Empirical Evaluation of Virtual Human Conversational and Affective Animations on Visual Attention in Inter-Personal Simulations

Matias Volonte*, Andrew Robb, Andrew T. Duchowski, and Sabarish V. Babu
Clemson University, Human Centered Computing Lab

ABSTRACT

Creating realistic animations of virtual humans remains comparatively complex and expensive. This research explores the degree to which animation fidelity affects users' gaze behavior when interacting in virtual reality training simulations that include virtual humans. Participants were randomly assigned to one of three conditions, wherein the virtual patient either: 1) was not animated; 2) played idle animations; or 3) played idle animations, looked at the participant when speaking, and lip-synced speech and facial gestures when conversing with the participant. Each participant's gaze was recorded in an inter-personal interactive patient surveillance simulation. Results suggest that conversational and passive animations elicited visual attention in a similar manner, as compared to the no animation condition. Results also suggest that when participants face critical situations in inter-personal medical simulations, visual attention towards the virtual human decreases while gaze towards goal directed activities increases.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Animations, Evaluation/methodology; I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality;

1 INTRODUCTION

Continuous advancements in computing technology, display costs, and open and accessible software platforms have dramatically lowered the barrier of entry to creating both immersive and non-immersive virtual reality (VR) environments. As these technologies become more accessible, their use for training simulations has expanded rapidly. Whereas even simple VR training environments once required dedicated labs and staff, consumer hardware and open game engines can now be leveraged by people with little experience to create complex training environments. These VR simulations provide users the possibility of learning procedures or techniques that can be transferable to world scenarios. For example, medical doctors can practice and learn complicated operations in virtual reality simulations before performing surgeries on human patients [1, 21].

In contrast to these VR environments, creating rich and life-like interactions with virtual humans remains complex and expensive, requiring motion capture hardware and/or trained animators. Humans simultaneously leverage multiple channels of communication during social interactions, including speech, turn taking, gaze, emotional expressions, hand gestures, facial expressions, body posture, and proxemics [7]. Each of these behaviors represent active communicative actions that are deliberately or unconsciously coordinated during human to human interactions. In addition to these active communicative actions, humans also engage in a range of non-communicative behaviors at all times, including blinking, postural shifts, head movements, and breathing [10]. While these behaviors are not as strongly linked to the conveyance of information, they are nonetheless important to give the impression of lifelikeness.

As a virtual human's visual or behavioral realism increases, user expectations of their capabilities also rise [7]. Because an agent looks like a human, people expect it to behave as such and will be disturbed by, or misinterpret, discrepancies from human norms. This phenomenon is often referred to as the *Uncanny Valley*, wherein differences between a real and virtual humans grow more and more disconcerting as the distance between the real and virtual shrinks [13]. Examples of this phenomenon can be found in consumer media, perhaps most notably as a comparison between two films: *The Incredibles* and *The Polar Express*, both of which were released by top movie studios in 2004. Despite the former's use of highly stylized, cartoonish human figures, the latter's much more realistic visuals and animation were judged much less appealing though lifelike [9]. Film critic Rob Blackwelder described this eerie experience by saying, "The glitch is in the eyes–there's just no life behind them. In this picture, they're pixel-driven doll orbs without personality or presence" ("Cold Eyes, Warm Heart"). Likewise, in the video game *King Kong*, the heroine's performance was described as unappealing due to her stiff and distorted facial expressions, even though her appearance was attractive [9]. Given the difficulty of generating high quality animation and behaviors that are capable of avoiding the *Uncanny Valley*, many training simulations opt to use either basic animation and behavior, or none at all. Low fidelity simulations are particularly common in medical skills training, wherein the virtual humans are often merely a static image on a website while listing the actions that a student can perform at a given point in time [8].

The current contribution extends two previous investigations that used a medical training system. The first studied the impact of virtual human animation on the emotional responses of participants in a medical virtual reality system for education when looking for signs and symptoms of patient deterioration [23]. In this study, participants were presented with a non-animated or animated character while the participant's psycho-physical Electro Dermal Activity (EDA) was measured, and subjective measures of affect (such as the Differential Emotions and Positive and Negative Affect Survey) were obtained. Findings suggest that participants in the dynamic condition with animations exhibited greater emotional response as compared to participants in the static condition. The second study using this system found that different rendering styles of virtual human agents affected users differently on an emotional level when measured using EDA and affect questionnaires [22]. Participants interacted with one of three anthropomorphic virtual characters namely sketch, cartoon and human-like characters. Results from this study suggest that participants in the human-like condition exhibited the least negative affect corresponding to the deterioration of the virtual patient, as compared to the non-photorealistic cartoon virtual agents.

The current study differs significantly from the previously mentioned evaluations, as we explore how different levels of conversational and affective behavioral animations of a virtual human impact users' *visual attention* in inter-personal simulations. We collect participants' gaze data with an eye tracker during their interactions with the virtual patient simulation designed to teach nurses patient interviewing, surveillance, and monitoring techniques. Participants were tasked with collecting a range of medical information from the patient at four sampling intervals, during which the patient's health

*e-mail: mvolont@clemson.edu

2018 IEEE Conference on Virtual Reality and 3D User Interfaces
18-22 March, Reutlingen, Germany
978-1-5386-3365-6/18/$31.00 ©2018 IEEE

deteriorates. Participants' were randomly assigned to one of three conditions: 1) the *No Animation* [NA] condition, in which the virtual patient was not animated at all, 2) the *Non Communicative* [NCA] condition, in which the virtual patient expressed basic passive animations, including idle motion, breathing, and blinking, and 3) the *Communicative* [CA] condition, in which the virtual patient also gazed at participants, displayed communicative lip-synced speech, facial expressions and gestures, in addition to the passive animations from the previous NCA condition.

