Empirical Evaluation of the Interplay of Emotion and Visual Attention in Human-Virtual Human Interaction

Matias Volonte
Clemson University
mvolont@clemson.edu

Reza Ghaiumy Anaraky
Clemson University
rghaium@clemson.edu

Bart P. Knijnenburg
Clemson University
bartk@clemson.edu

Andrew T. Duchowski
Clemson University
aduchow@clemson.edu

Sabarish V. Babu
Clemson University
sbabu@clemson.edu

ABSTRACT

We examined the effect of rendering style and the interplay between attention and emotion in users during interaction with a virtual patient in a medical training simulator. The virtual simulation was rendered representing a sample from the photo-realistic to the non-photorealistic continuum, namely Near-Realistic, Cartoon or Pencil-Shader. In a mixed design study, we collected 45 participants' emotional responses and gaze behavior using surveys and an eye tracker while interacting with a virtual patient who was medically deteriorating over time. We used a cross-lagged panel analysis of attention and emotion to understand their reciprocal relationship over time. We also performed a mediation analysis to compare the extent to which the virtual agent's appearance and his affective behavior impacted users' emotional and attentional responses. Results showed the interplay between participants' visual attention and emotion over time and also showed that attention was a stronger variable than emotion during the interaction with the virtual human.

CCS CONCEPTS

• **Human-centered computing → Human Computer Interaction**; • **Interaction paradigms →** *Virtual reality*.

KEYWORDS

Virtual Human, Human-Computer Interaction, eye tracking

ACM Reference Format:
Matias Volonte, Reza Ghaiumy Anaraky, Bart P. Knijnenburg, Andrew T. Duchowski, and Sabarish V. Babu. 2019. Empirical Evaluation of the Interplay of Emotion and Visual Attention in Human-Virtual Human Interaction. In *SAP '19: ACM Symposium on Applied Perception (SAP '19), September 19–20, 2019, Barcelona, Spain.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343036.3343118

SAP '19, September 19–20, 2019, Barcelona, Spain

ACM ISBN 978-1-4503-6890-2/19/09…$15.00
https://doi.org/10.1145/3343036.3343118

1 INTRODUCTION

Virtual humans (VHs) have been successfully employed for training by multiple industry fields. For example, the medical arena incorporated this type of tool for instructing their employees. Healthcare trainees can learn procedures and gain experience on identifying signs and symptoms of deterioration by interviewing virtual patients (VP) [Jung et al. 2005]. However, in order to build an effective virtual training tool, it is important to understand and control the artistic and technical aspects of it. One key factor of the virtual environment is the rendering algorithm that it will possess.

The appearance and behaviors of the virtual human will impact the users' visual attention and emotion distinctively. An investigation examined the effect of photo-realistic and non-photorealistic render styles on participants' emotions during interaction with a virtual patient [Volante et al. 2016]. Results showed that users' positive emotions decreased while negative emotions increased over time in all the render conditions. Also, there is evidence that humans' visual attention is affected by the different rendering styles of a VH [Carter et al. 2013]. Furthermore, there are indications that emotional cues from virtual agents can influence human attention and behavior [Beale and Creed 2009] and that emotion can influence the speed with which affective non-verbal expressions are identified or recognized [Eastwood et al. 2001]. This is important, considering that generating an emotional response from participants when working with learning systems is critical, as it seems to strongly influence memory retention [Dunsworth and Atkinson 2007].

Additionally, studies showed that human emotion and attention are closely linked together. Reports suggest that emotional responses can be regulated through visual attention manipulation. Experiments in the neuroscience field demonstrated that users' regulated their emotions when exposed to stressful stimuli by shifting their visual attention [Corbetta and Shulman 2002]. This mechanism is defined as "affect-biased attention" and refers to the selective attention process by which sensory systems are tuned to favor certain categories of affectively salient stimuli over others. For example, by biasing perception towards certain types of positive or negative stimuli, habitual affect-biased attention may modulate emotional responses to stressful events [Wadlinger and Isaacowitz 2011].

Our current investigation focuses on the interplay between the user's visual attention and emotion during interaction with a virtual patient in a medical training scenario. In a mixed design study, users interacted with a VP presented in a distinct rendering style,

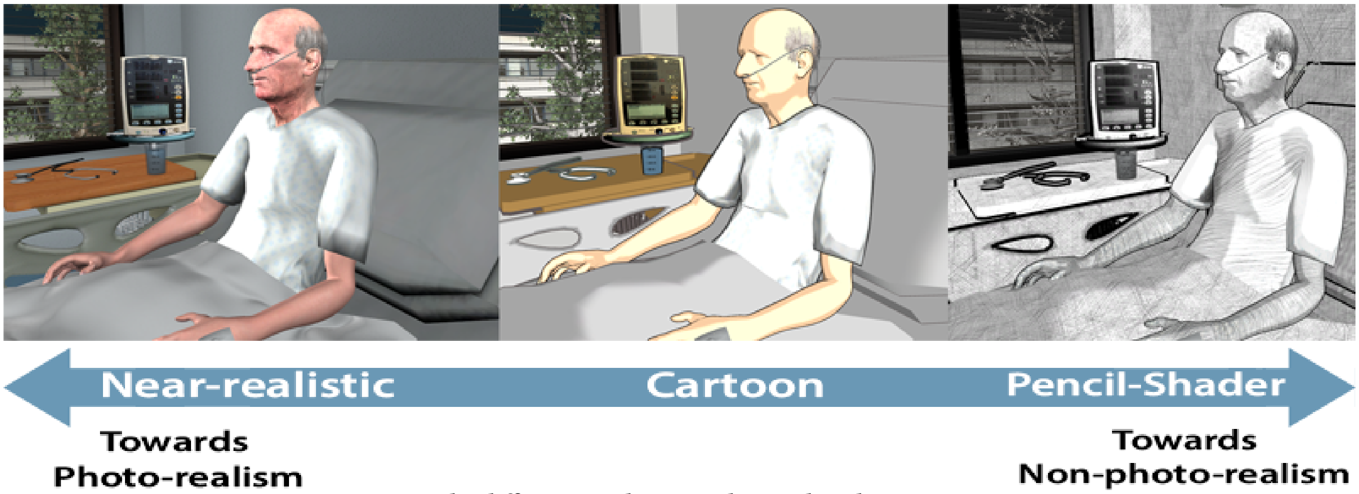

Figure 1: The different rendering styles used in the experiment.

namely Near-Realistic (NR), Cartoon (CT) or Pencil-Shader (PS). In this medical interactive simulation, the virtual patient's vital signs, appearance and cognitive processes decline across time. We collected users' emotional response and gaze in order to analyze their reciprocal relationship over time. We also produced a mediation analysis between attention and emotion over different conditions of virtual human rendering.

