Mould, & Li, 2011) studied subjective emotional responses to multiple non photorealistic rendering (NPR) approaches of virtual humans in digital videos, and found that NPR algorithms dampened users' emotional responses in terms of arousal and valence. McDonnell et al. (McDonnell, Breidt, & Bülthoff, 2012) carried out experiments to determine the effect of render styles on perception of personality of virtual humans. They found that the characters that appear highly abstract or highly realistic are considered equally appealing. Extremely unappealing characters appear even less appealing when moving than when still. Ellis et al. (Ellis & Bryson, 2005) demonstrated that the extent to which subjects ascribe emotions to VR faces is highly dependent on textures (photo-realistic vs. non-photorealistic) applied to the face. They found no significant difference for the happy or surprised conditions and all the variance resulted from the sad and angry conditions, both of which favor the photo-realistic face.

Kwon et al.(Kwon, Powell, & Chalmers, 2013) recently performed an experiment studying the effect of level of graphical realism on job interview anxiety in a VR exposure environment. They found that even though graphically realistic virtual human delivered a higher sense of presence, the sense of anxiety was not correlated to realism.

Ruhland et. al.(Ruhland et al., 2014) investigated the effects of eye gaze and visual realism of virtual humans on the perception of personality. They found that if the eye gaze and blinking behavior of a virtual character were accurately modeled then users perceive the characters personality to be robust and consistent, despite the differences in visual realism of the character models. In our experiment, we strived to maintain consistency in the animation quality of facial expressions and gestures, such that our experiment setup can allow us to examine the effects of visual appearance fidelity alone on the emotional reactions of users in interactions with a virtual human

Of relevance is the research of Zibrek and McDonnell (Zibrek & McDonnell, 2014) in which the authors found that the rendering style of the virtual character have the potential to affect the perceived personality of the character. They found that a visually realistic but ill-looking rendering style evoked a less desirable (less agreeable) personality than the cartoon style presentation of a virtual human. These results have implications to how users may perceive ill-looking yet visually realistic virtual humans in clinical deterioration scenarios in interactions with medical virtual reality systems like our RRTS.

Although emotion contagion and virtual human perception studies have been done in the past, most of them have used virtual characters in the form of still images or in short video clips. To our knowledge, this is the first research to study the effect of realistic vs stylized appearance on users' emotion contagion to interactive virtual humans in an IVE.

CHAPTER THREE

EXPERIMENT DESIGN

Setup and Simulation

The Rapid Response Training system (RRTS) (Cairco et al., 2012) was designed to train nurses in recognizing the signs and symptoms of rapid deterioration in patients. The RRTS simulates the daily duties of a nurse including visiting multiple patients typically four times a day, gathering the vital signs of the patients and reporting them in an Electronic Health Record System (EHR). To provide this experience our system is setup in a dual screen configuration. A large screen display is used to show the virtual hospital environment which includes the patient and the various instruments present in a general ward in a life-size view. The second monitor displays the Electronic Health Record (EHR) which is used to report the quantitative and qualitative vital signs gathered while interacting with the virtual patient.

Virtual humans used in the RRTS have been modeled after real life patients who have undergone rapid deterioration, over the course of a nurse's shift. The signs and symptoms of deterioration have been carefully modeled and animated in our virtual patients with the help of medical experts. Several modes of interaction with the patient are provided. The users can ask a set of predefined medically relevant questions (via a dialogue box). They can also use one of the instruments in the patient's environment to measure his vital signs, and record the observations in the simulated EHR system. Detailed information regarding the RRTS design, virtual human, and instrument interaction and implementation can be found in (Cairco et al., 2012; Wu et al., 2014). In this research, the RRTS served as a rich experiment platform for empirical examination of how factors associated with the visual appearance of the virtual human can affect the users' emotional state and responses (see Figure 3.1).

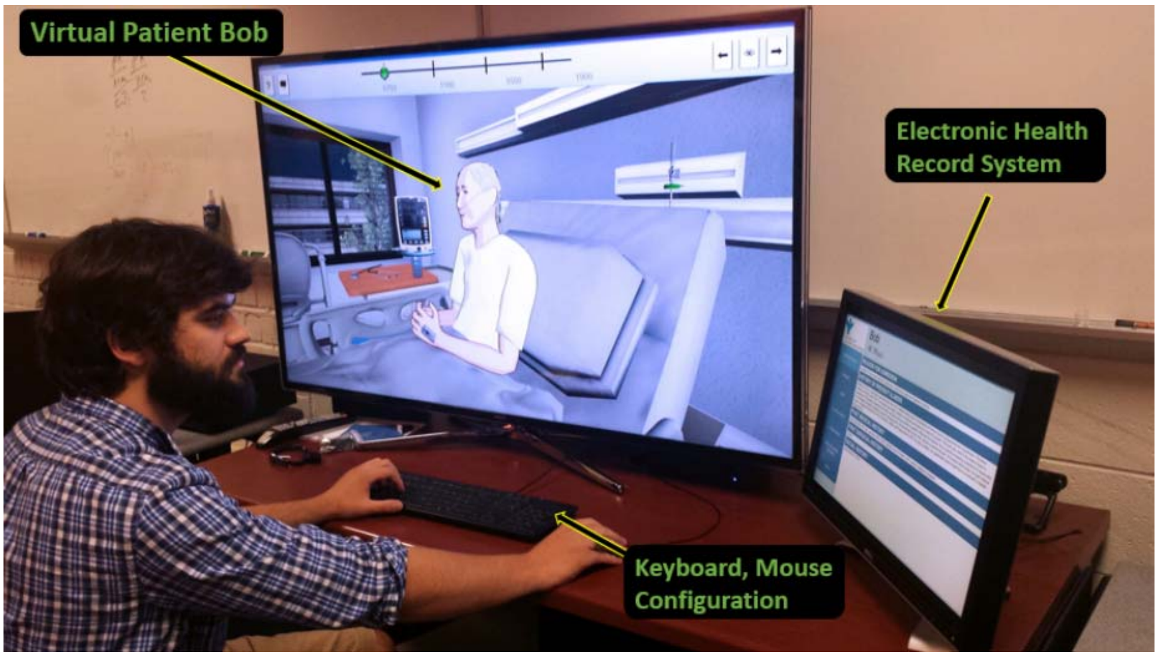


Figure 3.1: Screenshot shows a participant interacting with the virtual patient Bob in the RRTS, and recording his vitals in the EHR screen

The experiment featured three conditions, one photo-realistic and two non-photorealistic (Cartoon and Sketch). When considering the rendering style of our non-photorealistic (NPR) conditions, we used two very distinct techniques. After assessing many possibilities, we modeled one NPR condition to be cartoon-like (cel shaded), and the other to appear like a hand drawn as a sketch. We aimed to provide a consistent user experience among experiment conditions with the RRTS, and avoid any possible confounds due to differences in the virtual environment or instrument appearance. Therefore, the only modeled difference in the experiment conditions was the rendering techniques applied to the virtual patient; all virtual human animations and behaviors, as well as the look and feel of the virtual environment and instrument interactions in RRTS remained constant (see Figure 3.2).

In the Realistic condition, the skin textures of the virtual patient were modeled after real patients including visual details such as wrinkles and blemishes to add to the realism. Custom diffuse shaders were used to adjust the appearance of the skin texture as the patient deteriorates. The Cartoon, or cel shaded, condition used Unity3D's Basic Outline Toon shader that gives the character a uniform outline and simplified two color shading. Only the skin and eyeball textures were kept in detail in order to provide an outline of the eyes, eyebrows, and pupils. Specular highlights provided visual clues about the material type (skin, clothing etc.) of the model. Unlike standard shaders, there are clear borders with respect to shadows on cel shaded objects. The Sketch condition was implemented using a custom shader to give the character a hand-drawn charcoal sketch like appearance. A drawing technique called "hatching" was used, which refers to closely spaced parallel strokes that follow the curvature of a surface to define volume and materiality. No color is used in this condition, instead the density of the hatch marks defines tone, where less dense strokes denote a lighter tone and denser strokes a darker tone. Cross-hatching is also applied to create darker tones by layering strokes at different angles. We created a texture based shader that used a pre-rendered sequence of mipmapped hatch stroke images which correspond to different tones. A multi-texturing algorithm was implemented to calculate the lighting tones for each face, and then blend the appropriate hatching images to render the proper shading to the model of the virtual patient. Each texture in the sequence contains the hatch marks from all the previous images. This nesting property helped to provide seamless blending between tones.