We found that, while participants' visual attention shifted over time, the types of animations displayed had little effect on visual attention. Animations had no observed effect when participants were engaged in goal-directed activities (e.g. using tools to record a patient's vital signs), but they did have a significant effect when participants conversed with the patient: participants spent more time looking at the virtual human when engaging in conversation if the virtual human was more animated. Across all conditions, participants' gaze gradually shifted away from the virtual human and towards the interface elements. We hypothesize that this gradual shift may be attributed to increasing emotional intensity as the patient's health deteriorates rapidly. Animation quality does have an impact on visual attention during conversational moments, and it is plausible that animation quality will gain importance in simulations that are more strongly focused on conversational skills.

2 Related Work

Attention is the means by which information is filtered and selected for processing via stimulus-driven and voluntary actions [16]. On the one hand, stimulus-driven processes are strongly influenced by the properties of the stimulus and are often involuntary and automatic. On the other hand, voluntary processes may be seen as mostly goal-directed. There is evidence that social and emotional cues from virtual agents can influence human attention and behavior [5, 22, 23]. Although a major part of these responses are involuntary and automatic, we may also become aware of them by voluntary reflection of our reactions.

Research has explored the relationship between emotions and animation and their effect on emotional contagion in VR. It has been suggested that an animated virtual agent can have a positive impact in mediating the *Uncanny Valley effect* [23]. Moreover, research suggests that virtual agent animation is important in eliciting authentic emotional responses in participants in simulated human-virtual human interaction.

Gaze serves many functions in human-human communication. Gaze can be used to obtain information and feedback about the reactions of other people, provide attentional cues, help to regulate the flow of conversation, communicates emotion and relationships [2–4], and provides turn taking cues such as looking away at the beginning of an utterance and looking back at the end of it [4]. Also, people look nearly twice as much while listening as while speaking, they look longer, and their glances away are shorter.

Engaging in mutual gaze is considered important for successful social conversation. Interactants that exhibit high amounts of mutual gaze are perceived as competent, attentive, and powerful [3]. Research also suggests that successful conversation includes gaze aversion.

The multifaceted nature of gaze in human-human communication suggests that gaze with virtual humans will also be of great importance. In particular, virtual humans that engage in mutual gaze may seem more alive and capable than agents who do not engage in mutual gaze behaviors. Similarly, context-aware displays of gaze aversion may also be important.

Relatively little research has explored eye gaze behavior with virtual agents, relative to the importance of gaze in human-human communication. At least two studies have explored the relationship between gaze and animation in virtual humans. Martinez [11] used an eye tracker to assess how the presence of animation influenced gaze fixations, specifically comparing static images, alternating images without animation, and a fully animated animation of head turning. The presence of full animation drew an observer's attention the fastest. Prendinger [17] explored eye gaze in an online setting with a virtual agent against a standard webpage. Results showed that visual attention was drawn to the virtual agent, particularly when the agent made deictic gestures, and that users directed their attention to the objects the agent gestured towards. Additionally, visual attention was particularly drawn to the agents' face. Other research has shown that eye gaze with virtual humans mimics gaze with real humans, though there are also key differences [18, 19]. People look at virtual agents when they speak to them, and when the agents speak. However, people spend more time looking at virtual agents than they do at other humans during these moments, possibly due to the limited non-verbal cues expressed by virtual humans.

Of particular relevance to this paper is work performed by Pence et al. [15], who explored gaze with virtual agents in a pediatric interviewing application using a tablet platform. They studied participants' visual attention on visual interface layout configurations in a desktop simulation that taught pediatric interviewing techniques to nursing students. Surprisingly, these results showed that little visual attention was directed towards the virtual patients in this training context, even when animations were present. The majority of time was spent gazing at interface elements that were used to perform the interview task with the virtual patient, rather that at the patient. Though the participants reported enjoying interacting with the virtual patient, this was not reflected in their gaze behavior.

3 Methods

3.1 Materials and Apparatus

This experiment was conducting using the Rapid Response Training System (RRTS), a system originally developed to train nurses and nurse practitioners in recognizing the signs, symptoms, and behaviors of patients who suffered from rapid health deterioration. Trainees interacted with the patient four times, corresponding to four different assessment periods evenly spaced throughout a nurse's shift (we refer to these four sampling times as time-steps). The patient's physical and mental health deteriorated rapidly as the trainee progressed from one time-step to the next.

The RRTS teaches trainees to follow procedural interviewing and diagnostic steps nurses need to perform every time they visit their patient during medical rounds. These tasks include: 1) checking the patient's vital signs through the use of multiple medical devices (stethoscope, vitals monitor, input and output intake, O2 meter); 2) observing the patient for visual signs of deterioration; 3) checking the patient's cognitive and mental reflexes and his health by asking specific types of questions; and 4) entering the collected information into an Electronic Health Record Form (EHR).

Trainees interact with the RRTS using dual displays (see Figure 1). The first, a 65″ TV, displays the patient life-sized (as if sitting across the table from them) and presents tools used to assess his health. The second, a 21″ touchscreen monitor displays a simulated EHR that trainees can use to record the patient's health information (this simulated EHR was based on the system used at a regional hospital located near where this system was developed). Participants interact with the system using a keyboard and mouse. The 65″ display shows a visualization of a typical patient's room, along with interface elements used to complete the necessary surveillance and monitoring tasks. Trainees can ask the patient questions by selecting options from a list. This list is divided into categories that assess seven different aspects of the patient's health: general health, information about his situation, respiratory problems, cardiac problems, gastrointestinal problems, genitourinary problems, and his physical activity. The patient verbally responds after a nurse selects a question.