The results from this study will be beneficial for the virtual reality (VR) community and for developers. Also, they will be useful for simulations in which participants are expected to visually attend the virtual human behavior since selecting the appropriate rendering algorithm can deter or enhance users' emotions and gaze which in turn directly impacts the simulation success.

Finally, measuring how different rendering styles of virtual human alters the interplay between the users' emotions and attention has not been explored deeply yet. To the best of our knowledge, the existing literature is mostly focused on understanding how users' attention or emotion was altered during interaction with such stimulus.

2 RELATED WORK

Attention is the means by which information is filtered and selected for processing stimulus-driven and voluntary tasks [Posner 1982]. On the one hand, stimulus-driven tasks are strongly influenced by the properties of the stimulus and are often processed involuntarily and automatically. On the other hand, goal-directed tasks are mostly processed voluntarily. For example, stimuli with strong exogenous properties are often used as distractors while subjects perform voluntary tasks [Vanhala et al. 2010].

Emotion is an event-focused process consisting of specific elicitation mechanism based on the relevance of a stimulus that shapes an emotional response across several systems (behavioral, physiological, etc). Emotions are elicited as the individual evaluates situations relevant to his/her needs, goals, values, and general well-being. The detection of a relevant event elicits an adaptive emotional response that mobilizes resources that allow the individual to cope with the situation [Brosch et al. 2013].

Futhemore, emotion can also serve to draw or focus a user's attention, in both conscious and unconscious ways. Significant emotional stimuli such as food, mating partners, or signals of threat should be particularly effective cues to capture attention [Lang et al. 2013]. For instance, it was found that fear-related pictures of snakes and spiders were detected faster in grid-pattern arrays of fear-irrelevant pictures (flowers and mushrooms [Öhman 2005]. Also, emotional expressions can influence the speed with which emotional faces are identified or recognized [Eastwood et al. 2001]. Nevertheless, studies suggest that humans tend to be predisposed to attend to certain categories of affectively salient stimuli over others. Studies argue that this is a strategy humans use to increase, maintain, or decrease components of an emotional reaction [Gross 2001; Todd et al. 2012]. This regulation encompasses a wide array of mechanisms, characterized from implicit to explicit and reactive to effortless, that can be applied in anticipation of, or in response to, an emotional stressor [Eisenberg et al. 2010].

Furthermore, has been hypothesized that visual attention can be used as a tool for emotion regulation [Wadlinger and Isaacowitz 2011]. Individuals can focus the sensory stimuli in their environments activating a selective attention mechanism [Parkhurst et al. 2002]. This is because users attentional resources are limited; Therefore, it is likely that the salient stimuli that capture our attention will also direct our future choices and behaviors [Pashler et al. 2001]. Which stimuli individuals find to be salient is not merely random or accidental but is related to person-level variables such as motivation. Also, research states that salient sensory, emotional, and mental information is filtered, processed, and analyzed through various attentional structures, which can be automatically or consciously regulated [Calvo and Nummenmaa 2007].

The literature focused on the interplay between visual attention and emotion is vast. In a study conducted by Schupp et.al [Schupp et al. 2007], participants were exposed to a stream of high arousing and low arousing visual material. The study showed a variation of brain event related measures on the participants based on the the different presented stimuli. These results seem to indicate that emotion augmented attention effects.

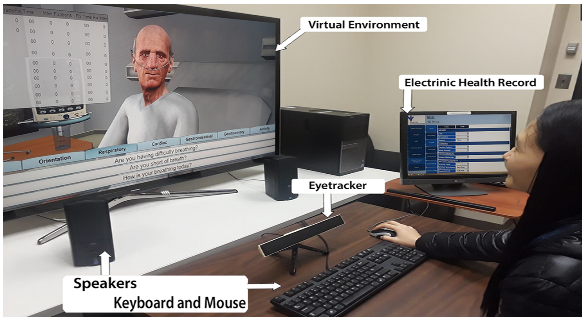

Figure 2: Participant during interaction with the virtual patient in the Rapid Response Training System.

Other research has explored the relationship between emotion and animation and their effect on emotion contagion in VR. This research suggests that animation of a virtual agent can have a positive impact in mediating the *Uncanny Valley* effect with virtual agents [Wu et al. 2014]. Moreover, this research propose that virtual agent animation is important in eliciting emotional responses of participants in simulated human-virtual human interaction. This elucidates the importance of virtual human animation for generating compelling and emotionally engaging reactions in users.

Prendinger et al. [Prendinger et al. 2006] examined users' gaze behavior when interacting with life-like interface agents in a website platform. Results showed that visual attention was drawn to the virtual agent, particularly when the agent produced deictic gestures, and that users directed their attention to the objects the agent gestured towards. Additionally, visual attention was mostly drawn to the agent's face.

Studies have shown that during conversations humans tended to gaze to VHs in a similar way they do to humans [Rehm and André 2005; Robb et al. 2016]. However, participants in this study spent significantly more time looking at virtual agent than the time they usually do to other humans. A possible explanation could be that due to the limited non-verbal cues expressed by virtual humans, participants gazed at them for longer time.