The RRTS was deployed in a dual screen system that used a 60" TV as the main screen for the simulation to display the interactive virtual patient (Bob) and a 21" LCD Monitor for the EHR system (as shown in Figure 3.1). The participant interacted with the system using a standard mouse and keyboard. They wore a sensor (Qsensor from Affectiva Inc.) that measured the skin electro-dermal activity on the wrist of their non-dominant hand. A separate laptop was provided in between time-steps for gathering responses to the surveys.

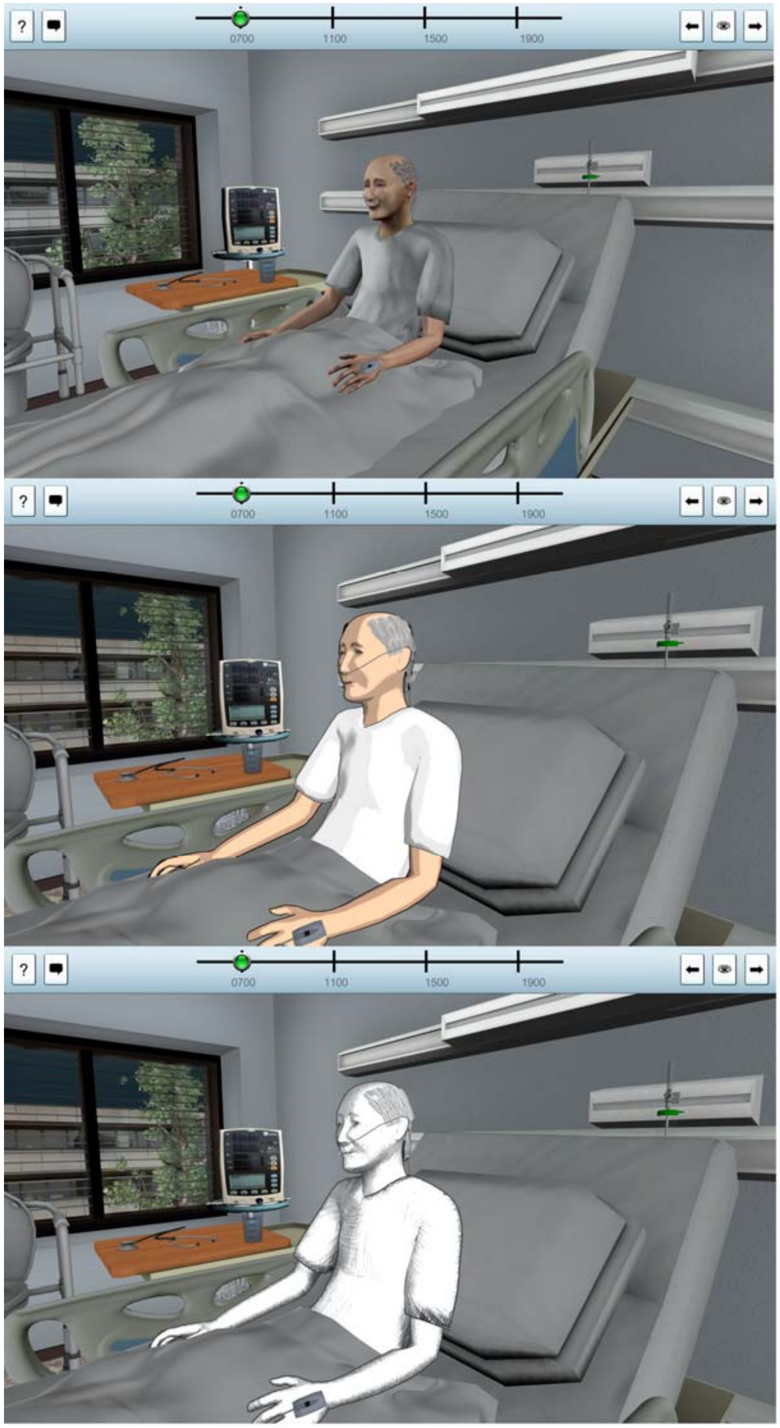


Figure 3.2: Screenshots depicting the realistic (top) versus stylized cartoon like (center) and sketch like (bottom) representation of a virtual humans.

Hypothesis and Research Questions

A good deal of time and forethought went into creating realistic human-like appearance for our virtual characters in the RRTS. We found ourselves wondering what effect all this extra effort had on the user. We narrowed down three essential questions that are currently unanswered in the research literature:

• Can virtual human visual realism affect the emotions of users?

• To what extent does a realistic (PR) human-like appearance of a virtual human versus two levels of a non-photo-realistic (NPR) appearance of a virtual human, namely a cartoon like rendering and a charcoal sketch like rendering, affect the emotional responses of users?

• To what extent does amplifying the emotional behaviors of the virtual human impact the users' affective reactions in each of the rendering conditions?

Study Design

In order to empirically examine each of these questions, we developed a 3x4 experiment. We had three between-subjects conditions: Realistic rendering, Cartoon rendering, and Sketch rendering. There were four within subject conditions. These were the distinct time-steps that had increasing levels of negative stimuli. Each participant was exposed to the same virtual agent, Bob, in each of the four time-steps. The declining health condition of Bob was expressed through his vital signs, verbal feedback, and non-verbal behaviors. The animated behaviors of the agent changes between time stamps to reflect this decline in the medical condition of Bob.

We recruited both male and female participants, between the ages of 18 and 35, who had a basic knowledge of the medical terminology present in the simulation. We ran a total of 35 participants (11 in Realistic, 12 in Cartoon, and 12 in Sketch rendering conditions), who were recruited from Clemson University. We had a near equal distribution of gender in our participant pool that included 20 males and 15 females.

Methodology

Participants first listened to a brief explanation regarding the design and objectives of the RRTS system. After consent was obtained, we placed an EDA Q Sensor on the wrist of the participant's non-dominant hand. The participant was then asked to fill out a series of questionnaires regarding his or her background and current disposition. The participant was trained on the system using a virtual human other than the one that would be used throughout the experiment, but in the same visual rendering condition (Realistic or Cartoon or Sketch) exhibiting a stable medical condition (time-step one). The experimenter introduced the participant to every interaction in the virtual environment during the training phase before initiating the study. Then the participant was introduced to the first time-step and asked to interact with Bob by asking as many questions as possible, use as many virtual instruments as necessary to medically assess Bob's condition, and record the observations in the EHR system. At the end of each time-step, the participant filled out a Differential Emotions Survey (DES) and a Positive and Negative Affect Schedule (PANAS) questionnaire to reflect his or her current emotional state. The participant was also given a short quiz on the condition of the patient after every time-step as an incentive to increase interactions with Bob. Finally, at the end of the fourth time-step, after they were administered the PANAS and DES surveys, they completed the co-presence survey, and were debriefed and thanked for their time.

Measures

The independent variables are visual appearance conditions of the virtual human, and the intensity of the negative emotions of the virtual agent from one time-step to the next of the simulated surveillance scenario of the virtual patient. There were a number of dependent variables used to obtain the emotional reactions of the user.

Quantitative Measures

Any physiological arousal, both external and internal, affects the sympathetic nervous system, which controls the sweat producing Eccrine glands and makes the skin a better conductor of electricity (Daily, Meyers, Darnell, Roy, & James, 2013). This measurable quantity of electrical conductance is known as Electrodermal activity (EDA), and can be measured as skin conductance using wrist worn sensors. In this case, the wireless Q sensor, manufactured by Affectiva, was placed on the ventral side of the participant's wrist of the non-dominant hand. To mark the different time steps in this study (Figure 3.3), the timestamp button on the sensor was used to indicate the start and the stop time. Data collected from the participant, was transferred from the sensor to a computer using a USB cable. To account for between subject variance and baseline fluctuations the EDA data was normalized, using the following formula, as described by Healy et al (Healey & Picard, 1998).

$$(SCR - MEAN(SCR)) / (MAX(SCR) - MIN(SCR))$$

SCR refers to Skin Conductance Response.