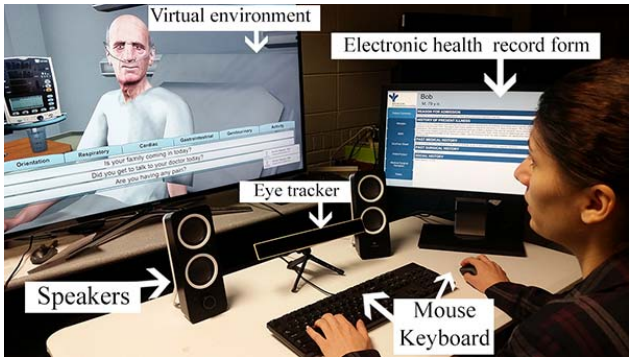

Figure 1: Screenshot shows a participant interacting with the virtual patient in the RRTS, and recording his vitals in the EHR screen.

3.2 Research Question and Expected Outcomes

We examined how the virtual patient's conversational and affective animations impacted visual attention. We asked the following:

1. Overall, how does visual attention differ between the virtual human, the environment, tools, and the simulation's user interface during simulated medical surveillance interactions?
2. To what extent does visual attention towards the virtual human differ from one simulation time to another during which the virtual patient's affective and behavioral animation changes?
3. How is visual attention towards a virtual human affected by the conversational and affective behavioral animations of the virtual entity?
4. To what extent do users visually attend to the virtual human when engaged in conversational tasks, and does this differ as a function of animation fidelity?

We hypothesized that more complex behaviors and animations would cause participants to increase their visual attention towards the virtual patient. We hypothesized that irrespective of the animation fidelity of the virtual entity, visual attention towards the virtual human would increase from one simulation time step to the next. Finally, in a manner similar to human-human social conversation, we hypothesized that during simulated conversation visual attention would be drawn towards the virtual human.

3.3 Experiment Design

3.3.1 Conditions

To answer our research questions, we designed a between-subjects experiment with three conditions, wherein the patient displayed different behaviors and animations in each condition. The differences between these three conditions are summarized in Table 1.

In the *Communicative Animations* (CA) condition, the patient continued to display all of the animations the RRTS is capable of simulating. These behaviors include head motion, body movements, dynamic eye gaze, mutual gaze with the participant when speaking, random idle motions involving interactions with the environment (e.g. hand and head scratching, looking at TV), lip-synced speech, conversational facial expressions, and spoken audio responses. All of these motions were carefully scripted based on actual data collected from real actors at a local regional hospital.

The *Non-Communicative animations* (NCA) condition removed all animations that were related to communications between the participant and the patient. In particular, mutual eye gaze, lip synced speech, and conversational facial expressions were removed, while

Table 1: The behaviors and animations expressed by the patient in each of the three conditions.

	CA	NCA	NA
Audio	Yes	Yes	Yes
Idle animations	Yes	Yes	No
Environmental behaviors	Yes	Yes	No
Non-mutual gaze behavior	Yes	Yes	No
Mutual gaze behavior	Yes	No	No
Lip syncing	Yes	No	No

all other animations remained as is. This gave the impression of a human-like patient, but one who was unaware of the participant and largely did not respond to conversational actions performed by the participant, except for eliciting audio responses to questions asked by the participant (without lip-synced speech and facial expressions). Eliminating mutual gaze was expected to reduce the social engagement between the virtual agent and the participant, as engaging in mutual gaze is considered important for a successful social conversation [3].

The final condition, *No Animation* (NA), involved no animated behaviors at all. The patient was presented in a fixed pose that changed between time-steps, while showing the patients' increasing discomfort and deterioration. These static poses were taken from the animations on display by the virtual patient in the other conditions. The virtual patient continued to express the same audio responses found in the other two conditions, including coughing, breathing, discomforts or pain sounds. Speech audio was also played when the user asked questions, but the virtual patient did not engage by gaze or perform any lip synced speech or conversational gestures.

3.3.2 Participants

We recruited 12 male and 21 female participants between the ages of 18 and 50 from the Clemson University campus. This experiment included 33 participants balanced with 11 per condition (CA=11, NCA=11, NA=11).

3.3.3 Methodology

Figure 2 visualizes the flow of the study. Upon arrival, participants were given a brief description of the virtual reality system and its use as a tool to train nurses to recognize rapid deterioration in their patients. Participants were then told what they would be asked to do and were asked to sign a consent form. If participants gave their consent, they then completed a survey regarding their demographics and current disposition. Once the surveys were completed, the training session started. In this phase of the study, the study proctor thoroughly explained how to use the RRTS system to the participant. After the explanation, participants practiced with the system until they felt fully acclimated to its use.

After the training phase ended, the study proctor calibrated the eye tracker using a nine-point calibration routine. This procedure was repeated until it was successful, usually within three attempts. Upon successful gaze adjustment, participants were instructed to remain in the same relaxed, adopted posture for the remainder of the experiment, so as to ensure that the eye tracker continued to function properly throughout the entire study.

Next, the participant was introduced to the first time-step and was asked to interact with the virtual patient by asking as many questions as they felt necessary, to use as many virtual instruments as necessary to medically assess the virtual patient, Bob, his condition, and record the collected data in the EHR system. At the end of each time-step, the participant filled out a survey with questions regarding the vital signs of the virtual patient in order to check for the participant's learning outcomes and to encourage the user to be engaged in the simulation. Finally, at the end the fourth time-step, the participant

Figure 2: The time-line of the experiment from left to right. The surveys administered after each time-step assessed the participant's knowledge of the patient's condition. The final evaluation included additional surveys assessing general impressions about the patient.