Pence [Pence et al. 2014] examined different layouts configuration and the effect on participants' visual attention in a pediatric VR taining simulator. Results showed that little visual attention was directed towards the virtual patients. The majority of time was spent gazing at the interface elements that were used to perform the interview task with the virtual patient.

3 EXPERIMENT SIMULATION AND SETUP

The current study was conducted in a medical training simulator conceived for training nurses in the recognition of signs and symptoms of rapid deterioration of patients. This simulator, named the Rapid Response Training System (RRTS), reproduced nurses' daily tasks which include performing medical rounds four times a day, gathering the vital signs of the patients and reporting them in an Electronic Health Record System (EHR).

The experiment setup consisted of a 65 inches display and a 21 inches monitor dual screen configuration. The large screen displayed the virtual environment and the virtual human at life-sized scale while the monitor showed the EHR form where users could input the quantitative and qualitative data gathered during the interaction with the patient (see Figure 2).

All the elements and procedures in this simulation were carefully created with the advice of medical experts. These includes but are not limited to the behaviors of the virtual patient (VP), the digital instrumentation, and the requested medical procedures users performed. For example, the course of the simulation was divided in four distinct time-steps (TSs) since these are the periods that nurses visit their patients in this medical facility. Finally, the virtual patient's health, appearance and cognitive reflexes declines after each time-step. This medical deterioration corresponds to real medical situations that nurses experienced during the course of their daily shift.

In this system, participants could interact with the VP by accessing a set of pre-defined medically relevant questions or by utilizing medical instruments present in the patient's environment to measure his vital signs. The information gathered through the questionnaire and by the use of medical instrumentation was recorded in the simulated EHR system. In this research, the RRTS served as a rich experiment platform for empirically studying how factors associated with the appearance of the virtual human can affect the users' emotional state and visual attentional response.

The current experiment presented three conditions of render styles of the virtual human and the virtual environment. These are Near-Realistic (NR), Cartoon (CT), and Pencil-Shader (PS). All the elements in the simulation in all conditions remained legible and clear and none of the participants reported any problem on understanding the displayed information.

For the photorealistic (PR) sample, we created the Near-Realistic condition. For this scenario the skin textures of the virtual patient included details such as wrinkles and blemishes to increase realism. Furthermore, the virtual environment used realistic texture maps and high definition range image based lighting. For the non-photorealistic (NPR) conditions, we rendered the virtual human and the environment using Cartoon and a Pencil-Shader materials (see Figure 1). The Cartoon condition used Unity3D's Basic Outline Toon shader. This material gives the character a uniform outline with simplified two-color shading. Only the skin and eyeball textures maintained 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. The Pencil-Shader 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 denotes a lighter tone and denser strokes a darker tone. Cross-hatching is also applied to create darker shades by layering strokes at different angles.

We aimed to provide a consistent user experience among experiment conditions within the RRTS, to avoid any possible confounds due to differences in virtual environment and instrument appearance. Thus, the only controlled difference in the experimental conditions were the rendering styles; all virtual human animations and behaviors, as well as the instrument interactions in the RRTS were

Figure 3: Timeline of the experiment that participants from the Near-Realisitic (NR), Cartoon (CT) and Pencil-Shader (PS) experienced in this study.

consistent. Since our main interest is in investigating how different rendering styles affect users' visual attention and emotion, we selected these specific rendering styles based on the existing literature of realistic and non-photorealistic computer graphics [Halper et al. 2003; MacDorman et al. 2009; McDonnell et al. 2012; Zibrek and McDonnell 2014].

3.1 Research Questions and Hypotheses

In the present examination we are interested in answering the following questions:

(1) which one was the stronger variable between the users' emotion and attention during the interaction with the virtual human at different time-steps of the experiment?. This question seeks to answer if users' attention influenced their emotion or if the users' emotion influenced their attention during the different time-steps of the simulation.
(2) Over the four time-steps of the simulation, to what extent did the negative affective behaviors of the virtual humans alter users' emotion and attention?.

Moreover, based on existing literature on virtual humans we produced the following hypothesis:

(1) Users' emotions would be the stronger response than attention over the four time-steps of the simulation. We based this hypothesis on the literature that suggest that increasing affective content seems to produce higher levels of arousal in humans [Bernat et al. 2006].
(2) The virtual patient's incremental negative affective behavior will increase user's emotion and visual attention scores. We based this hypothesis in literature that states that emotional content seems to have more influence on the users' attention over neutral imagery [Nummenmaa et al. 2006].

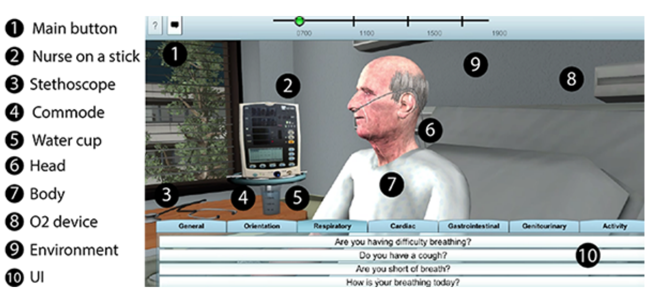

Figure 4: Locations where gaze allocation was tabulated.

(3) The CT rendering style condition will score higher visual attention values than the PS or NR rendering styles. We base this decision in the current literature that suggests that agents that possess cartoon like rendering style are perceived to be more amicable and agreeable [McDonnell et al. 2012; Zibrek and McDonnell 2014].

3.2 Participants

We recruited a total of 45 participants, male (N=28) and female (N=17), between the ages of 18 and 50 (mean age 26.17) from a University Campus.