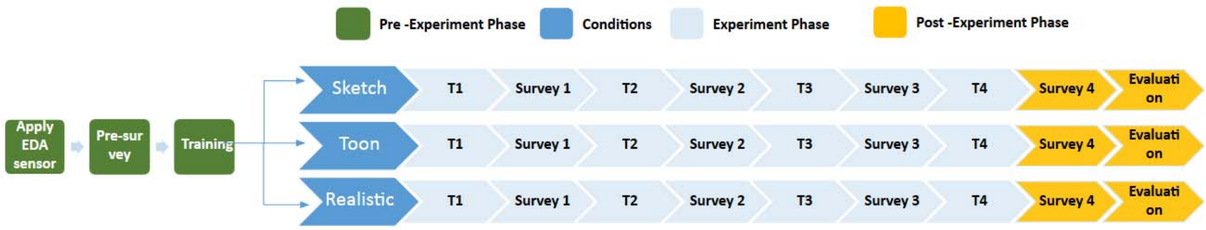


Figure 3.3: The time-line of the experiment from left to right.

For every participant, the mean and standard deviation values were calculated from the normalized EDA data set. The window for each mean and standard deviation were the timestamps mentioned in the *Methodology* section (demographic questionnaire, training phase, and timestamp1 to timestamp 4). The same process was followed for all the three experimental conditions (Toon, Sketch and Realistic).

The Positive and Negative Affect Schedule was used to evaluate an overall valence of emotion (Watson, Clark, & Tellegen, 1988). This is a 20 item survey, with each item being scored on a Likert scale from 1 to 5. The score reflects how strongly the participant felt an emotion at a given time, 1 being not at all and 5 being extremely. The results were grouped by positive and negative items.

The Differential Emotions Scale (DES) questionnaire in our experiment was based on the DES IV with a modification to reduce the item count to 30 (Van Der Schalk et al., 2011). There are 10 categories scored in the DES (3 items per category). The participant used a 0-9 Nominal style scale to express how strongly they felt each item. Again, 0 being Never and 9 being Extreme. The results were grouped into categories, and each category received its own score. The ten categories used in our questionnaire are as follows: Interest, Enjoyment, Surprise, Sadness, Anger, Contempt, Fear, Guilt, Shame and Shyness.

In addition to the repeated measures above, there was a final evaluation of the immersive virtual patient Bob. This survey contained 5 items based off the social presence survey (Bailenson, Blascovich, Beall, & Loomis, 2003). These were scored on a 1-7 Nominal style scale, 1 being not at all and 7 being a great deal.

Qualitative Measures

In an attempt to better interpret the quantitative items, we used a number of open-ended discussion questions. These questions were used to assess their overall experience in the simulation. Two examples of these questions are: "what did you like most about Bob?" and "how would you describe Bob's personality and appearance?"

CHAPTER FOUR

RESULTS

Objective measure: Electrodermal Activity

In order to measure if the arousal levels of the participants' differed by the rendering style of the virtual human and to examine the differences in arousal based on the deterioration scenario presented, we conducted a 3 x 6 mixed model repeated measures Analysis of Variance (ANOVA) on the mean EDA scores with sampling time across the six different period of the experiment, namely the pre-experiment survey, training, and time-steps 1 to 4, as a within-subjects repeated measures variable. The conditions Realistic, Sketch and Cartoon shading were the between-subjects variable. In all the ANOVA analyses conducted in our empirical evaluation, we first tested to make sure that Mauchley's test of sphericity has been met. In cases where the test of sphericity was significant, Greenhouse-Geisser corrected degrees of freedom was used to assess the significance of the corresponding F values. Tukey HSD post-hoc was used to evaluate the effects of the between subjects variable of rendering style within each time-step, and Bonferroni adjusted type 1 error (alpha) was used for within subjects post-hoc comparisons across time within scores in a rendering style.

The 3 x 6 ANOVA did not reveal any significant difference in the main effect of condition, or the time of sampling, or the interaction between time and condition. In order to examine the effects of gender on arousal levels of participants overall, we conducted a 2 x 6 mixed model repeated measures ANOVA on the mean EDA scores, with sampling time across the six different period of the experiment as a within-subjects repeated measures variable. Gender was compared as a two level between-subjects variable. The ANOVA analysis did not reveal a significant difference in the main effect of sampling time or interaction of gender and time, but revealed a significant main effect of gender. Overall, male participants (M=0.322, SD=0.032) exhibited significantly higher levels of arousal overall than female participants (M=0.222, SD=0.039), $F(1, 32) = 4.135$ with a $p = 0.045$.

Subjective Measure: Positive and Negative Affect Schedule

In order to assess if there were any significant effects of condition or gender on positive and negative affect scores across the five sampling times, namely pre-experiment, and after time steps 1 through 4 corresponding to Bob's deterioration, we performed a 3 x 2 x 5 mixed model repeated measures ANOVA. The between subjects variables were condition (3 levels) and gender (2 levels), and the within subjects variable were the sampling times. The ANOVA analysis did not reveal any significant main effects or interaction effects for positive affect scores. With respect to negative affect, the ANOVA analysis did not reveal a significant main effect of gender or a significant main effect of condition. However, the ANOVA analysis revealed a significant main effect of negative affect scores overall, $F(4, 128) = 36.83$ and $p < 0.001$, and a significant interaction effect of negative affect scores by condition, $F(8, 128) = 5.123$ and $p = 0.004$ (See Figure 4.1). Pairwise comparisons between the repeated levels of negative affect, with p values adjusted based on Bonferroni correction, revealed that overall each of the levels of negative affect were significantly different than the other. Negative affect was the least at the baseline, prior to the experiment, and was significantly highest after time step 4 after witnessing that Bob had significantly deteriorated. All the mean negative affect scores were significantly different than the others and higher than the previous time-step. Therefore, the interactive medical deterioration scenario with Bob produced a gradual increase in the mean affect scores of participants from one time-step to the next overall, regardless of the appearance conditions of Bob. These effects will be explored in further detail in the analysis of the DES results.

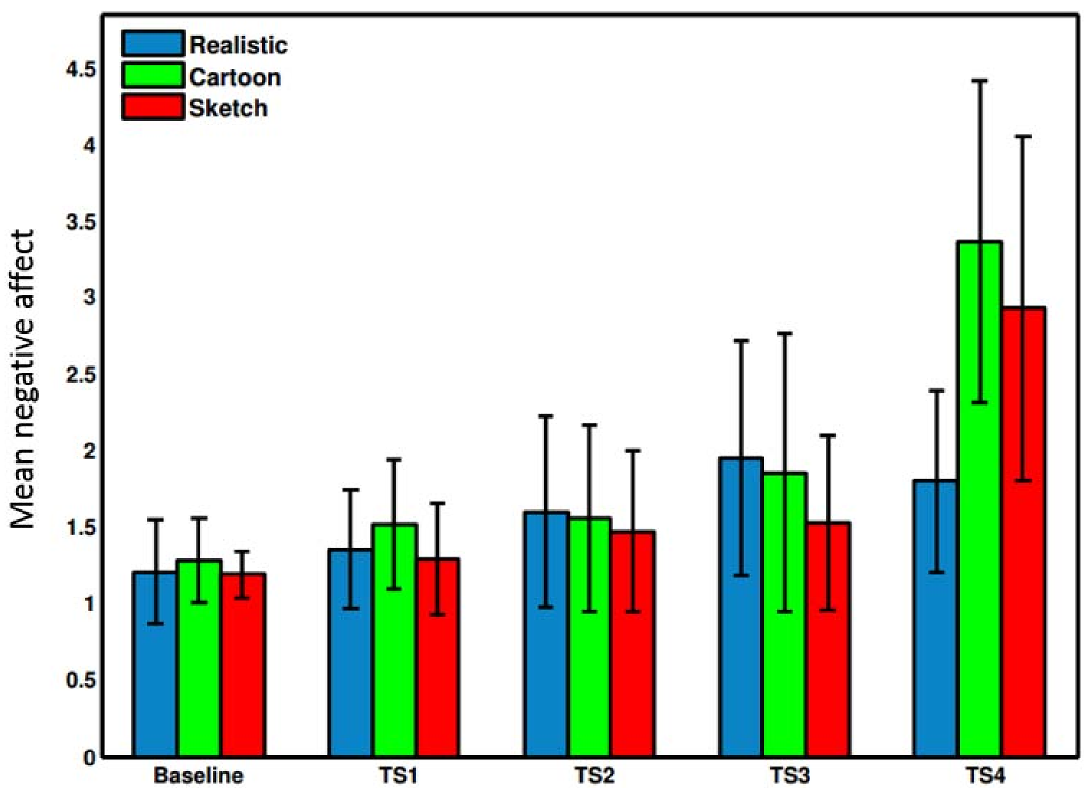


Figure 4.1: Bar graph showing the mean and SEM of negative affect scores across the different rendering conditions of the virtual human across time-steps.

