The Effect of An Animated Virtual Character on Mobile Chat Interactions

Sin-Hwa Kang, Andrew W. Feng, Anton Leuski, Dan Casas, and Ari Shapiro
USC Institute for Creative Technologies
Playa Vista, USA
kang,feng,leuski,casas,shapiro@ict.usc.edu

ABSTRACT

This study explores presentation techniques for a 3D animated chat-based virtual human that communicates engagingly with users. Interactions with the virtual human occur via a smartphone outside of the lab in natural settings. Our work compares the responses of users who interact with no image or a static image of a virtual character as opposed to the animated visage of a virtual human capable of displaying appropriate nonverbal behavior. We further investigate users' responses to the animated character's gaze aversion which displayed the character's act of looking away from users and was presented as a listening behavior. The findings of our study demonstrate that people tend to engage in conversation more by talking for a longer amount of time when they interact with a 3D animated virtual human that averts its gaze, compared to an animated virtual human that does not avert its gaze, a static image of a virtual character, or an audio-only interface.

ACM Classification Keywords

I.2.12 Artificial Intelligence: Distributed Artificial Intelligence—Intelligent agents; J.4 Computer Applications: Social and Behavioral Sciences—Psychology.

Author Keywords

virtual humans, agents, smartphones, chat applications,
nonverbal behavior, self-disclosure, rapport, reciprocity,
facial expressions, saccade, gaze aversion, speech
recognition

INTRODUCTION

3D content, including virtual worlds and virtual characters are increasingly being utilized in video games, simulations, and film and television. Virtual humans, in particular, have emerged from this digital space to capture the attention of many social scientists, psychologists, and computer scientists as an important area of research. This is because humans often perceive and are impacted by virtual humans in the same way that they perceive and are impacted by real humans [7]. Thus virtual humans prove to be a valuable source with which to elicit emotions, manipulate and simulate real humans. Mobile platforms such as smartphones and tablets are a convenient, pervasive technology capable of running the software components necessary for displaying a convincing and interactive virtual human. Anyone with a smartphone can now interact with a virtual human. It has been shown that people have a unique, strong emotional connection to their mobile phones compared to other types of computers, such as desktop computers [14]. The mobile nature of a smartphone additionally facilitates long-term interactions with virtual humans in ways not previously possible with desktop or other fixed installations.

In this study, we are interested in exploring whether people would talk with 3D animated virtual humans using a smartphone for a longer amount of time as a sign of feeling rapport [7], compared to non-animated or audio-only characters in everyday life. Based on previous studies [2, 10, 19], users prefer animated characters in emotionally engaged interactions when the characters were displayed on mobile devices, yet in a lab setting. We aimed to reach a broad range of users outside of the lab in natural settings to investigate the potential of our virtual human on smartphones to facilitate casual, yet emotionally engaging conversation.

We hypothesized that users would talk more to a virtual human that presents nonverbal behavior in an appropriate manner compared to a virtual human that does not have an image or animated behavior. We further examined whether presenting gaze aversion or constant mutual gaze on the small screen of smartphones would affect a user's responses to virtual humans when the gaze movements were used while the virtual humans were listening. We examined our hypothesis and question using a 3D animated and chat-based virtual human which presented emotionally expressive nonverbal behaviors such as facial expressions, head gestures, gaze, and other upper body movements. To explore the question of optimal communicative medium, we distributed our virtual human application to users via an app store for Android-powered phones (i.e. Google Play Store) in order to target users who owned a smartphone and could use our application in various natural settings.

We found a trend that users interacted with a 3D animated virtual human with gaze aversion while listening more, compared to communicating with a 3D animated virtual human

HAI 2015, October 21-24, 2015, Daegu, Kyungpook, Republic of Korea.
Copyright © 2015 ACM. ISBN 978-1-4503-3527-0/15/10$15.00.
DOI: http://dx.doi.org/10.1145/2814940.2814957

without gaze aversion while listening, a virtual human with a static visage, or an audio-only interface. We are reporting these findings that were obtained by analyzing objective data that we argue are more meaningful reflections of users impressions compared to subjective data. We describe our study and findings in detail in the following sections.