3.3 Study Design

The experiment was of a 3×4 mixed factorial design. Independent variables were the rendering styles of the virtual simulation, and the four time-steps (TSs) during which the virtual patient's distress increased. The between-subjects condition was the rendering style: Near-Realistic (N=15), Cartoon (N=15) and Pencil-Shader (N=15). The within-subjects condition was the virtual patient emotional distress at each time-step of the simulation. As time progressed, the virtual human's vital signs declined rapidly, and this was reflected by his mood and declination of his cognition, and his appearance.

3.4 Methodology

After obtaining consent, participants filled a demographic questionnaire and surveys regarding their current disposition. Next, the experimenter carefully explained the study and showed how the simulation functioned. Then, users practiced the procedures needed to perform in the study until they stated they had fully understood it. Then, the experimenter proceeded to calibrate the eye tracker using a 9-point calibration sequence provided by the Gazepoint GP3 eye tracker software. Following successful calibration, participants started the first time-step and were asked to interact with the VP, Bob, by asking questions, using the virtual instruments as needed to medically assess his condition, and to record their observations in the EHR system. At the end of each TSs, participants filled out a Differential Emotion Survey (DES) in order to keep track of their current emotional state after the interaction. Finally, at the end of the fourth time-step, users filled out the DES survey [Van Der Schalk et al. 2011] and were debriefed. Figure 3 shows the experiment time-line for the different conditions.

3.4.1 Statistical analysis. In order to determine if it is attention affecting or predicting the users' emotion or vice-versa over the course of the experiment, we implemented a cross-lagged panel model (CLPM) [Rogosa 1980]. This is a statistical method that can be used to analyze reciprocal relationship or directional influences between variables over time [Kearney 2017; Kenny 1975]. The CLPM estimates relationship between one variable to another ("crossed") over different periods of time ("lagged"). This method provided us with statistical evidence about the relationship and directionality between variables over the time-steps of the experiment. The result of the CLPM will show which was the stronger predictor between the variables of study (Attention and Emotion). This was important for producing a data driven mediation model between different rendering styles, Attention and Emotion.

4 DEPENDENT MEASURES

During the experiment we collected the participants' gaze using a Gazepoint Eye Tracker GP3 sampling at 60 Hz. Then, in an off-line process we replayed it and emitted a ray using the average of their left and right eye positions (see Figure 5). In case of a collision between the emitted ray and a geometry of concern we computed our variables on interest. We grouped the scene objects into 10 categories, shown in Figure 4. For the final analysis, we collapsed them into four groups: Virtual Human (VH), User Interface (UI),

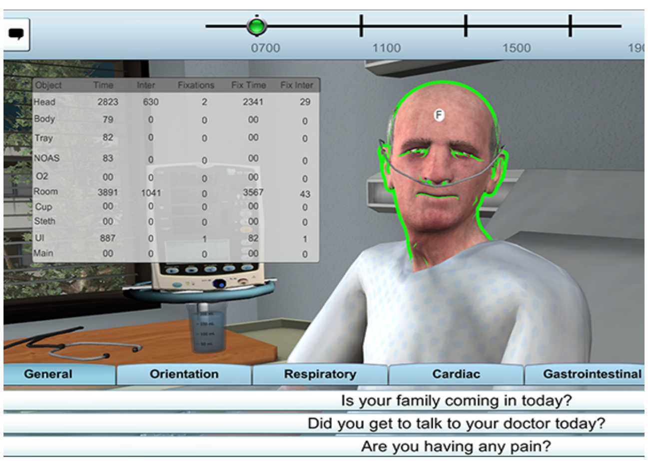

Figure 5: Image of the post visualization tool interface.

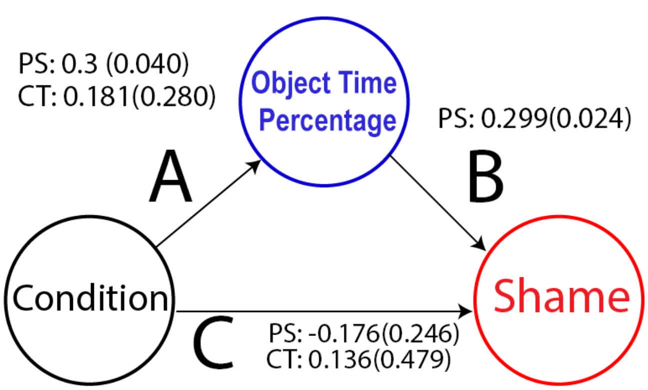

Figure 6: Graph showing the Beta coefficients and P-values in parenthesis of the mediation effects between *Object Time Percentage* and *Shame* on time-step 1 of the simulation.

Tools, and Environment. However, we focused our analysis studying how users' visual attention and emotional response varied between conditions only towards the virtual human.

Users' gaze yielded several dependent variables used to measure visual attention of the users during the interaction with the virtual human.

4.1 Fixation Categorization

We processed the collected data and categorized human eye motion into fixations and saccades [Duchowski 2017]. Fixations occur when gaze remains focused on a specific area for longer than a given threshold while saccades are rapid, discontinuous motions where the eyes move from one fixation point to another. We intended to detect where the eye fixates and when the motion signal changes abruptly, indicating the end of a fixation and the onset of a saccade.

In our analysis, fixations and saccades were detected by a velocity-based algorithm using the Savitzky-Golay filter as suggested by Nyström and Holmqvist [Nyström and Holmqvist 2010]. This algorithm functions assuming that the eye movement signal is recorded at a uniform sampling rate. Successive samples are differentiated to estimate eye movement velocity. Fixations are either implicitly detected as the portion of the signal between saccades, i.e., the portion of the signal where the velocity falls below a threshold. We used the Savitzky-Golay filter for differentiation [Savitzky and Golay 1964], preceded by smoothing of the raw data with a 2nd order low-pass Butterworth filter with sampling and cutoff frequencies set to 60 and 1.65 Hz, respectively.