In order to examine the interaction effect between condition and negative affect scores, post-hoc analysis by blocks of scores across the different sampling times of negative affect was examined in each of the three virtual human rendering conditions (Realistic, Cartoon and Sketch), by first performing a one-way repeated measures ANOVA within the block of trials. In the Realistic condition, the one-way repeated measures ANOVA across mean negative affect scores at the different sampling times was significant, $F(4, 40) = 7.58$ and $p = 0.004$. Post-hoc pairwise comparisons with Bonferroni correction revealed that the negative affect scores of participants in the Realistic condition at the baseline (M=1.209, SD=0.103) were significantly lower than the scores after time-step 4 (M=1.8, SD=0.180), $p = 0.010$. Post-hoc pairwise comparisons also revealed that the negative affect scores of participants in the Realistic condition after time-step 1 (M=1.35, SD=0.18) were also significantly lower than time-step 3 (M=1.96, SD=0.23), $p = 0.43$.

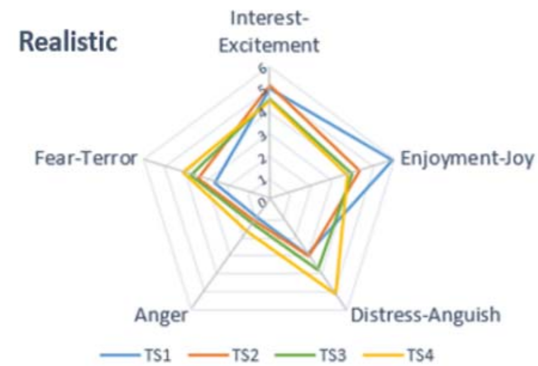
In the Cartoon condition, the one-way repeated measures ANOVA across mean negative affect scores at the different sampling times was also significant, $F(4, 44) = 18.75$ and $p < 0.001$. Post-hoc pairwise comparisons with Bonferroni correction revealed that the negative affect scores of participants in the Cartoon condition at the baseline (M=1.28, SD=0.08) was significantly less than after time-step 4 (M=3.37, SD=0.30), $p < 0.001$. Negative affect scores of participants in this condition after time-step 1 (M=1.52, SD=0.12) was also significantly lower than time-step 4 (M=3.37, SD=0.30), $p = 0.002$. Finally, negative affect scores of participants in the Cartoon condition after time-step 2 (M=1.59, SD=0.17) were also significantly lower than after time-step 4 (M=3.37, SD=0.30), $p = 0.011$.

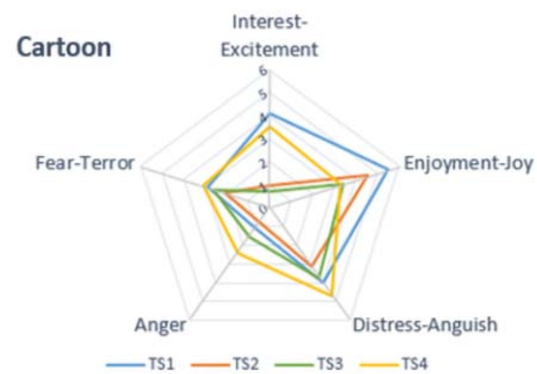
In the Sketch condition, the one-way repeated measures ANOVA across mean negative affect scores at the different sampling times was significant, $F(4, 44)=16.96$ and $p = 0.001$. Post-hoc pairwise comparisons with Bonferroni correction revealed that the negative affect scores of participants in the Sketch condition at the baseline (M=1.19, SD=0.04) was significantly lower than after time-step 4 (M=2.93, SD=0.33), $p = 0.002$. Negative affect scores of participants in the Sketch condition after time-step 1 (M=1.29, SD=0.11) were also significantly lower than after time-step 4 (M=2.93, SD=0.33), $p = 0.010$. Likewise, negative affect scores of participants in this condition after time-step 2 (M=1.29, SD=0.11) were also significantly lower than after time-step 4 (M=2.93, SD=0.33), $p = 0.029$. Finally, negative affect scores of participants in the Sketch condition after time-step 3 (M=1.53, SD=0.165) were significantly lower than after time-step 4 (M=2.93, SD=0.33), $p = 0.036$.

<u>Differential Emotions Survey</u>

In order to examine the dimensions of emotion response of participants to the deterioration scenario of Bob at a more granular level, we administered the Differential Emotions Survey (DES) at the end of every time-step of interaction with Bob. We also wanted to examine if the visual rendering conditions of the virtual human had any effect on the affective responses of participants, corresponding with the medical deterioration scenario, in each of the 10 dimensions of the DES survey. In each of ten granular dimensions of the DES questionnaire, we employed the following analyses. We first performed a 3 x 2 x 4 mixed model repeated measures ANOVA, with gender (2 levels) and rendering condition (3 levels) as the between-subjects variables, and the scores on each of the DES dimensions after each time-step of interaction with Bob, as a within-subjects repeated measures variable. This tested for the main effect of condition, gender, DES dimension, gender by DES interaction, and the condition by DES interaction terms. Appropriate post-hoc and block analyses were conducted for follow-ups on main and interaction effects respectively (same as previous section). We found significant differences in 5 out of 10 DES dimensions, 3 out of which correspond to negative emotions namely Anger, Distress-Anguish, and Fear-Terror. We believe that we observed significant effects only in these 3 negative emotions due to the nature of the application. The application portrayed a deteriorating patient who does not depict negative emotions such as shyness, guilt or shame. Hence, this could be the reason why no significant effects were found for transference of these negative emotions. Figure 4.2 shows the mean scores of the significant DES measures across the four time-steps in each rendering condition.

In the dimension interest-excitement, the mixed model ANOVA analysis revealed a significant main effect of interest-excitement scores across time-steps, $F(3, 87) = 22.5$ and $p < 0.001$, and a significant interest-excitement sampling times by condition interaction, $F(6, 87) = 6.92$ and $p < 0.001$. The main effect of gender and the gender by interest-excitement interaction was not significant. Post-hoc comparisons using Tukey HSD revealed that overall participants in the Realistic rendering condition scored significantly higher (M=4.80, SD=0.44), than participants in the Cartoon rendering condition (M=2.31, SD=0.43) in the dimension interest-excitement, $p < 0.001$. Post-hoc comparisons using Tukey HSD also revealed that overall participants in the Realistic rendering condition scored significantly higher (M=4.80, SD=0.44), than participants in the Sketch rendering condition (M=2.06, SD=0.57), p = 0.001.





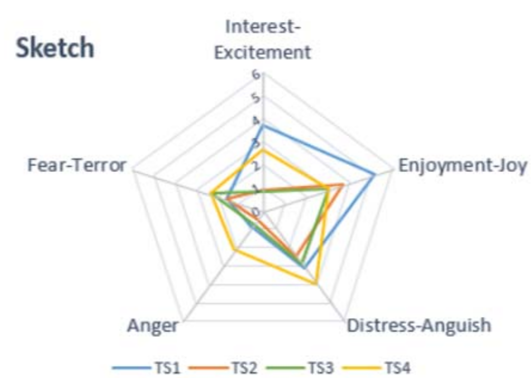


Figure 4.2: Radial graph shows the mean scores of the significant DES measure (interest-excitement, enjoyment-joy, distress-anguish, anger, fear-terror) across the four time-steps of interaction with Bob by each of the rendering conditions.