RELATED WORK AND CHALLENGES

Virtual humans with different modalities on mobile devices

The majority of smartphone based virtual human applications have been created for the purpose of casual chats or personal assistance. Ally [13], an intelligent personal medical assistant, incorporates language analysis, dialogue management, nonverbal response generation and presentation. Another example of a virtual human application is Siri, the consumer-level Apple companion using natural language technology. One notable downside of Siri is she does not have an image or human-like face that users can identify or connect with. Appearance is known to play a significant role in perception of virtual humans [6, 10]. Despite this, there are few studies that have manipulated the animation or fidelity of virtual humans displayed on mobile devices [10, 2, 19]. All of these studies explored different modalities of animated characters in the form of a virtual humans displayed on mobile devices.

However, these studies were conducted using characters that lacked certain facets of a real human figure or behavior in a lab setting and with only college student population. The details of the studies are described below:

Kang and colleagues [10] investigated users feelings of social copresence with their interaction partner presented using an avatar (controlled by a real human), specifically in emotionally engaged interactions, using a mockup of a cellphone on a laptop computer. The researchers found that human users were more engaged in conversations with animated characters, compared to static characters or no images at all. Although the avatars facial expressions matched the human communicators in real-time using a Logitech QuickCam Orbit MP camera equipped with face-tracking function, their eye movements were not mimicked. Rincon-Nigro and Deng [19] explored users' rapport with a communication partner in the form of a 3D animated conversational avatar. The avatar was created using FaceGen software, and driven by text on Nexus One HTC phones. The researchers discovered that most of the participants enjoyed talking with the 3D animated avatar more than the text-only version. Bickmore and Mauer [2] also investigated users' ability to connect with a relational agent presented on a Personal Digital Assistant (PDA). The authors discovered that users built stronger bonds with their agent that displayed animation with text and conversational nonverbal behavior. The agent presented facial expressions, head nods, eye gaze, posture shifts, and visemes. However, the agent was a low-fidelity 3D animated character with behavior that did not match the character's speech.

We describe why the nonverbal behavior of a virtual human plays a critical role in facilitating emotionally engaging social interactions in the following section.

Virtual humans' nonverbal behavior in social interactions

In both human-to-human and human-to-virtual human communication, researchers address the importance of nonverbal behavior in emotionally engaging communication [8, 10]. Poggi and Pelachaud [18] point out that it is critical to define how to express conversational agents' communicative intentions via a harmonious combination of verbal and nonverbal behavior in virtual human applications. Pelachaud's work [17], in particular, explores how nonverbal behavior represents the psychological state and affects communication between real humans and virtual agents in social interactions. Gratch and colleagues [7] argue that there are positive associations between the coordinated, responsive nonverbal feedback of virtual agents and real human users' perceived rapport. Among the various nonverbal behaviors, it is well known that eye gaze plays a prominent role in smooth communication and rapport building in virtual human applications [20]. For instance, Wang and Gratch [24] argue that the mutual gaze of a virtual human listener could help users increase rapport with a virtual human when continual mutual gaze is accompanied by other nonverbal feedback in a timely manner, compared to continuously mutual gaze with no nonverbal feedback or random gaze aversion. However, Andrist and colleagues [1] posit that appropriately-timed gaze aversion of a virtual human could induce greater disclosure and smoother turn-taking in conversations, compared to inappropriately-timed gaze aversion or no gaze aversion.

Although the literature has not reached a consensus regarding the ideal gaze patterns for a virtual human, one thing researchers agree on is that inappropriate gaze could negatively impact conversations at times, even worse than receiving no visual feedback at all [6, 1]. Everyday life may bring the experience of awkwardness or uncomfortable sentiments in reaction to continuous mutual gaze. On the other hand, total gaze aversion could also make a speaker think their partner is not listening. Previous work [24] further states that appropriate gaze behavior can be interpreted in different ways based on the nature of the social context. Our work aims to address this question of what constitutes appropriate eye gaze of a virtual human in emotionally engaged interactions.

In our study, the virtual human displayed gaze consisted of either constant mutual gaze or gaze aversion based on a statistical model of saccadic eye movement while listening [12]. Both gaze patterns were accompanied by other forms of appropriate nonverbal feedback. We describe how we created our animated virtual human and utilized it in this study in the following section.

3D CHAT-BASED ANIMATED VIRTUAL HUMANS

System Capabilities

We built our mobile virtual human using 3D techniques that are common in desktop applications. We used a 3D character with a human-like, but stylized (not photorealistic) mesh with artist-generated textures for skin hair, clothing and face coloring. The character consisted of approximately 15,000 polygons, with texture images of size 512x512. We use the terms 'character' and 'virtual human' similarly; the character is the 3D embodiment of the virtual human.