4.2 Visual Attention Categorization

We categorized visual attention as *Object Time Percentage*, calculated as the amount of time gaze was elicited towards the virtual human during conversation divided by the total conversation time in that time-step. We also derived the *Fixated Time Percentage* which is the total time fixation towards the virtual human divided by the total amount of time participants fixated on all the objects in that time-step. This variable was implemented to analyze the Fixated Time users gazed towards the virtual human.

The difference between Object Time Percentage and Fixated Time Percentage is that the first is the percentage of all the time (independently if there was a fixation or not) that users gazed towards the virtual human while the second is only the percentage of time calculated only when there was a gaze fixation toward the VH.âĂİ Finally, the *Transitions per Minute* was the total number of gaze transitions from the conversational UI to the virtual human during conversation per minute at that time-step. This variable was conceived to examine the number of "quick gazes" users performed during simulated dialogue on the experiment.

4.3 Emotional response Measures

The Differential Emotion Scale (DES) questionnaire in our experiment was based on the DES IV modified to reduce the item count to 30 [Van Der Schalk et al. 2011]. There are 10 emotion categories scored in the DES: ***Interest, Enjoyment, Surprise, Sadness, Anger, Contempt, Fear, Guilt, Shame and Shyness.*** Participants used a 0-9 Nominal scale to express how strongly they felt about each item, with 0 indicating Never and 9 indicating Extreme.

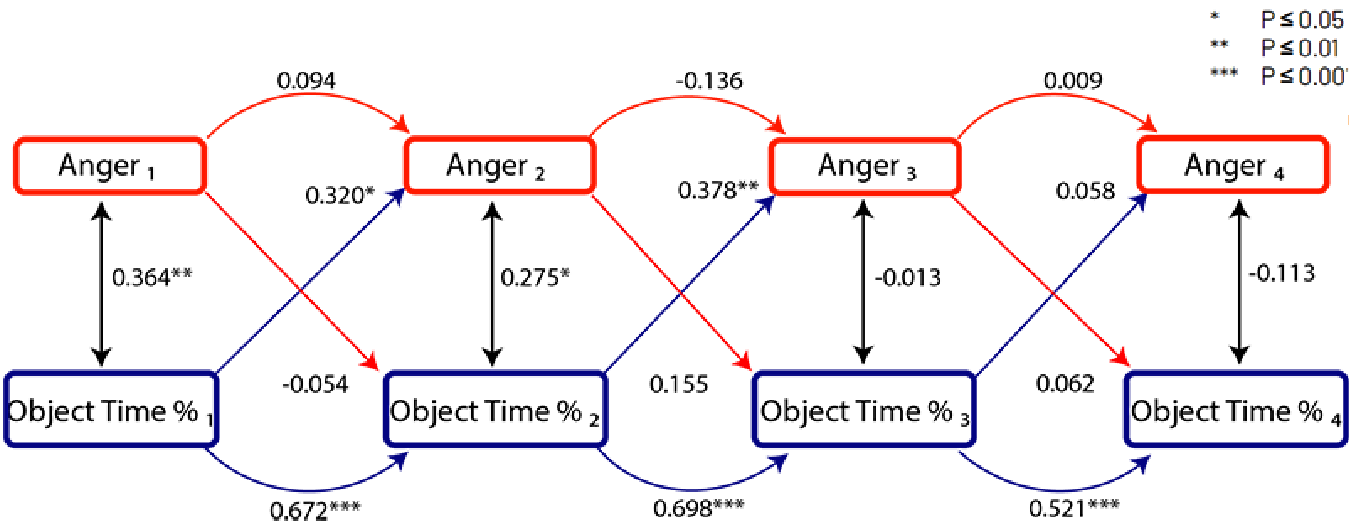

Figure 7: Graphical representation of the cross-lagged panel model used in this study. We presented the *P* and Beta values. Black arrows indicate correlation between emotion and attention and the colored arrows show regressions during time-steps between the different conditions of the Virtual Human.

Table 1: Results from the regressions showing Beta coefficients and P values in parenthesis.

Object Time Percentage (OTP)

	Condition on Attention					Attention to Emotion			Condition to Emotion	
	TS 1	TS 2	TS 3	TS 4		TS 1	TS 2		TS 1	TS 2
OTP	PS, 0.3(0.04)	PS, 0.426(0.002)	PS, 0.276(0.047)	PS, 0.3(0.037)	Interest		0.378(.009)	Interest		
					Enjoy		0.312 (0.5)	Enjoy		PS, 0.385 (0.016)
					Surp		0.385 (0.008)	Surp		
					Distr	0.246 (0.05)		Dist		CT, 0.361 (0.010)
					Anger	0.459 (<0.001)	0.385 (0.008)	Anger	PS, -0.414 (0.003)	
					Cont		0.368 (0.043)	Cont		
					Fear	0.64 (<0.001)		Fear		
					Guilt	0.337 (0.015)		Guilt	PS, -0.314 (0.031)	
					Shame	0.299 0.024)		Shame		
					Shy	0.38 (0.05)		Shy	PS, -0.320 (0.018)	

Table 2: Results from the regressions showing Beta coefficients and P values in parenthesis.

Fixated Time Percentage (FTP)

	Condition on Attention			Attention to Emotion			Condition to Emotion	
	TS 1	TS 2		TS 1	TS 2		TS 1	TS 2
FTP	PS, 0.289 (0.05)	PS, 0.385 (0.004) CT, 0.325 (0.05)	Interest		0.337 (0.021)	Interest		
			Enjoy		0.333 (0.038)	Enjoy		PS, 0.289 (0.05)
			Surp		0.375 (0.012)	Surp		
			Dist			Dist	CT, 0.358 (0.012)	
			Anger	0.474, (<0.001)	0.446 (0.001)	Anger	PS, -0.413 (0.003)	
			Cont			Cont		
			Fear	0.595 (<0.001)		Fear		
			Guilt	0.307 (0.023)		Guilt	PS, -0.302 (0.037)	
			Shame	0.295 (0.025)		Shame		
			Shy	0.335 (0.017)		Shyn	PS, -0.302 (0.037)	

5 RESULTS

5.1 Quantitative results

5.1.1 Cross-lagged Panel Model. Figure 7 shows the result of our CLPM for "Anger" DES dimension. In this model attention in time 1 and 2 could predict emotion in time 2 and 3 respectively. We observed the same pattern for all the emotions. Also, we observed the same pattern for the different rendering style conditions so we reported the overall model. Based on these results, in the mediation analysis we produced a mediation path from Attention to Emotion.