Pairwise comparisons examining the overall effects of sampling times of mean scores in the dimensions interest-excitement across the four time-steps of interaction with Bob were conducted using pairwise comparisons of means with Bonferroni adjusted p values. The pairwise comparison revealed that participants scored significantly higher after time-step 1 (M=4.25, SD=0.39), as compared to time-step 2 (M=2.356, SD=0.22) $p < 0.001$, and time-step 3 (M=1.99, SD=0.20) $p < 0.001$. Participants in time-step 2 (M=2.356, SD=0.22) scored significantly higher after time-step 2, as compared to time-step 3 (M=1.99, SD=0.20) $p = 0.002$, but significantly less than after time-step 4 (M=3.63, SD=0.47) $p = 0.002$. Participants scored significantly less after time-step 3 (M=1.99, SD=0.20) as compared to time-step 4 (M=3.63, SD=0.47) $p < 0.001$.

In order to thoroughly examine the interaction effect between interest-excitement means scores across each time step and condition, we conducted independent sample t-tests to compares the mean interest-excitement scores within each time-step (2-tailed hypothesis test), and performed block analysis to compare interest-excitement scores between time-steps in each condition. After time-step 2 participants in the Realistic condition (M=5.33, SD=1.50) scored significantly higher than participants in the Sketch condition (M=0.88, SD=0.74), $p < 0.001$. Also, after time-step 2 participants in the Realistic condition (M=5.33, SD=1.50) scored significantly higher than participants in the Cartoon condition (M=1.0, SD=0.94), $p < 0.001$. In time-step 3, participants in the

Realistic condition (M=4.57, SD=1.45) scored significantly higher than participants in the Sketch condition (M=0.83, SD=0.65), $p < 0.001$. Also in time-step 3, participants in the Realistic condition (M=4.57, SD=1.45) scored significantly higher than participants in the Cartoon condition (M=0.72, SD=0.80), $p < 0.001$. Finally in time-step4, participants in the Realistic condition (M=4.45, SD=1.74) scored significantly higher than participants in the Sketch condition (M=2.66, SD=2.31), $p = 0.48$.

The block analysis was conducted between the mean scores of interest-excitement after each time-step in each of the condition separately, as a one-way repeated measures ANOVA followed by post-hoc analysis. The one-way within subjects ANOVA on interest-excitement scores in the Sketch condition was significant, $F(3, 33) = 18.14$ and $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni adjustment revealed scores after timestep 1 (M=3.69, SD=0.53) were significantly higher than time-step 2 (M=0.88, SD=0.22) $p < 0.001$, and after time-step 3 (M=0.83, SD=0.19) $p < 0.001$. Post-hoc pairwise comparisons revealed that scores after time-step 2 (M=0.88, SD=0.22) were significantly lower than time-step 4 (M=2.66, SD=0.67) $p = 0.036$. Also, post-hoc pairwise comparisons revealed that scores after time-step 3 (M=0.83, SD=0.19) were significantly lower than after time-step 4 (M=2.66, SD=0.67) $p = 0.039$. The one-way within subjects ANOVA on interest-excitement scores in the Cartoon condition was significant, $F(3, 33) = 20.504$ and $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni adjustment revealed scores after time-step 1 (M=4.08, SD=0.61) were significantly higher than time-step 2 (M=1.0, SD=0.27) $p < 0.001$, and after time-step 3 (M=0.72, SD=0.23) $p < 0.001$. Post-hoc pairwise comparisons revealed that scores after time-step 2 (M=1.0, SD=0.27) were significantly lower than time-step 4 (M=3.5, SD=0.82) $p = 0.031$. Also, post-hoc pairwise comparisons revealed that scores after time-step 3 (M=0.72, SD=0.23) were significantly lower than after time-step 4 (M=3.50, SD=0.82) $p = 0.01$. The one-way within subjects ANOVA on interest-excitement scores in the Realistic condition was not significant.

In dimension enjoyment-joy, we found the main effect of sampling time of mean scores across the four time steps was significant, $F(3, 87) = 25.82$ and $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni adjustment revealed that overall scores after time-step 1 (M=5.52, SD=0.39) were significantly higher than time-step 2 (M=4.41, SD=0.42) $p = 0.005$, time-step 3 (M=3.65, SD=0.39) $p < 0.001$, and after time-step 4 (M=3.62, SD=0.44) $p < 0.001$. Post-hoc pairwise comparison revealed that scores after time-step 2 (M=4.41, SD=0.42) were significantly higher than time-step 3 (M=3.65, SD=0.39) $p = 0.007$, and after time-step 4 (M=3.62, SD=0.44) $p = 0.013$. Overall, enjoyment-joy seemed to decrease from the first time-step to the last, corresponding with the medical deterioration of Bob.

In dimension distress-anguish, we found a significant main effect of sampling time of mean scores across the four distinct time steps, $F(3, 87) = 10.25$ and $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni adjustment revealed that overall scores after time-step 1 (M=3.56, SD=0.41) were significantly lower than after time-step 4 (M=4.88, SD=0.45) $p < 0.001$. Post-hoc pairwise comparison also revealed that scores after time-step 2 (M=2.98, SD=0.35) were significantly lower than after time-step 4 (M=4.88, SD=0.45) $p < 0.001$. Finally, post-hoc pairwise comparison revealed that scores after time-step 3 (M=3.80, SD=0.41) were significantly lower than after time-step 4 (M=4.88, SD=0.45) $p = 0.043$. Overall, distress-anguish seemed to increase from one time-step to the next, with the participants exhibiting the highest distress-anguish after the last time-step, corresponding with Bob's medical deterioration.

In dimension anger, we found a significant main effect of sampling time of mean scores across the four distinct time steps, $F(3, 87) = 17.31$ and $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni adjustment revealed that overall scores after time-step 1 (M=0.95, SD=0.25) were significantly lower than after time-step 4 (M=2.56, SD=0.48) $p = 0.019$. Post-hoc pairwise comparison also revealed that scores after time-step 2 (M=0.83, SD=0.26) were significantly lower than after time-step 4 (M=2.56, SD=0.48) $p = 0.001$. Finally, post-hoc pairwise comparison revealed that scores after time-step 3 (M=1.29, SD=0.38) were significantly lower than after time-step 4 (M=2.56, SD=0.48) $p = 0.003$. Overall, anger seemed to increase from one time-step to the next, with the participants exhibiting the highest feeling of anger after the last time-step, corresponding with Bob's condition.

In dimension fear-terror, we found a significant main effect of sampling time of mean scores across the four distinct time steps, $F(3, 87) = 4.92$ and $p = 0.010$, and a significant interaction effect of gender by fear-terror scores, $F(3, 87) = 2.96$ and $p = 0.049$. Posthoc pairwise comparisons using Bonferroni adjustment revealed that overall scores after time-step 1 (M=2.32, SD=0.33) were significantly lower than after time-step 4 (M=3.26, SD=0.41) $p = 0.035$. Post-hoc pairwise comparison also revealed that scores after time-step 2 (M=2.26, SD=0.37) were significantly lower than time-step 3 (M=3.03, SD=0.43) p = 0.002, and after time-step 4 (M=3.26, SD=0.41) p = 0.002. Overall, fear-terror seemed to increase from one time-step to another, with the participants exhibiting the highest feeling of fear-terror in the last time-step, as Bob's condition deteriorated.

In order to thoroughly examine the interaction effect between fear-terror mean scores across each time step and gender, we conducted independent sample t-tests to compares the mean fear-terror scores within each time-step (2-tailed hypothesis test), and also performed block analysis to compare fear-terror scores between time-steps in each gender. After time-step 3, males (M=2.04, SD=0.53) showed significantly lower fear-terror scores than females (M=4.03, SD=0.67), p = 0.02. Also after time-step 4, males (M=2.47, SD= 0.51) showed lower fear-terror scores than females (M=4.05, SD=0.65), p = 0.023. The block analysis with the one-6 way repeated measure ANOVA on male scores did not reveal a significant difference. However, the block analysis of mean fear-terror scores across the distinct time-steps in females was significant, F(3, 42) = 4.78 and p = 0.014. Post-hoc pairwise comparisons using Bonferroni adjustment revealed that mean fear-terror scores in females after time-step 2 (M=2.84, SD=0.65) was significantly less than time-step 3 (M=4.02, SD=0.67) p = 0.02, and time-step 4 (M=4.18, SD=0.61) p = 0.026. It seems that females exhibit greater change in feat-terror emotions in response to Bob's medical deterioration in later time-steps than males.