We animated a 3D character using a hierarchical set of joints representing the characters skeleton, which in turn deformed the mesh using linear blend skinning techniques. The software moved the characters' face using a set of joints that controlled the movement of the lips, mouth, tongue, eyes, eyebrows and cheeks. We used a set of static facial poses to represent FACS [5] units, and a separate set of static facial poses to define the lip and mouth movements necessary for lip syncing. The software produced the final face animation by combining the individual FACS Action Units and lip poses, weighted by a control signal.

We generated the control signal for lip syncing by recording the speech of a female actress saying each utterance, then producing the sequence of phonemes from a forced alignment process [3]. The phoneme sequence was then used in combination with a two-diphone based lip animation method [25] to generate the animation sequences for each utterance. Non-lip animation was generated automatically by an offline nonverbal behavior generator [16] which produces a set of behavior instructions in the Behavioral Markup Language (BML) [11] including head movements (nods, shakes, tilts), timed gesture instructions, such as beat, metaphorical, or deictic gestures. The lip syncing and behavior instructions were generated offline, and interpreted during runtime by a BML realizer [22]. Saccadic eye movements were used during the listening phase of one of the study conditions. Such movements were controlled by a statistical model of eye movements generated from a data set of listening behaviors taken from a human listener [12].

In addition, instructions for timing facial movements via AUs, such as raising or lowering eyebrows were specified. A BML instruction set consists of a set of behavioral instructions associated with one utterance, such as "move the head up and down", "make a pointing gesture", or "blink your eyes", along with timings in order to coordinate the various movements. The lip syncing and behavior instructions are generated offline, and interpreted during runtime by a BML realizer [22]. The animation system compiles those instructions into a set of coordinated movements, including handling conflicts between instructions, or movement descriptions that violate the movement capability of the character. The movements are then interpreted as a time series of translations, rotations and activation values which control the characters body, which in turn moves and alters the appearance of the 3D model.

Application Design

The character used two types of behaviors: 1) speaking behaviors and 2) listening behaviors with backchannel feedback. The speaking behaviors were generated by recording the voice of a female actor, then generating the behavioral description, lip syncing information, and recorded audio file for processing during runtime. The speaking behaviors were triggered when the virtual human described itself to a user, or asks a question. The listening behaviors were also generated by recording a set of responses such as 'uh huh', 'okay', 'I understand' and so forth and were triggered during an interaction while the user was speaking. Listening behaviors also consisted of head nodding and eyebrow raises that were triggered after a pause was detected in the user's response. In addition, the virtual human performed saccadic eye motions during the listening phase. These saccadic eye movements during the listening phase were timed according to a statistical model of listening, but not generated in response in a specific user action [12]. The saccadic eye movement was also used to display the gaze aversion of a virtual human in the study of Andrist et al [1]. This was done in contrast to an explicit model of eye gaze as a function of turn-taking and discourse structure as in [23]. For descriptive purposes, the use of a statistical model of saccadic eye movements during the listening turn will be described as 'gaze aversion', while the lack of such movement (and a fixed gaze) will be called 'no gaze aversion'.

The user interacted with the application in our study by pressing a button marked "Click and hold to speak" during his/her speaking turn. We had the user explicitly indicate when he/she was speaking in order to reduce errors during regulation of speaking turns between the virtual human and the user. The speech was captured by the mobile device's microphone, then sent to the Google Speech engine for transcription. If the speech translation engine was unsuccessful at interpreting the speech, the character responded by saying "Sorry, can you say that again?" and asked the user to enter another answer before proceeding to the next question. The character's responses were generated by natural language processing and dialogue management. The transcription of the user's speech and the duration of the speech turn were then stored in a database [4] along with the total length of time the user spent during the speaking turn.

Figure 1 describes the offline and online application process flow. We describe how we designed and conducted an experimental evaluation using our 3D chat-based virtual human in the following section.

METHOD

In contrast to most previous studies (see section 2-1), we argue that research using smartphone applications should be designed for their native setting (i.e. outside of the lab) in order to fully explore the potential of the device. Few studies have utilized this approach owing to a lack of state of the art virtual human applications. Our study evaluates a more flexible and advanced stand-alone application that was available on Google Nexus (4 or 5) or Samsung Galaxy (S4 or 5) and released via the Google Play Store.