5.1.2 Mediation Analysis. We implemented a path analysis to evaluate our independent and dependent variables. This analysis is a graphical representation of multiple regression models that shows the relationship between dependent and independent variables. Figure 6 depicts the multilevel path model used for analyzing the causal (reciprocal) relationship between emotion and attention during the different TSs of the experiment. We formed this model based on the results of our cross-lagged panel model that showed that Attention was a predictor of Emotion during the different time-steps of the simulation. Figure 7 shows the Beta coefficients and P-Values from the regressions.

The graph contains the dependent variables in circles and the arrows show linear regressions. The arrows include the standardized beta coefficient that represents the effect size and also a *p*-value showing significance of the effect.

5.1.3 Object time Percentage. In our mediation analysis, a regression was used to investigate the extent that *Attention* mediates the effect of *Render Style* of virtual human on *Emotion*. Results indicated that the *Render style* of the virtual patient was a positive significant predictor of *Attention* in all the time-steps of the simulation. Results also showed that *Attention* was a positive significant predictor for *Emotion* in time-steps 1 and 2 of the simulation. Furthermore, analysis revealed that the PS condition had a direct negative effect on *Anger*, *Guilt* and *Shyness* in time-step 1 and a negative direct effect of all render styles on *Anger*. In time-step 2, The PS style had a direct effect on *Enjoyment/Joy* and the cartoon condition directly affected users' distress. Refer to the Table 1 for a detailed description of the results.

5.1.4 Fixated time Percentage. Results indicated that the *Render style* of the virtual patient had a significant positive effect on *Attention* in time-steps 1 and 2 of the simulation. On time-step 1, only the PS render style had an effect on Attention while on time-step 2, the PS and CT render styles had a direct positive effect on *Fixated Time Percentage* metric.

The analysis also revealed that *Attention* had a positive causal influence on *Anger*, *Fear* and *Terror*, *Guilt*, *Shame* and *Shyness* categories of *Emotion* in time step 1. In time step 2, Attention positively influenced *Interest*, *Enjoyment/Joy*, *Surprise*, and *Anger*. Furthermore, analysis revealed that the PS render condition had a direct negative effect on *Anger*, *Guilt* and *Shyness* in time-step 1. In time step 2, the PS style had a direct effect on *Enjoyment/Joy*. Refer to the Table 2 for a detailed description of the results.

5.1.5 Transitions per Minute Results. During simulated conversations, the PS condition had a significant positive effect on users' Attention in time-steps 2, 3 and 4 of the simulation.

Table 3: Shows the significant findings between the different conditions, attention and emotion during the different time-steps of the simulation.

Transitions per Minute (TPM)

	Cond. to Atte.			Atte. to Emo.
	TS 2	TS 3-4		TS 1
TPM	PS 0.343(0.019)	PS 0.343(0.017)	Int	0.315(0.026)
			Enj	
			Surp	0.352(0.006)
			Dist	0.247(0.037)
			Ang	0.177(0.05)
			Cont	
			Fear	0.253(0.02)
			Gui	
			Sha	
			Shy	

Analysis also revealed that Attention was a significant positive influence on *Interest*, *Surprise*, *Distress*, *Anger* and *Fear* categories of Emotion but only in time-step 1. Table 3 shows results of the regression analysis.

6 DISCUSSION

We empirically analyzed how different render styles of a virtual human and environment altered users' attention and emotion in an interactive virtual training simulator. The render styles conditions represented a sample from the photorealistic to non-photorealistic continuum, namely Near-Realistic (NR), Cartoon (CT) or Pencil-Shader (PS) styles (See Figure 1). In a mixed designed study, users related with a virtual human that exhibits negative affective behaviors due to health declination over the course of four time-steps.

During the experiment, we collected users' emotional disposition through the DES questionnaire and their gaze by an eye tracker. After off-line processing, we produced a cross-lagged panel analysis in order to determine which was the stronger predictor between attention and emotion during the course of the experiment (see Figure 7). Based on the the cross-lagged results we analyzed how emotion varied based the different rendering condition styles and how this was mediated by attention (See Figure 6). Finally, it is important to point out that we interpret increased attention as a sign of attraction to the stimulus, however, it could also interpret this as a sign of repulsion[Cheetham et al. 2013].

To the question: *"which one was the stronger variable between the users' emotion and attention during the interaction with the virtual human at different time-steps of the experiment?"*, results showed that attention was a stronger causal influence of emotion for the subsequent time-steps of the simulation. We could produce such statement since no differences between rendering groups were found. This means that users from the near realistic to the non-photorealistic conditions reacted in a similar manner (Figure 7). The results from the cross-lagged panel suggest that the path in the mediation model should derive from Attention to Emotion.

To the question: *"Over the four time-steps of the simulation, to what extent did the negative affective behaviors of the virtual humans alter users' emotion and attention?"*, our results suggest that the negative affective behaviors altered the users' emotions and visual attention significantly. Recall that the distress of the virtual patient increases as time progresses, showing a very noticeable declination of his vital signs, behavior, and appearance in TS 3 and 4. This evident negative conduct of the virtual human seems to have impacted the users' attention and emotion since the most significant results are found in the earlier time-steps of the simulation (See Table 1 and 2). During TS 1 and TS2, when the virtual human's vital signs are still not severely negative, results are mostly significant.