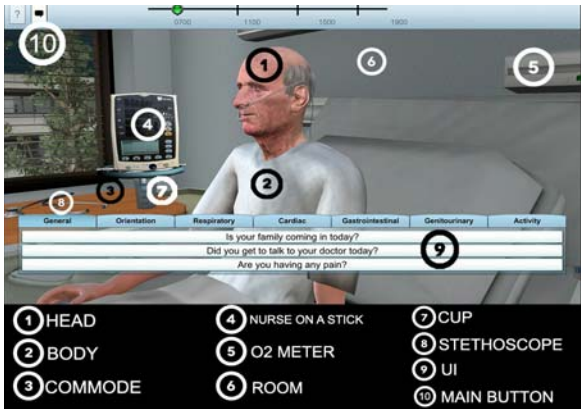

Figure 3: Regions of interest checked for gaze intersection.

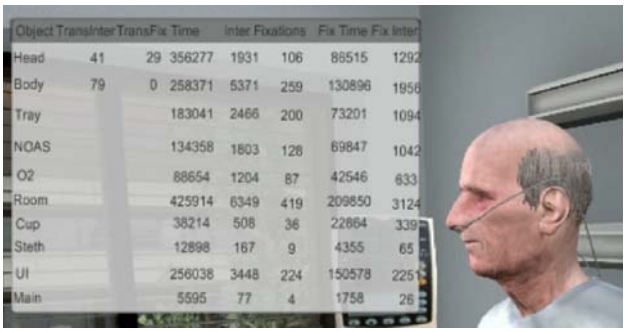

Figure 4: Post processing tool replaying a participant's gaze and displaying calculations of gaze intersection with regions of interest.

was debriefed and thanked for participating in the study. Participants joined the study on a volunteer basis and no compensation was provided for participation in this study.

3.4 Gaze Measures

A Gazepoint eye tracking system was used in this experiment for gathering participants' gaze data. This device is a tabletop eye tracker that is placed in front of the display of interest (see Figure 1). The motion of both eyes was sampled at 60 Hz. Participants were instructed to adopt a comfortable natural posture before the Gazepoint tracking system was repositioned to accurately track the users' gaze while using the simulation, and the eye tracking system was subsequently calibrated.

Human eye motion can be modeled using three behaviors: fixations, smooth pursuits, and saccades. Fixations occur when gaze remains focused on a specific point of interest for longer than a fixed value. Smooth pursuits occur when the eye smoothly moves to follow a moving object or point of interest. Saccades are rapid, discontinuous motions where the eye rapidly jumps from one fixation point to another. When computing visual attention, the goal is typically to detect where the eye is gazing and when the motion signal changes abruptly, indicating the end of a fixation, the onset of a saccade, and the beginning of a new fixation [6]. Smooth pursuits may also be of interest if objects in the scene are moving.

In this research, fixations and saccades were detected using Nyström and Holmqvist's algorithm for velocity-based detection (I-VT event detection) [14]. A high-level description of this algorithm is as follows: assuming that the eye movement signal is recorded at a uniform sampling rate, successive samples are subtracted to estimate eye movement velocity. Fixations are either implicitly detected as the portion of the signal between saccades, or the portion of the signal where the velocity falls below a threshold. Because the data produced by the gazepoint tracker is noisy, the Savitzky-Golay filter [20] was applied to the data before analysis, along with a 2nd order low-pass Butterworth filter to smooth the raw gaze data with sampling and cutoff frequencies of 60 and 1.65 Hz.

Once fixations were extracted from the eye gaze data, we then needed to associate these fixations with specific points on the RRTS display. To accomplish this, the participant's eye gaze was recorded, along with mouse input and all data concerning actions performed in the RRTS during the experiment. This allowed the interactions to be replayed later for off-line analysis. In this post-experiment step, participants' eye gaze was replayed so as to determine which objects were being observed by the participant at any given point in time. This was accomplished by averaging the left and right eye screen space coordinates and casting a ray into the virtual environment from this position. This ray was used to check for collisions with the 3D geometry in the scene. When a collision was detected, this object was registered as the object the participant was looking at that particular moment in time. All of the objects in the scene were grouped into 10 categories, shown in Figure 3. These categories were later collapsed to five for analysis: head, body, UI, tools, and environment. A visualization tool was also developed that visualized participants' eye gaze, showing the screen space intersection, mouse motion, and highlighting the object being visually attended to. This visualization tool was used to further explore the users' visual attention during the experiment (see Figure 4).

4 Results

4.1 Comparing Visual Attention by Conditions of Animation Fidelity across Time-steps on Objects

Visual attention data such as proportion of time gazed using the Butterworth filter smoothing algorithm (% of time gaze drawn towards), fixations per minute (via Savitzky-Golay algorithm), and proportion of time fixated on (% of time fixations elicited towards) were separately treated with a $3 \times 4 \times 4$ mixed model analysis of variance (ANOVA). The between-subjects factors were animation fidelity conditions at 3 levels (Communicative Animations (CA),

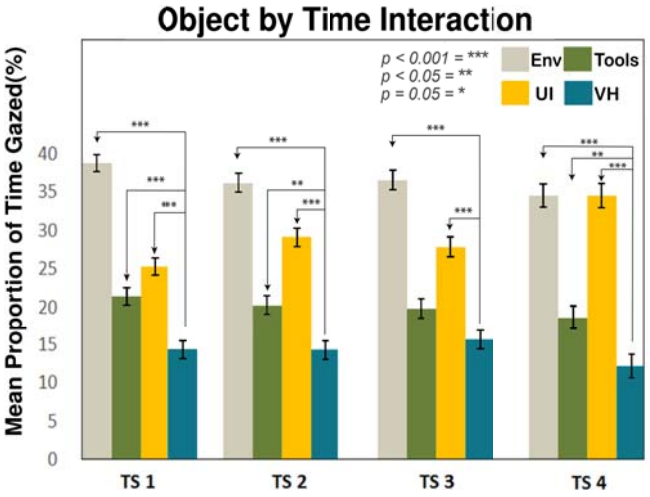

Figure 5: Proportion of time gazed at objects over time-steps (Mean and Standard Error). The arrows show significant Tukey's HSD pairwise comparisons between the virtual human and other objects within a time-step.