We collected data in two notable ways in an effort to broaden the range of participants we could reach outside of a lab setting. First, we conducted the study by recruiting paid participants via Qualtrics who had access to a smartphone, were willing to participate in a chat with a virtual human, and could fill out online questionnaires on their smartphone before and after each interaction (Study A). Our second mode of recruitment relied upon users who downloaded the application via the Google Play Store and wanted to participate without getting compensation for their participation (Study B). The ap-

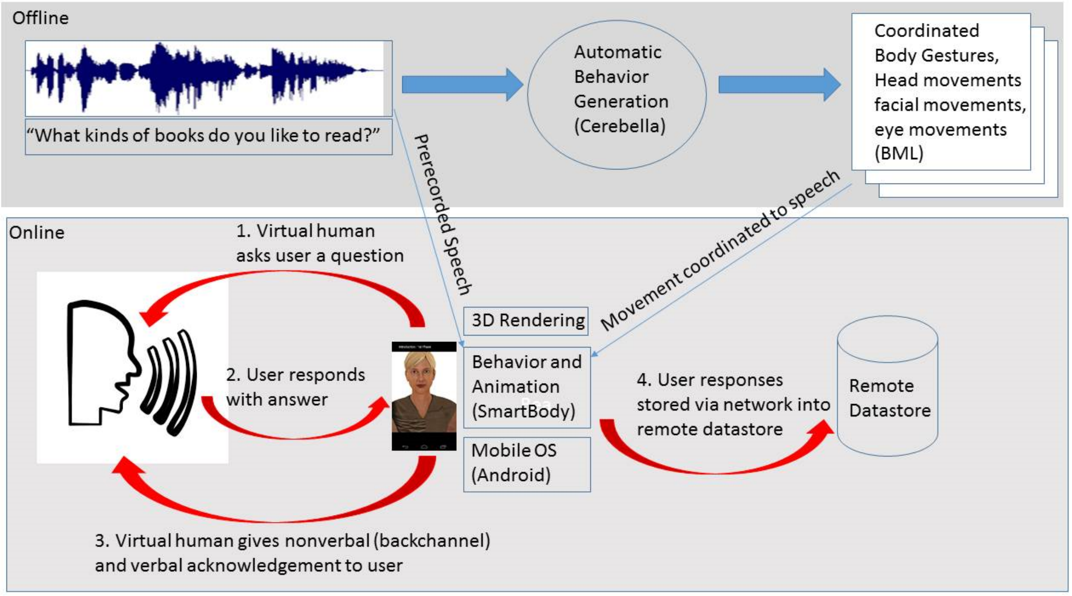

Offline
"What kinds of books do you like to read?"
Automatic Behavior Generation (Cerebella)
Coordinated Body Gestures, Head movements facial movements, eye movements (BML)
Prerecorded Speech
Movement coordinated to speech
Online
1. Virtual human asks user a question
2. User responds with answer
3D Rendering
Behavior and Animation (SmartBody)
Mobile OS (Android)
4. User responses stored via network into remote datastore
Remote Datastore
3. Virtual human gives nonverbal (backchannel) and verbal acknowledgement to user

Figure 1: [Top] Offline, a set of utterances are recorded and then processed by a non-verbal behavior generator (Cerebella) and a lip sync process. The results are stored in a BML file for later use during runtime. [Bottom] Online, a user listens to a virtual human, then responds by holding the 'Press to Speak' button, causing the virtual human to backchannel. The user responses to questions are stored in a remote datastore (Amazon Web Services). The system runs on an Android device using the SmartBody animation system.

proach of Study B allowed us to reach out to wider and more general population of smartphone users who were willing to participated in the study without being paid, compared to the population contacted via Qualtrics (Study A).