Analysis of *Object Time Percentage* and *Fixated Time Percentage* showed a similar pattern between the different pathways of rendering styles on emotion with mediation of attention. Overall, the Pencil-Shader condition drew more attention than the Near-Realistic style in all the time-steps (*Object time percentage*) of the simulation and in the first two time periods (*Fixated Time Percentage*).

Furthermore, attention showed a stronger influence on emotion in the fist two time-steps with very mild influence in the third time period. Finally, users in the PS rendering style condition showed a negative relationship between *Anger*, *Guilt* and *Shyness* while users shown the Cartoon style showed a positive relation to *Distress*.

The *Transition per Minute* measure is a measure that indicates the number of times users gazed towards the virtual human from the conversational User Interface questionnaire during simulated dialog. A granular analysis of this data shows that rendering style was significant in the PS over the NR condition in time-steps 2, 3 and 4. Additionally, attention significantly influenced *Interest*, *Surprise*, *Distress*, *Anger* and *Fear* only in time-step 1 of the simulation. In this period the virtual patient presents a very mild negative affective behavior during the simulation.

Based on our findings, we cannot support our hypothesis that users' emotions would be a stronger causal influence than attention during the four time-steps of the simulation since our cross-lagged model showed that attention produced the stronger influence.

Based on our results from analysis of *Object time Percentage* and *Intersections per Minute*, we support our hypothesis stating that negative emotional behavior of the virtual patient over time would impact visual attention

Finally, the results from our mediation analysis showed that the most significant effects occurred in the PS condition; Therefore, we reject our hypothesis that the virtual human Cartoon condition would elicit the higher attention scores.

7 CONCLUSION

To our knowledge, this research is one of the first in empirically examining the mediation of realistic and stylized virtual human on users' emotion and attention in an interactive training simulator.

First, our cross-lagged analysis showed that attention had higher influence over the users' emotions during the different time-steps of the simulation. Participants' visual attention, independently on the condition that participants were exposed to, prioritized their visual input over their inner emotional response. This finding highlights the importance that the visual stimuli had over emotion during the extent of an interaction with a virtual human.

Second, our mediation analysis showed that users gazed more at the virtual human depicted using a pencil shader rendering style during the four time-steps of the simulation. In addition, visual attention only influenced users' emotions during time-step 1 and 2, when the virtual patient presents the lowest negative affect behaviors. However, during the later time-steps when the virtual human presents the higher emotional stress behaviors, users' attention did not alter their emotions.

Our findings have direct impact for designing virtual humans in medical training and social simulations. Selecting the correct rendering style algorithm is critical since it can enhance or discourage users' visual attention towards specific objects in the simulation. Specifically, this decision is imperative in medical training scenarios where users are expected to gaze to the virtual human for monitoring and examining closely their behaviors and manifestations of illnesses for data gathering. Moreover, visual attention did not alter users' emotions in the later and critical time-steps of the simulation. This is an unfavorable effect in educational training simulators since emotion and cognition are closely linked [Lazarus 1982].

Our results have implications to the *Uncanny Valley* effect existing in synthetic digital characters. Presenting anthropomorphic virtual agents depicted realistically, can affect users' expectations not only about their technical capabilities but also users' can notice imperfections on the animations, facial expressions and cognition. This can have a detrimental effect on visual attention towards characters presenting realistic rendering styles. Therefore, in simulations where the virtual character mimics a real human but does not possess the technical abilities to show realistic animation and capabilities close to a real human, a non-realistic appearance might be more effective [Knijnenburg and Willemsen 2016].

Several hypothesis could explain the users' emotional diminishing during time-step 3 and 4. Perhaps users used their gaze as an adaptive strategy to regulate their emotional experience while interacting in the simulation [Todd et al. 2012; Wadlinger and Isaacowitz 2011]. Since the virtual patient shows the highest levels of negative emotional distress in the time-steps 3 and 4, maybe users' shifted their gaze away from the virtual human and focused their efforts on completing the task. This type of tunnel vision effect has been noticed in medical practitioners when handling critical situations in a hospital setting and could potentially explain the results found in this study [Ellis and Bryson 2005]. Finally, another possibility could be that prolonged exposure to the simulation might have caused a reduction on sensitivity to that stimulus; therefore, a decrease in the users' emotions [Cheetham et al. 2013].

8 LIMITATION AND FUTURE WORK

A limitation of the current investigation is that the rendering quality of the Near-Realistic condition style could be enhanced. Despite the fact that non of the participants reported any complains about the appearance of this condition, current technology provide the tools for creating more realistic virtual humans. Moreover, the dialogue metaphor for interacting with the virtual patient could be natural speech rather than a graphical user interface.

In a future study, we would like to compare learning outcomes between the current desktop VR scenario setup and an immersive virtual reality environment. In this experiment we would like to compare social presence, task performance and rapport.

REFERENCES

Russell Beale and Chris Creed. 2009. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies* 67, 9 (2009), 755–776.

Edward Bernat, Christopher J Patrick, Stephen D Benning, and Auke Tellegen. 2006. Effects of picture content and intensity on affective physiological response. *Psychophysiology* 43, 1 (2006), 93–103.

Tobias Brosch, Klaus R Scherer, Didier Maurice Grandjean, and David Sander. 2013. The impact of emotion on perception, attention, memory, and decision-making. *Swiss medical weekly* 143 (2013), w13786.

Manuel G Calvo and Lauri Nummenmaa. 2007. Processing of unattended emotional visual scenes. *Journal of Experimental Psychology: General* 136, 3 (2007), 347.

Elizabeth J Carter, Moshe Mahler, and Jessica K Hodgins. 2013. Unpleasantness of animated characters corresponds to increased viewer attention to faces. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 35–40.

Marcus Cheetham, Ivana Pavlovic, Nicola Jordan, Pascal Suter, and Lutz Jancke. 2013. Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Frontiers in psychology* 4 (2013), 108.

Maurizio Corbetta and Gordon L Shulman. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience* 3, 3 (2002), 201.