Non-Communicative Animations (NCA), or No Animations (NA)), and objects at 4 levels (virtual human, tools (cup, nurse-on-a-stick NOAS, commode, O2, stethoscope), environment, or the conversational UI). The within-subjects repeated measures factors were the distinct sampling times of interaction with the virtual patient (at 4 levels), during which his medical condition gradually worsened, which is also considered as the emotionally distressing dimension of the progression of time.

4.1.1 Comparison of Proportion of Time Gazed

The proportion of time gazed is a normalized measure of the percentage of time participants visually attended to different objects including the virtual human at every simulation time-step. ANOVA revealed a significant main effect of objects F(3, 120) = 107.65, p<0.001, η2= 0.73 and time by object interaction F(8.274, 330.95) = 7.50, p <0.001, η2= 0.16 in the mean proportion of time the participant visually attended to various objects in the simulation. Post-hoc Bonferroni pairwise comparisons showed significant results in mean proportion of time visually attended to various objects between simulation time-steps is given in Table 2. Overall mean proportion of time visually attended to objects and post-hoc Tukey's HSD comparisons of differences in visual attention to the virtual human as compared to other objects is illustrated in Figure 5.

4.1.2 Comparison of Fixations/Minute Data

The total number of fixations on each object that were detected were converted to fixations per minute by dividing the fixations counts by the total time fixated on all objects in the scene (virtual human, environment, UI and virtual instruments) at that time-step for each participant. ANOVA on the mean number of fixations per minute revealed a significant main effect of objects F(3, 120) = 128.465, p <0.001, η2 = 0.76, and time by object interaction F(6, 360) = 5.51, p <0.001, η2 = 0.12. Post-hoc Bonferroni pairwise comparisons showed significant results on mean fixation per minute to various objects between simulation time-steps, shown in Table 3. Overall mean fixations per minute on objects and post-hoc Tukey's HSD comparisons of differences in visual attention in fixations per minute to the virtual human as compared to other objects is illustrated in Figure 6. There were no significant differences from one simulation time-step to another during which Bob medically deteriorated in the mean fixations per minute on the virtual instruments such as the nurse-on-a-stick (NOAS), O2 meter, stethoscope, etc.

Table 2: Table showing mean proportion of time gaze and standard deviations to objects by time-step and significant Bonferroni's pairwise comparison effects between time-steps on different objects.

Object	Time-step	Mean% (SD)
VH		
	2	14.3% (5.3)
	3	15.7% (6.8)
	4	12.2% (6)
Bonferroni's comparisons		P-Value
	TS4 <TS2	0.021
	TS4 <TS3	<0.001
UI	1	25.3% (9.2)
	3	27% (8.8)
	4	34.5% (11.9)
Bonferroni's comparisons		P-Value
	TS4 >TS1	0.004
	TS4 >TS3	0.024
Tools	1	21.4% (4.8)
	2	20.2% (4.5)
	4	18.6% (5.6)
Bonferroni's comparisons		P-Value
	TS4 <TS1	0.015
	TS4 <TS2	0.048
Env	1	38.8% (6.1)
	4	34.5% (8.8)
Bonferroni's comparisons		P-Value
	TS4 <TS1	0.006

4.1.3 Comparison of Gaze Transitions

Overall, we also measured the number of gaze transitions per minute towards the virtual human at any time during the simulation time-step derived from the Butterworth-filtered smooth gaze data across conditions. ANOVA of the overall gaze transitions per minute data revealed a significant main effect of time F(2.28, 992.69)=14.32, p <0.001, η2= 0.32. Mean gaze transitions per minute towards the virtual human were significantly higher in time-step 4 (M=12.04,

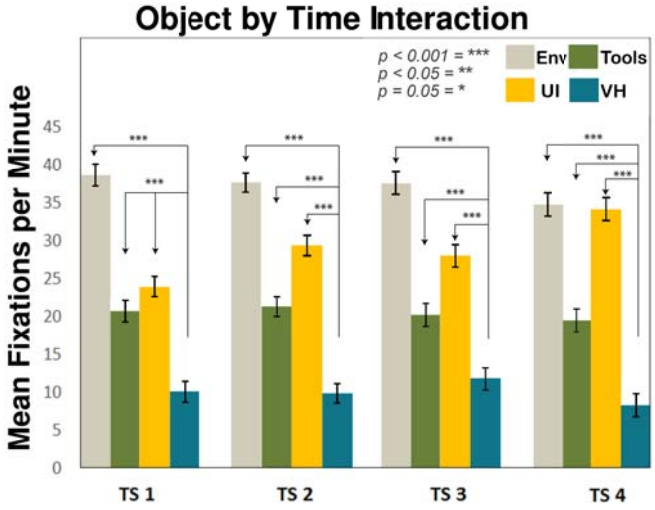

Figure 6: Overall fixations per minute, simulation time-steps by object interaction graph (Mean and SEM). Arrows show Tukey's HSD pairwise significant differences of fixations per minute between the virtual human and other objects within a time-step.