Social Presence

We treated the set of scores on each of the responses to the questions on the social presence or co-presence questionnaire with a 2 x 3 two-way ANOVA analysis. The two independent variables were the gender and virtual human rendering condition that the participant experienced. Out of a total of 5 questions, we found a significant effect of condition on two of them. In response to the question (To what extent did you feel that Bob was sentient, conscious, and alive?), the ANOVA analysis revealed a significant main effect of condition, $F(2, 34) = 3.535$ and $p = 0.041$. Post-hoc Tukey HSD revealed that participants in the Realistic condition (M=5.81, SD=0.43) perceived Bob as a sentient, conscious and alive significantly higher than participants in the Cartoon condition (M=4.25, SD=0.41), $p = 0.012$. In response to the question (To what extent did you feel Bob was only a computerized image and not a real person?), the ANOVA analysis revealed a significant main effect of condition, $F(2, 34) = 3.47$ and $p = 0.048$. Post-hoc Tukey HSD revealed that participants in the Realistic condition (M=2.54, SD=0.48) rated Bob significantly lower in response to this question than the Sketch condition (M=4.0, SD=0.45), $p = 0.039$. The scores for the Cartoon condition (M=3.41, SD=0.46) on this question fell between the Realistic and Sketch conditions.

Based on the log files of the interactions of the participants with Bob in the rapid response simulation, we examined a number of performance related variables namely the average time in minutes spent with the Bob, average questions asked, and the average instrument interactions in measuring Bob's vital signs. These numerical dependent variables were treated with a 2 x 3 two way ANOVA analyses with gender (2 levels) and condition (3 levels) as independent variables. The ANOVA analysis revealed a significant main effect of condition with respect to the average questions asked to the virtual patient Bob, $F(2, 28) = 4.87$ and $p = 0.015$. Post-hoc Tukey HSD revealed that participants asked significantly less number of questions to Bob in the Realistic condition

(M=9.12, SD=0.99) as compared to the Cartoon condition (M=12.28, SD=0.92) p = 0.027. Also, participants asked significantly less number of questions to Bob in the Realistic condition (M=9.12, SD=0.99) as compared to the Sketch condition (M=13.74, SD=1.22), p = 0.007.

Qualitative Results

In order to assess the perceived differences between Realistic, Cartoon and Sketch conditions, at the end of the study we asked participants to report on their overall impressions of interacting with the virtual patient Bob. In response to the question, “what did you like most about Bob?” Participants in the Realistic condition mentioned that Bob, “seemed and felt like a real patient,” ”seemed realistic and interactive, immersive, and felt very real,” and “Bob interacts and answers promptly.” Whereas, participants in the Cartoon condition said, “he could have more facial expressions”, “he was trying to look at the bright side,” ”he looked cartoonish.” Participants in the Sketch condition mentioned, “he was patient”, “his voice made him sound real”. In response to the question, “what did you like least about Bob?” Participants in the Sketch condition said, “I can’t think of Bob as someone in the flesh,” and “can’t think of him as someone exists.” Whereas, participants in the Cartoon condition said that Bob, “was not so cheerful.” Participants in the Realistic condition said, “he followed me with his eyes,” ”his negative attitude”, “he was rude and curt when answering questions” and “his appearance was a bit disconcerting.” In response to the question, “how would you describe Bob’s personality and appearance?” Participants in the Sketch condition said he was, “old, weak and slim” and “old and sick.” Whereas participants in the Cartoon condition were more ambivalent, “he is pleasant, not annoying at all”, “he is trying to be cheerful”, “seemed like a kind person”. Participants in the Realistic condition said, ”friendly old guy under other circumstances, but he seemed to appear more and more agitated, distressed and unhealthy”, “was weary and weak, his personality seemed very grumpy and kind at the same time”, “seemed understanding at first, but then he started acting rude as you can see that he was experiencing more pain.”

CHAPTER FIVE

COCLUSION AND FUTURE WORK

Conclusion

To our knowledge, this research is one of the first in empirically examining the role of realistic versus stylized appearances of a virtual human in a dynamic virtual reality simulation on the emotional reactions of users. In our empirical evaluation, we found that the visual realism of the virtual human had a significant effect on participants' positive and negative emotional reactions in several dimensions of emotion. Our findings have important implications to the design of virtual humans in complex medical trainers and social simulations. Our research suggests that developers of affective agents in VR simulations for entertainment as well as goal-oriented interactions should pay careful attention to the visual realism of the interactive virtual humans as they may have a strong impact on emotional reactions of learners. The results of this study have implications to the Uncanny Valley effect (McDonnell et al., 2012) that we notice with virtual entities in movies and virtual simulations. It has been suggested that abstractions and stylized rendering can lower our expectation of virtual entities, so that we are not detracted by the lack of human idiosyncrasies and subtle appearance cues. Our research shows that interactions with stylized virtual humans in virtual reality systems can also subdue levels of emotional bonding and alter affective reactions to the virtual humans in simulated inter-personal trainers. The diminished responses in interest and other critical affective dimensions may negatively impact higher level functions such as rapport, empathy, and trust formation.

Future Work

To take this research forward, we plan to apply Structural Equation Modelling to understand the relations between the various dependent variables in our study. It would also be interesting to study the effect of the personality and mood of the users on the emotions aroused. One of the limitations of our study is that in the NPR conditions, the virtual environment was not rendered according to the NPR algorithm as well. Instead, we strived to maintain consistency in the look and interactions of participants with the environment, medical instruments, and EHR across all three virtual human rendering conditions, Real, Cartoon and Sketch. Even though none of the participants in the stylized rendering conditions reported any perceived discrepancy in rendering between the virtual human and the environment, future studies will compare the results of the emotional responses to the NPR renderings of the virtual human alone, to conditions in which the virtual human as well as the environment are rendered in a similar manner. Future work will also include an empirical evaluation of the effects of visual realism on participants' perception of sociality of agents, such as persuasion, rapport formation and trust in simulated task-oriented encounters in IVEs.

APPENDICES

Appendix A

Demographics Survey

1. Participant ID:
2. Age:
3. What degree(s) are you currently seeking?
4. Gender:
5. To what extent do you use a computer in your daily activities?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal of use

6. How many hours, on average, do you spend each day on a computer?
7. To what extent have you been exposed to 3-D technology (gaming, movies etc.)?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal of exposure

8. How many hours, on average, do you spend each day interacting with 3-D technology?
9. How many months of hands-on nursing experience have you had (this may include simulations)?

   <1   2-4   4-6   6-12   12-18   >18

10. Have you ever interacted with real patients in a hospital setting?

    Yes   No

11. If you answered yes to the above question, approximately how many months of experience do you have?

    <1   2-4   4-6   6-12   12-18   >18

Appendix B

International Personality Item Pool Interpersonal Circumplex Survey

Answer the following questions on the below scale:

1-Very inaccurate 2-Moderately inaccurate 3-Neither inaccurate nor accurate

4-Moderately accurate 5-Very accurate

	Questions	Response
1.	Am quiet around strangers	
2.	Speak softly	
3.	Tolerate a lot from others	
4.	Am interested in people	
5.	Feel comfortable around people	
6.	Demand to be the center of interest	
7.	Cut others to pieces	
8.	Believe people should fend for themselves	
9.	Am a very private person	
10.	Let others finish what they are saying	
11.	Take things as they come	
12.	Reassure others	
13.	Start conversations	
14.	Do most of the talking	
15.	Contradict others	
16.	Don't fall for sob-stories	
17.	Don't talk a lot	
18.	Seldom toot my own horn	
19.	Think of others first	
20.	Inquire about others' well-being	
21.	Talk to a lot of different people at parties	
22.	Speak loudly	
23.	Snap at people	
24.	Don't put a lot of thought into things	
25.	Have little to say	
26.	Dislike being the center of attention	
27.	Seldom stretch the truth	
28.	Get along well with others	
29.	Love large parties	
30.	Demand attention	
31.	Have a sharp tongue	
32.	Am not interested in other people's problems	

Appendix C

Profile of Mood States

Directions: Describe how you feel right now.