Andrew Duchowski. 2017. *Eye tracking methodology: Theory and practice* (3rd ed.). Springer-Verlag, Inc., London, UK.

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

John D Eastwood, Daniel Smilek, and Philip M Merikle. 2001. Differential attentional guidance by unattended faces expressing positive and negative emotion. *Attention, Perception, & Psychophysics* 63, 6 (2001), 1004–1013.

Nancy Eisenberg, Tracy L Spinrad, and Natalie D Eggum. 2010. Emotion-related self-regulation and its relation to children's maladjustment. *Annual review of clinical psychology* 6 (2010), 495–525.

Paula M Ellis and Joanna J Bryson. 2005. The significance of textures for affective interfaces. In *International Workshop on Intelligent Virtual Agents*. Springer, 394–404.

James J Gross. 2001. Emotion regulation in adulthood: Timing is everything. *Current directions in psychological science* 10, 6 (2001), 214–219.

Nick Halper, Mara Mellin, Christoph S Herrmann, Volker Linneweber, and Thomas Strothotte. 2003. Psychology and non-photorealistic rendering: The beginning of a beautiful relationship. In *Mensch & Computer 2003*. Springer, 277–286.

Bernhard Jung, Abdul Ahad, and Matthias Weber. 2005. The affective virtual patient: An E-learning tool for social interaction training within the medical field. In *Proceedings of International Conference TESI-Training Education & Education*.

Michael W Kearney. 2017. Cross lagged panel analysis. *The SAGE encyclopedia of communication research methods* (2017), 1–6.

David A Kenny. 1975. Cross-lagged panel correlation: A test for spuriousness. *Psychological bulletin* 82, 6 (1975), 887.

Bart P Knijnenburg and Martijn C Willemsen. 2016. Inferring capabilities of intelligent agents from their external traits. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 28.

Peter J Lang, Robert F Simons, Marie Balaban, and Robert Simons. 2013. *Attention and orienting: Sensory and motivational processes*. Psychology Press.

Richard S Lazarus. 1982. Thoughts on the relations between emotion and cognition. *American psychologist* 37, 9 (1982), 1019.

Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.

Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. 2012. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 91.

Lauri Nummenmaa, Jukka Hyönä, and Manuel G Calvo. 2006. Eye movement assessment of selective attentional capture by emotional pictures. *Emotion* 6, 2 (2006), 257.

Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods* 42, 1 (2010), 188–204.

Arne Öhman. 2005. The role of the amygdala in human fear: automatic detection of threat. *Psychoneuroendocrinology* 30, 10 (2005), 953–958.

Derrick Parkhurst, Klinton Law, and Ernst Niebur. 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision research* 42, 1 (2002), 107–123.

Harold Pashler, James C Johnston, and Eric Ruthruff. 2001. Attention and performance. *Annual review of psychology* 52, 1 (2001), 629–651.

Toni B Pence, Lauren C Dukes, Larry F Hodges, Nancy K Meehan, and Arlene Johnson. 2014. An eye tracking evaluation of a virtual pediatric patient training system for nurses. In *International Conference on Intelligent Virtual Agents*. Springer, 329–338.

Michael I Posner. 1982. Cumulative development of attentional theory. *American Psychologist* 37, 2 (1982), 168.

Helmut Prendinger, Chunling Ma, and Mitsuru Ishizuka. 2006. Eye movements as indices for the utility of life-like interface agents: A pilot study. *Interacting with Computers* 19, 2 (2006), 281–292.

Matthias Rehm and Elisabeth André. 2005. Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. In *International Workshop on Intelligent Virtual Agents*. Springer, 241–252.

Andrew Robb, Andrea Kleinsmith, Andrew Cordar, Casey White, Samsun Lampotang, Adam Wendling, and Lok Benjamin. 2016. Do Variations in Agency Indirectly Affect Behavior with Others? An Analysis of Gaze Behavior. *IEEE Transactions on Visualizations and Computer Graphics* 22, 4 (April 2016), 1336–1345. http://www.ncbi.nlm.nih.gov/pubmed/23428451

David Rogosa. 1980. A critique of cross-lagged correlation. *Psychological Bulletin* 88, 2 (1980), 245.

Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36, 8 (1964), 1627–1639.

Harald T Schupp, Jessica Stockburger, Maurizio Codispoti, Markus Junghöfer, Almut I Weike, and Alfons O Hamm. 2007. Selective visual attention to emotion. *Journal of neuroscience* 27, 5 (2007), 1082–1089.

Rebecca M Todd, William A Cunningham, Adam K Anderson, and Evan Thompson. 2012. Affect-biased attention as emotion regulation. *Trends in cognitive sciences* 16, 7 (2012), 365–372.

Job Van Der Schalk, Agneta Fischer, Bertjan Doosje, Daniël Wigboldus, Skyler Hawk, Mark Rotteveel, and Ursula Hess. 2011. Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion* 11, 2 (2011), 286.

Toni Vanhala, Veikko Surakka, Harri Siirtola, Kari-Jouko Räihä, Benoît Morel, and Laurent Ach. 2010. Virtual proximity and facial expressions of computer agents regulate human emotions and attention. *Computer Animation and Virtual Worlds* 21, 3-4 (2010), 215–224.

Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily, and Tracy Fasolino. 2016. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1326–1335.

Heather A Wadlinger and Derek M Isaacowitz. 2011. Fixing our focus: Training attention to regulate emotion. *Personality and Social Psychology Review* 15, 1 (2011), 75–102.

Yanxiang Wu, Sabarish V Babu, Rowan Armstrong, Jeffrey W Bertrand, Jun Luo, Tania Roy, Shaundra B Daily, Lauren Cairco Dukes, Larry F Hodges, and Tracy Fasolino. 2014. Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics* 20, 4 (2014), 626–635.

Katja Zibrek and Rachel McDonnell. 2014. Does render style affect perception of personality in virtual humans?. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 111–115.