Table 3: Table showing pairwise significant differences in mean fixations/minute on objects between simulation time-steps.

Object	Time-step	Mean/Min (SD)
VH		
	1	10.5 (3.2)
	2	9.85 (3.89)
	3	11.76 (5.73)
	4	8.29 (4.8)
Bonferroni's comparisons		P-Value
	TS3 >TS2	0.036
	TS3 >TS4	<0.001
	TS4 <TS1	0.031
	TS4 <TS2	0.015
UI	1	24.04 (9.43)
	2	29.43 (10.06)
	3	28.07 (10.33)
	4	34.16 (12.86)
Bonferroni's comparisons		P-Value
	TS4 >TS1	0.001
	TS4 >TS2	0.04
	TS4 >TS3	0.029
	TS1 <TS2	0.008
Env	1	38.7 (9.43)
	4	34.8 (8.51)
Bonferroni's comparisons	TS1 <TS4	0.019

SD=6.19) as the virtual patient Bobs health was greatly deteriorating as compared to time-step 1 (M=6.83, SD=2.9) p <0.001, time-step 2(M=8.35, SD=4.0) p=0.001, and time-step 3(M=9.2, SD=3.8) p=0.05. It is interesting to note that gaze transitions increase significantly with the patients deteriorating state in this latter time-step, while conversely, the proportion of gaze, and fixations per minute appear to be decreasing significantly.

4.2 Comparing Gaze Driven towards the Virtual Human during Conversation

In order to specifically compare visual attention to the virtual human between animation and emotional reaction conditions, we compared total time spent at each time-step, the proportion of time visually attending the virtual human, and the proportion of time fixating on the virtual human in a 3×4 mixed model ANOVA, specifically when the conversational GUI in the simulation was enabled. Similar to the overall analysis, the between-subjects factors were the animation fidelity at three levels (communicative animation (CA), non-communicative animations (NCA), and no animation static condition (NA)), and the within-subjects repeated measures variable was the distinct time-steps of interaction with the virtual patient Bob at four levels, during which the patient's medical condition deteriorates, which is also considered as the emotionally distressing dimension of the progression of time. Dependent variables used in this analysis were similar to the previous section as explained below.

4.2.1 Comparison of Proportion of Time Visually Attended

The proportion of time visually attended to the virtual human during conversation was calculated as the amount of time smoothed gaze was elicited towards the virtual human during conversation, divided by the total conversation time in that time-step. ANOVA revealed a significant main effect of simulation time F(3, 189) = 7.14, p <0.001, $\eta 2 = 0.10$, and time by condition interaction F(6, 189) = 2.27, $p = 0.039$, $\eta 2 = 0.067$ on the proportion of time spent visually attending to the virtual human during a conversational event (see Figure 7). Post-hoc significant Bonferroni comparisons of mean proportion of time visually attended to the virtual human in different animation conditions between time-steps are shown in Table 4. Post-hoc Tukeys HSD comparisons of mean proportion of time visually attended to the virtual human between animation conditions in time-steps 1, 2, 3, and 4 did not reveal any significant differences.

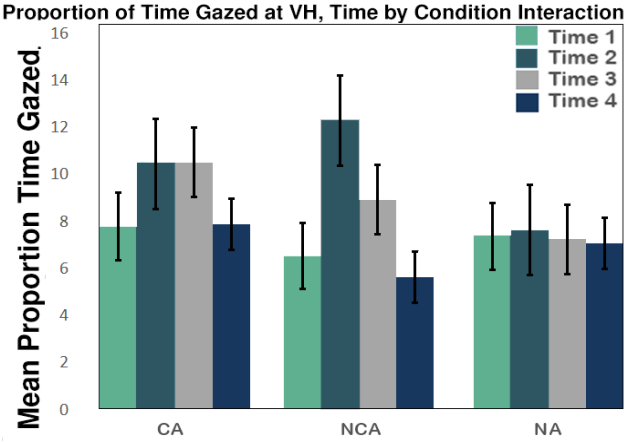

Figure 7: Proportion of time visually attended to the virtual human during conversation, time by condition interaction (Mean and SEM).

4.2.2 Comparison of Proportion of Time Fixated

The proportion of time fixated towards the virtual human during conversation was calculated as the total time fixation was elicited towards the virtual human during conversation, divided by the total amount of time participants fixated at objects in that time-step. The ANOVA analysis revealed a significant main effect of simulation time F(3, 189) =7.08, p <0.001, $\eta 2 = 0.10$, and a time by condition interaction F(6, 189) = 2.23, $p = 0.042$, $\eta 2 = 0.067$ on the proportion of time fixated at the virtual human during a conversational event (See Figure 8). Post-hoc significant Bonferroni comparisons of mean proportion of time fixated on the virtual human in different animation conditions between time-steps are shown in Table 5. Post-hoc Tukeys HSD comparisons of mean proportion of time visually fixated on the virtual human between animation conditions in time-steps 1, 2, 3, and 4 did not reveal any significant differences.