1-Not at all 2-A little 3-Moderately 4-Quite a bit 5-Extremely

	Questions	Response
1.	Friendly	
2.	Tense	
3.	Angry	
4.	Worn out	
5.	Unhappy	
6.	Clear-headed	
7.	Lively	
8.	Confused	
9.	Sorry for things done	
10.	Shaky	
11.	Listless	
12.	Peeved	
13.	Considerate	
14.	Sad	
15.	Active	
16.	On edge	
17.	Grouchy	
18.	Blue	
19.	Energetic	
20.	Panicky	
21.	Hopeless	
22.	Relaxed	
23.	Unworthy	
24.	Spiteful	
25.	Sympathetic	
26.	Uneasy	
27.	Restless	
28.	Unable to concentrate	

29.	Fatigued	
30.	Helpful	
31.	Annoyed	
32.	Discouraged	
33.	Resentful	
34.	Nervous	
35.	Lonely	
36.	Miserable	
37.	Muddled	
38.	Cheerful	
39.	Bitter	
40.	Exhausted	
41.	Anxious	
42.	Ready to fight	
43.	Good-natured	
44.	Gloomy	
45.	Desperate	
46.	Sluggish	
47.	Rebellious	
48.	Helpless	
49.	Weary	
50.	Bewildered	
51.	Alert	
52.	Deceived	
53.	Furious	
54.	Effacious	
55.	Trusting	
56.	Full of pep	
57.	Bad-tempered	
58.	Worthless	
59.	Forgetful	
60.	Carefree	
61.	Terrified	
62.	Guilty	
63.	Vigorous	
64.	Uncertain about things	
65.	Bushed	

Appendix D

Positive and Negative Affect Schedule

Now we want you to indicate to what extent you feel this way at the present moment OR indicate the extent to which you felt this way over the LAST experiment session

	Questions	Response
1.	Interested	
2.	Distressed	
3.	Excited	
4.	Upset	
5.	Strong	
6.	Guilty	
7.	Scared	
8.	Hostile	
9.	Enthusiastic	
10.	Proud	
11.	Irritable	
12.	Alert	
13.	Ashamed	
14.	Inspired	
15.	Nervous	
16.	Determined	
17.	Attentive	
18.	Jittery	
19.	Active	
20.	Afraid	

Appendix E

Differential Emotions Survey

We want you to tell us how strongly you felt each of these feelings during the previous experimental session regarding your interactions with Bob. You can tell us the intensity by assigning a value from 0-9 for each question.

0 (Never) 2 (Barely) 4 (Neutral-) 5 (Neutral+) 7 (Strong) 9 (Extreme)

	Questions	Response
1.	Felt like Bob is interesting	
2.	Felt so interested in interacting with Bob, that you were caught up in it	
3.	Felt alert, curious, or kind of excited about interacting with Bob	
4.	Felt glad about participating	
5.	Felt happy about interacting with Bob	
6.	Felt like Bob's situation is under your control	
7.	Felt surprised, like something suddenly happened with Bob that you didn't expect	
8.	Felt amazed, like you can't believe what's happened with Bob, it was so unusual	
9.	Felt like something unexpected happened with Bob	
10.	Felt unhappy, blue, downhearted about Bob	
11.	Felt sad and gloomy, almost like crying	
12.	Felt discouraged, like nothing is going right with Bob	
13.	Felt like screaming at somebody or banging on something	
14.	Felt angry, irritated, or annoyed	
15.	Felt mad at somebody	
16.	Felt like Bob is not worth your time	
17.	Felt like Bob is "good-for-nothing"	
18.	Felt like you're better than Bob	
19.	Felt scared, uneasy, like something bad might happen with Bob.	
20.	Felt fearful, like you're in danger, very tense	

21.	Felt afraid, shaky, or jittery	
22.	Felt regret, sorry about something you did with Bob	
23.	Felt like you did something wrong with Bob	
24.	Felt like you ought to be blamed for what you did	
25.	Felt embarrassed about making mistakes	
26.	Felt like Bob thought you were incompetent	
27.	Felt like Bob was watching you when something went wrong	
28.	Felt sheepish, like you did not want Bob to look at you	
29.	Felt shy, like you didn't want to interact with Bob	
30.	Felt bashful, embarrassed	

Appendix F

Co-presence questionnaire

1. To what extent did you feel you wanted to help Bob?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal

2. To what extent did you feel you were in the environment with Bob?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal

3. To what extent did you feel Bob was aware of your presence?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal

4. To what extent did you feel Bob was sentient, conscious, and alive?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal

5. To what extent did you feel Bob was only a computerized image and not a real person?

   Never 1 – 2 – 3 – 4 – 5 – 6 – 7 A great deal

Evaluation of Bob

1. What did you like most about Bob?
2. What did you like least about Bob?
3. How would you describe Bob's health and condition?
4. How would you describe Bob's personality and appearance?
5. What do you remember most about Bob?

Appendix G

IRB Approval

Investigator: ______________________________ Participant ID: _______________

Comparing Emotional Responses to Deteriorating Virtual Patients

Description of the Research and Your Participation

You are invited to participate in a research study conducted by Larry F. Hodges, Tracy Fasolino, Sabarish Babu, Shaundra Daily, Lauren Dukes, Jeff Bertrand, and Shelby Darnell. The purpose of this research is to evaluate the emotional impact of a virtual patient's behavior for a prototype system that would be used to train nurses in recognizing the symptoms of a rapidly deteriorating hospital patient. A large portion of the patients who die due to in-hospital adverse events display signs of rapid deterioration, however, registered nurses (RNs) fail to recognize the subtle changes and trends of deteriorating patients. One potential way to help RNs recognize this condition is to use a virtual environment simulation of a hospital and its patients to help in training them to recognize the corresponding symptoms.

The researchers will be happy to answer any questions for you. Your participation will involve:

1. You will be assigned a random number that all your responses will be associated with.
2. Your responses and actions will be audio and video recorded.
3. You will interact while wearing a sensor wristband.
4. You will complete a survey at the end of each timestamp.
5. You will fill out short questionnaires regarding your mood, personality, and demographic information before beginning the simulation.
6. You will complete an end of simulation feedback questionnaire.

The amount of time of your participation will be approximately 1 1/2 hours.

Potential Benefits

The benefits of this research are that you will be able to experience participation in a research study and have the opportunity to interact with virtual characters in a simulated environment. You will also be given the opportunity to be a part of a study that will help contribute to the broader questions of using virtual characters to represent virtual patients in educating nurses. The results of this research may have an impact on how people use virtual environments and virtual humans for education.

Risks and Discomforts

There are no known major risks associated with this research. There is a minor risk that you may experience minor eye discomfort, but no more than if you were playing a video game for an hour or watching an hour of television. Resting periods will be provided. If you experience any discomfort, you may discontinue participation at any time without penalty. Another minor risk is that your assigned number may become connected to your responses. To minimize this risk, the associated numbers and consent forms will be kept in a separate locked cabinet for one year and all digitized data will be stored on a password-protected computer, all of which researchers of this study only have access to. After the duration of approval of this study, all physical and digitized data will be destroyed. Another minor risk associated with wearing the sensor wristband, would be slight discomfort. To minimize this risk, please ask for assistance before adjusting it.

Protection of Confidentiality

We will do everything we can to protect your privacy. Consent forms and associated numbers will be locked in a cabinet and other coded digitized data will be stored on a password protected computer, all of which only researchers of this study have access to. Any information collected will be summarized

This form is valid only if the Clemson University IRB stamp of approval is shown here:

CLEMSON UNIVERSITY
IRB CONSENT FORM
APPROVED 8/11/2014
EXPIRES 8/10/2015

Page 1 of 2

across all participants in this research, so that no information will be presented that may identify you specifically. Your identity will not be revealed in any publication that might result from this study. In rare cases, a research study will be evaluated by an oversight agency, such as the Clemson University Institutional Review Board, Bon Secours St. Francis Health System (funder), or the federal Office for Human Research Protections, which would require that we share the information we collect from you. If this happens, the information would only be used to determine if we conducted this study properly and adequately protected your rights as a participant.