Study Design

This study examined users' perceptions and reactions to a virtual human based on various presentation types: (1) animation with gaze aversion, (2) animation with constant mutual gaze (no gaze aversion), (3) static image, and (4) no image (see Figure 2). The animation included facial expressions, head gestures, gaze, and other upper body movements using our 3D chat-based virtual human (see section 3). Because users were asked to use the button "Click and Hold to Speak" when they answered each question, we designed gaze aversion as a way to intentionally increase users' self-disclosure and comfort [1], rather than other functions such as turn-taking. We did not gauge users eye gaze as we could not control the users smartphone in its native setting. We further did not want the users to feel uncomfortable due to being recorded. Users answered a total of twenty four questions of increasing intimacy asked by the virtual human (e.g. "What are your favorite sports?"). We borrowed the structure and context of the questions from the studies of Kang and colleagues [9, 10]. Since smartphones were treated as an icon of emotionally engaged communication [10], the conversation scenario in our study imitated casual chats in the format of an interview in a counseling situation to maintain the emotionally engaged interaction. During the conversation, the virtual human responded to users' utterances with its own back stories in order to reciprocate intimate information sharing and advance the conversation (e.g. "I like to play very active sports like basketball and tennis."). The self-disclosure of the virtual human was pre-scripted, but other verbal responses were generated by natural language processing and dialogue management. Most experimental study designs [10] include approximately ten question items for a one time interaction. Thus, we designed each session to consist of twelve questions for our study. A total of twenty four questions in the two sessions allowed users to have enough interaction time with a virtual human.

Participants and Procedure

For Study A, a total of 89 participants (35% men, 65% women; *M*=38.8, *SD*=11.4) were randomly assigned to one of 4 conditions: animation with gaze aversion (N=22), animation without gaze aversion (N=21), static image (N=21), and no image (N=25). The participants were given $5 compensation when they completed the study. Table 1 demonstrates the procedure.

Participation required a total of 35 minutes on an individual basis. The pre-questionnaire included questions pertaining to users' demographics. There were two types of the post-questionnaires. All users received the first post-questionnaire, which included metrics to rate their perception of virtual rapport with and social attraction toward a virtual human. The second post-questionnaire was also given to all users regardless of participating in another conversation with a virtual human for the 12 additional questions. It gauged the driving factors behind the users' choice to continue or not continue conversing with the virtual human. It was mandatory to complete

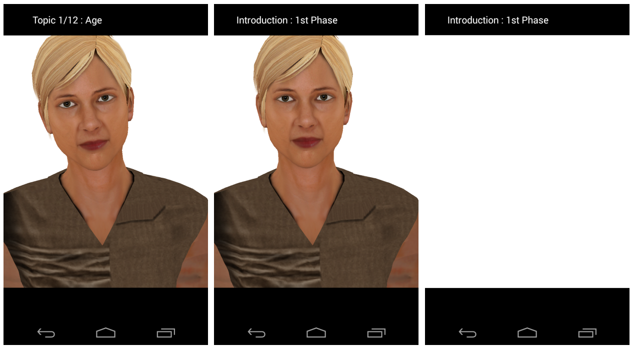


Figure 2: Screen captures from the app showing the 4 conditions of the study. The first and second condition (left) used an animated virtual human, (center) a static image, and (right) no image, only audio.

	Study A Procedure
i)	Participants fill out pre-questionnaire
ii)	Participants download and install mobile app
iii)	Participants answer 12 questions asked by a virtual human
iv)	Participants answer the first post-questionnaire
v)	Participants are asked if they would like to continue, and if so, answer up to 12 additional questions
vi)	Participants answer the second post-questionnaire

Table 1: Study A, 89 participants

	Study B Procedure
i)	Participants download and install mobile app
ii)	Participants answer 12 questions asked by a virtual human
iii)	Participants are asked if they would like to continue, and if so, answer up to 12 additional questions

Table 2: Study B, 233 participants

the first session and two post-questionnaires to get compensation, but the second conversation was optional. This was done in order to effectively observe whether users enjoyed conversing with the virtual human.

We were motivated to conduct a follow up study based on our results from Study A. Study B consisted of a total of 233 participants as the participants in Study A were also included. In Study B, we utilized the same mobile app and 4 conditions noted above. The only exception is that participants in Study B were not required to fill out a pre-questionnaire and post-questionnaires. Thus, we did not have participants' demographic information. Participants were also randomly assigned to one of the 4 conditions: animation with gaze aversion (N=66), animation without gaze aversion (N=55), static image (N=47), and no image (N=65). Table 2 outlines this procedure.

Measurements

For objective measures, in Study A, we analyzed users' feedback derived from their text input in the post-questionnaires. The data was categorized into three types: 1) the number of questions that a user answered (by asking the user to enter the last question that he/she answered), 2) the completion of all 24 questions, and 3) negative reasons for quitting the conversation. In Study B we analyzed users' transcription of their verbal responses to questions that included the transcription of Study A participants' verbal responses. We collected data for the total time spent answering questions.