Table 4: Comparison of Proportion of Time Visually Attended to Virtual Human during Conversation

Condition	Object	Time-step	Mean% (SD)
CA	VH		
		1	7.77% (8.2)
		2	10.47% (10.9)
Bonferroni's Comparisons			P-Value
		TS1 <TS2	0.035
NCA	VH		
		1	6.5% (6.0)
		2	12.3% (9.8)
		3	8.92% (6.12)
		4	5.63% (5.5)
			P-Value
Bonferroni's Comparisions		TS2 >TS1	0.006
		TS2 >TS3	0.05
		TS2 >TS4	0.003
		TS3 >TS4	0.009

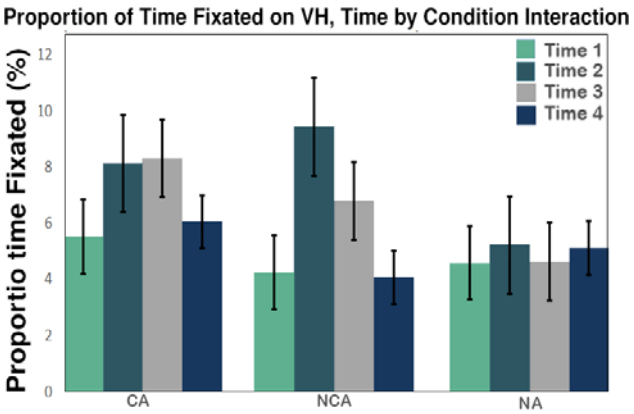

Figure 8: Proportion of time fixated on the virtual human during conversation, time by condition interaction (Mean and SEM).

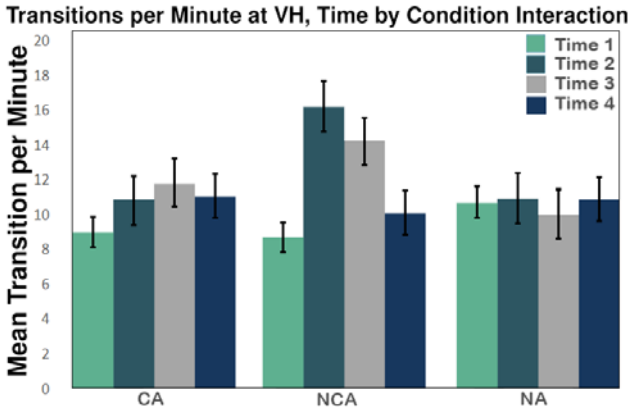

Figure 9: Gaze transitions per minute on the virtual human during conversation, time by condition interaction (Mean and SEM).

4.2.3 Comparison of Transitions Per Minute

The number of gaze transitions per minute was calculated as the total number of smooth gaze transitions from the conversational UI object to the virtual human during conversation per minute at that time-step. ANOVA revealed a significant main effect of simulation time $F(3, 189) = 6.75$, $p < 0.001$, $\eta 2 = 0.10$, and time by condition interaction $F(6, 189) = 4.29$, $p < 0.001$, $\eta 2 = 0.12$ on the mean number of fixations towards the virtual human during a conversational event (see Table 6 and Figure 9 for comparisons).

5 Discussion

In response to the question of how does users' visual attention differ between the virtual human, environment, tools and the simulation user interface, we found that users' visual attention differed significantly between the four elements of interest in the medical simulation. Overall, the proportion of time visually attended and fixations per minute data suggests that visual attention towards the virtual human decreases in the last time-step. Attention towards the virtual environment gradually decreases from the first to the last time-step possibly due to acclimation since, over time, users may know exactly where to focus their attention in order to accomplish the task. We also found that, overall, levels of visual attention towards the tools remains the same throughout the simulation, while visual attention towards the conversational UI significantly and drastically increases from one time-step to the next as the virtual patient deteriorates.

In response to the question to what extend does visual attention towards the virtual human differ over time as the virtual patient's behaviors changes due medical deterioration, we found that, overall, medical deterioration had a significant impact on the users' visual attention. Visual attention towards the virtual human was higher in the middle time-steps and decreased significantly at the last time step, when the virtual patient medically deteriorates the most. However, the proportion of time users spent visually attending to the UI increased significantly in time-step 4, as compared to the initial time-steps of the simulation experience. This result provides insights regarding the users' gaze behavior in that as the simulation time progressed, users became more focused on gathering data by asking questions of the virtual patient. Interestingly, gaze transition towards the virtual human overall was significantly higher in the last time step, when the virtual patient was critical, as compared to the previous time-steps. This provides evidence that as the participant experiences the medical simulation, they become more efficient in adapting their visual attention to task- or goal-oriented aspects such as the conversational UI, instead of focusing on the virtual patient for a long period of time. These results suggest that in goal-oriented inter-personal simulations, such as medical trainers, when presented with critical situations such as a failure-to-rescue scenario, participants tend focus on accomplishing tasks at the cost of minimizing social face-to-face interactions. This type of tunnel vision has been

Table 5: Comparison of Proportion of Time Fixated on the Virtual Human (%) during Conversation

Condition	Object	Time-step	Mean% (SD)
CA	VH		
		1	5.5% (7.45)
		2	8.1% (9.98)
Bonferroni's			P-Value
Comparissons		TS1 <TS2	0.030
NCA	VH		
		1	4.2% (5.7)
		2	9.4% (8.5)
		3	6.74% (5.5)
		4	4.1% (4.6)
Bonferroni's			P-Value
Comparissons		TS2 >TS1	0.003
		TS2 >TS4	0.005
		TS3 >TS1	0.011
		TS3 >TS4	0.009

Table 6: Comparison of Transitions Per Minute during Conversation

Condition	Object	Time-step	Mean/Min (SD)
CA	VH		
		1	9 (4)
		3	10.8 (8.3)
Bonferroni's			P-Value
Comparisons		TS3 >TS1	0.050
Condition	Object	Time-step	Mean/Min (SD)
NCA	VH		
		1	8.7 (4.1)
		2	16.2 (9.6)
		3	14.2 (8.3)
		4	10.1 (6.3)
Bonferroni's			P-Value
Comparisons		TS2 >TS1	<0.001
		TS2 >TS4	0.003
		TS3 >TS1	0.002
		TS3 >TS4	0.037

noticed in medical practitioners when handling critical situations in a hospital setting, and could potentially explain the results found in this study [12]. Similar results were also witnessed in desktop virtual patient simulations in procedural interview training tasks [15].