Voluntary Participation

Your participation in this research study is voluntary. You may choose not to participate and you may withdraw your consent to participate at any time. You will not be penalized in any way should you decide not to participate or to withdraw from this study. If you decide not to participate or to stop participating in this study, it will not affect your grade in any way.

If you choose to stop taking part in this study, the information you have already provided will be used in a confidential manner.

Contact Information

If you have any questions or concerns about this study or if any problems arise, please contact Dr. Larry F. Hodges at Clemson University at LFH@clemson.edu or 864.656.7552.

If you have any questions or concerns about your rights as a research participant, please contact the Clemson University Office of Research Compliance (ORC) at 864-656-6460 or irb@clemson.edu. If you are outside of the Upstate South Carolina area, please use the ORC's toll-free number, 866-297-3071.

Consent

I have read this consent form and have been given the opportunity to ask questions. I give my consent to participate in this study. To participate in this study, I assure that:

- **I am at least 18 years of age**
- **I have 20/20 vision or corrected to 20/20 vision.**
- **I use English as my first language and/or am able to communicate in English well.**
- **I have full use of hearing or corrected hearing with use of a hearing aid in at least one ear.**
- **I am a junior or senior level nursing student at Clemson University.**
- **I approve the use of audio/video recording of me while I complete my task.**

Participant's signature: ________________________________ Date: ______________

A copy of this consent form will be given to you.

This form is valid only if the Clemson University IRB stamp of approval is shown here:

CLEMSON UNIVERSITY
IRB CONSENT FORM
APPROVED 8/11/2014
EXPIRES 8/10/2015

REFERENCES

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments, 15*(4), 359-372.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality & Social Psychology Bulletin, 29*(7), 819-833. doi:10.1177/0146167203029007002 [doi]

Bouchard, S., Bernier, F., Boivin, E., Dumoulin, S., Laforest, M., Guitard, T., . . . Renaud, P. (2013). Empathy toward virtual humans depicting a known or unknown person expressing pain. *Cyberpsychology, Behavior, and Social Networking, 16*(1), 61-71.

Cairco, L., Bertrand, J., Gupta, M., Armstrong, R., Babu, S., Hodges, L., & Fasolino, T. (2012). Towards simulation training for nursing surveillance. *Carolinas Women in Computing,*

Campbell, J. C., Hays, M. J., Core, M., Birch, M., Bosack, M., & Clark, R. E. (2011). Interpersonal and leadership skills: Using virtual humans to teach new officers. Paper presented at the *Proc. of Interservice/Industry Training, Simulation, and Education Conference, Paper,* (11358)

Courgeon, M., Buisine, S., & Martin, J. (2009). Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. Paper presented at the *Intelligent Virtual Agents,* 201-214.

Daily, S. B., Meyers, D., Darnell, S., Roy, T., & James, M. T. (2013). Understanding privacy and trust issues in a classroom affective computing system deployment. *Distributed, ambient, and pervasive interactions* (pp. 414-423) Springer.

De Melo, C. M., & Gratch, J. (2009a). Creative expression of emotions in virtual humans. Paper presented at the *Proceedings of the 4th International Conference on Foundations of Digital Games,* 336-338.

De Melo, C. M., & Gratch, J. (2009b). Expression of emotions using wrinkles, blushing, sweating and tears. Paper presented at the *Intelligent Virtual Agents,* 188-200.

Dehghani, M., Carnevale, P. J., & Gratch, J. (2014). Interpersonal effects of expressed anger and sorrow in morally charged negotiation. *Judgment and Decision Making, 9*(2), 104-113.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., . . . Lhommet, M. (2014). SimSensei kiosk: A virtual human interviewer for healthcare decision support. Paper presented at the *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems,* 1061-1068.

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education, 49*(3), 677-690.

Ehrlich, S. M., Schiano, D. J., & Sheridan, K. (2000). Communicating facial affect: It's not the realism, it's the motion. Paper presented at the *CHI'00 Extended Abstracts on Human Factors in Computing Systems,* 251-252.

Ellis, P. M., & Bryson, J. J. (2005). The significance of textures for affective interfaces. Paper presented at the *Intelligent Virtual Agents,* 394-404.

Gaggioli, A., Mantovani, F., Castelnuovo, G., Wiederhold, B., & Riva, G. (2003). Avatars in clinical psychology: A framework for the clinical use of virtual humans. *Cyberpsychology & Behavior, 6*(2), 117-125.

Garau, M., Slater, M., Pertaub, D., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments, 14*(1), 104-116.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 529-536.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1), 1-22.

Healey, J., & Picard, R. (1998). Digital processing of affective signals. Paper presented at the *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference On, , 6* 3749-3752.

Kätsyri, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies, 66*(4), 233-242.

Kwon, J. H., Powell, J., & Chalmers, A. (2013). How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies, 71*(10), 978-987.

Levy, S. (2004). **Why Tom Hanks is less than human**. http://www.newsweek.com/why-tom-hanks-less-human-124405 Accessed Oct 2015 *Newsweek*

Mandryk, R. L., Mould, D., & Li, H. (2011). Evaluation of emotional response to non-photorealistic images. Paper presented at the *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering,* 7-16.

McDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real?: Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG), 31*(4), 91.

Mori, M. (1970). *Bukimi no Tani [the Uncanny Valley], Energy,[Online] 7 (4), 33-35,*

Obaid, M., Damian, I., Kistler, F., Endrass, B., Wagner, J., & André, E. (2012). Cultural behaviors of virtual agents in an augmented reality environment. Paper presented at the *Intelligent Virtual Agents,* 412-418.

Pan, X., Banakou, D., & Slater, M. (2011). Computer based video and virtual environments in the study of the role of emotions in moral behavior. *Affective computing and intelligent interaction* (pp. 52-61) Springer.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies, 62*(2), 231-245.

Qu, C., Brinkman, W., Ling, Y., Wiggers, P., & Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior, 34*, 58-68.

Robb, A., Kopper, R., Ambani, R., Qayyum, F., Lind, D., Su, L., & Lok, B. (2013). Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences. *Visualization and Computer Graphics, IEEE Transactions On, 19*(4), 662-670.

Roter, D. L., Frankel, R. M., Hall, J. A., & Sluyter, D. (2006). The expression of emotion through nonverbal behavior in medical visits. *Journal of General Internal Medicine, 21*(S1), S28-S34.

Ruhland, K., Andrist, S., Badler, J., Peters, C., Badler, N., Gleicher, M., . . . Mcdonnell, R. (2014). Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. Paper presented at the *Eurographics State-of-the-Art Report,* 69-91.

Silverman, B. G. (2004). Toward realism in human performance simulation. *The Science and Simulation of Human Performance, 5*, 469-498.

Slater, M., Khanna, P., Mortensen, J., & Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *Computer Graphics and Applications, IEEE, 29*(3), 76-84.

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior, 27*(2), 741-749.

Tsai, J., Bowring, E., Marsella, S., Wood, W., & Tambe, M. (2012). A study of emotional contagion with virtual characters. Paper presented at the *Intelligent Virtual Agents,* 81-88.

Van Der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion, 11*(2), 286.

Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., & Slater, M. (2006). Building expression into virtual characters.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.

Wellerdiek, A. C., Breidt, M., Geuss, M. N., Streuber, S., Kloos, U., Black, M. J., & Mohler, B. J. (2015). Perception of strength and power of realistic male characters. Paper presented at the *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception,* 7-14.

Wu, Y., Babu, S. V., Armstrong, R., Bertrand, J. W., Luo, J., Roy, T., . . . Fasolino, T. (2014). Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *Visualization and Computer Graphics, IEEE Transactions On, 20*(4), 626-635.

Yu, I., Mortensen, J., Khanna, P., Spanlang, B., & Slater, M. (2012). Visual realism enhances realistic response in an immersive virtual environment-part 2. *Computer Graphics and Applications, IEEE, 32*(6), 36-45.

Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., & Hodges, L. F. (2007). Social responses to virtual humans: Implications for future interface design. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 1561-1570.

Zibrek, K., & McDonnell, R. (2014). Does render style affect perception of personality in virtual humans? Paper presented at the *Proceedings of the ACM Symposium on Applied Perception,* 111-115.