With regards to the question of how is visual attention directed towards a virtual human affected by the conversational and affective behavioral animations of the virtual entity, we found that the conversational and behavioral animations played a significant role in allocation of the participants' visual attention. When examining both the overall as well as the conversational event results, we noticed that participants in the CA and NCA conditions elicited a larger proportion of time and fixations towards the virtual human (especially in the middle time-steps), than the no animation condition. The non-verbal behaviors of the virtual patient elicited higher visual attention towards the virtual human, as compared to the no animation condition in which visual attention towards the virtual patient was low and did not differ one time-step to another.

With regards to what extent users visually attend to the virtual human when engaged in conversational tasks and the impact of animation fidelity on the same, we found that the presence of conversational animations had a significant impact on visual attention to the virtual human during conversation. Our results do not support the hypothesis that conversational animations will elicit increased visual attention as compared to non-conversational animations as the visual attention towards the NCA condition was the highest among all three conditions. However, animations do elicit visual attention as compared to no-animations. Interestingly, the number of gaze transitions towards the virtual human during conversation was significantly higher in the NCA condition in the middle time-steps as compared to the CA and NA conditions, which could be due to the lack of conversational verbal and non-verbal behaviors causing confusion in participants in assessing the patient's state between the presence of passive life-like animations and the absence of active conversational behaviors.

6 Conclusion and Future Work

In an empirical evaluation, we examined the impact of conversational and affective animations of a virtual human on participants' visual attention using eye tracking in a medical inter-personal skills simulation. We found that as the virtual patient's emotionally distressing animations pertaining to medical deterioration intensified, participants tended to shift visual attention from the virtual human towards goal-oriented tasks of surveying and monitoring the vital signs of the patient. Our results suggest that in a manner similar to real world social interactions, when faced with a critical situation, participants may intently switch visual attention to goal-oriented tasks, rather than engaging in social face-to-face gaze behaviors. Evidence also suggests that conversational and non-conversational animations elicit visual attention in a similar manner as compared to no animations which elicited the least visual attention to the virtual human overall.

Future work will focus on evaluating the effects of natural non-goal oriented dialogue versus task-oriented social interactions on visual attention to virtual humans. Future work will also focus on validating our findings regarding gaze behaviors to virtual humans in inter-personal simulations, against gaze behaviors directed towards real humans in real-world goal-oriented or task-oriented encounters.

References

[1] R. Aggarwal, J. Ward, I. Balasundaram, P. Sains, T. Athanasiou, and A. Darzi. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Annals of surgery*, 246(5):771–779, 2007.

[2] M. Argyle. *Social interaction*, vol. 103. Transaction Publishers, 1973.

[3] M. Argyle and M. Cook. Gaze and mutual gaze. 1976.

[4] M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, pp. 289–304, 1965.

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

[6] A. Duchowski. *Eye tracking methodology: Theory and practice*. 3rd ed. London, UK: Springer-Verlag, Inc., 2017.

[7] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. Badler. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent systems*, 17(4):54–63, 2002.

[8] G. Huang, R. Reynolds, and C. Candler. Virtual patient simulation at us and canadian medical schools. *Academic Medicine*, 82(5):446–451, 2007.

[9] F. Kaba. Hyper realistic characters and the existence of the uncanny valley in animation films. *International Review of Social Sciences and Humanities*, 4(3):188–195, 2013.

[10] N. Magnenat-Thalmann and D. Thalmann. *Handbook of virtual humans*. John Wiley & Sons, 2005.

[11] S. Martinez, R. J. Sloan, A. Szymkowiak, and K. C. Scott-Brown. Using virtual agents to cue observer attention. In *CONTENT 2010: The Second International Conference on Creative Content Technologies*, pp. 7–12, 2010.

[12] C. McLaughlin. An exploration of psychiatric nurses' and patients' opinions regarding in-patient care for suicidal patients. *Journal of Advanced Nursing*, 29(5):1042–1051, 1999.

[13] M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

[14] M. Nyström and K. Holmqvist. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, 42(1):188–204, 2010.

[15] T. B. Pence, L. C. Dukes, L. F. Hodges, N. K. Meehan, and A. Johnson. An eye tracking evaluation of a virtual pediatric patient training system for nurses. In *International Conference on Intelligent Virtual Agents*, pp. 329–338. Springer, 2014.

[16] M. I. Posner. Cumulative development of attentional theory. *American Psychologist*, 37(2):168, 1982.

[17] H. Prendinger, C. Ma, and M. Ishizuka. Eye movements as indices for the utility of life-like interface agents: A pilot study. *Interacting with Computers*, 19(2):281–292, 2006.

[18] M. Rehm and E. André. Where do they look? gaze behaviors of multiple users interacting with an embodied conversational agent. In *International Workshop on Intelligent Virtual Agents*, pp. 241–252. Springer, 2005.

[19] A. Robb, A. Kleinsmith, A. Cordar, C. White, S. Lampotang, A. Wendling, and L. Benjamin. Do Variations in Agency Indirectly Affect Behavior with Others? An Analysis of Gaze Behavior. *IEEE Transactions on Visualizations and Computer Graphics*, 22(4):1336–1345, Apr. 2016.

[20] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[21] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. Obrien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458–464, 2002.

[22] M. Volante, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics*, 22(4):1326–1335, 2016.

[23] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino. Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics*, 20(4):626–635, 2014.