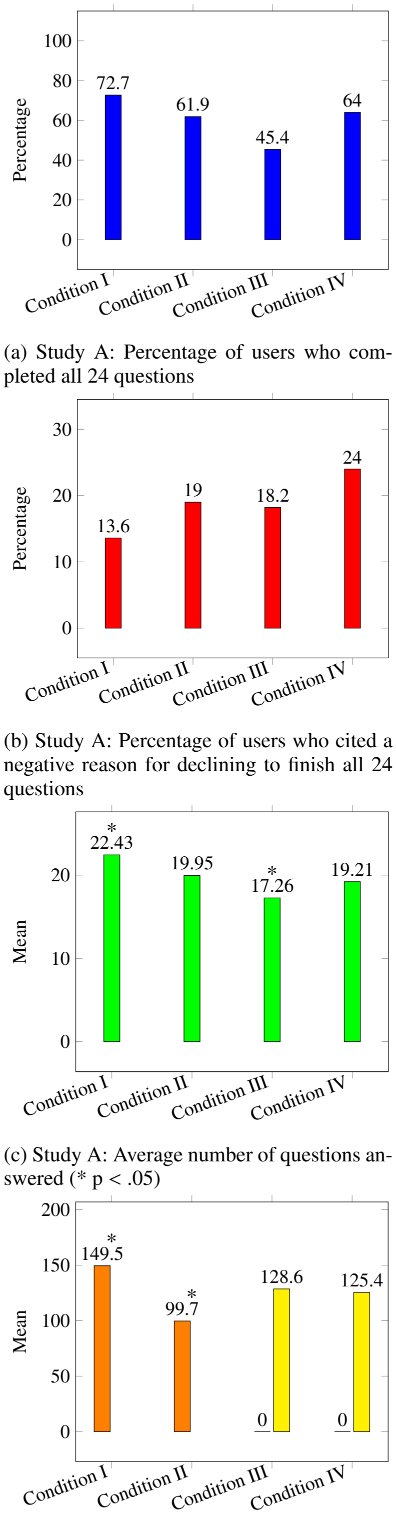
For subjective measures, in the first post-questionnaire, we utilized Social Attraction to measure users' feelings of attraction toward a virtual human. We also measured Virtual Rapport to assess users' feelings of rapport with a virtual human. In the second post-questionnaire, we further asked an additional question related to likelihood to converse with the virtual human in the future (e.g. "I would look forward to another conversation with the virtual human."). These scales contained a Likert-type 5-point metric for items. All the scales described above showed good reliability (Social Attraction: Cronbach's alpha = .923, Virtual Rapport: Cronbach's alpha = .959).

RESULTS

Results for the objective measures (Study A and B)

Study A.

We coded the completion of the conversation with a binary measure (completion and no completion). We also examined



(a) Study A: Percentage of users who completed all 24 questions

(b) Study A: Percentage of users who cited a negative reason for declining to finish all 24 questions

(c) Study A: Average number of questions answered (* p < .05)

(d) Study B: Average amount of time (seconds) spent answering questions (* p < .05)

Figure 3: Results for the objective measures in each of 4 conditions (Conditon I: animation with gaze aversion, Condition II: animation without gaze aversion, Condition III: static image, Condition IV: audio-only)

users' reasoning for quitting the conversation using a binary measure (negative reasons and non-negative reasons). The items for a negative reason of quitting conversation included: "I feel it was canned responses," "Talking to AI is silly and dumb," and "I'm no longer interested." To analyze this data, we ran a Chi-square test to explore the associations between the two categorical variables. We did not find statistically significant differences among the conditions. However, we discovered notable trends that are presented in Figure 3, (a) and (b). The results show that users were more likely to complete all 24 questions when interacting with an animated character with gaze aversion. The results further demonstrate that users were more likely to cite negative reasons for declining to answer all 24 questions when interacting with an audio-only interface. The items for a non-negative reason of quitting conversation included: "I ran out of time" and "The conversation was over."

To measure the length of the conversation, we used the number of the last question that the user answered before stopping. We had to eliminate the data for six participants in our study given that they did not remember what question they last answered. To analyze the remaining data, we performed a Between-Subjects ANOVA. Our results [F(3, 79)=2.89, p=.040] with Tukey HSD Test demonstrate that users answered more questions when they interacted with animated characters that demonstrated gaze aversion (M=22.43, SD=3.79), compared to interacting with static characters (M=17.26, SD=6.61). Results are reported in Figure 3, (c). There was no other significant difference between the other conditions, however there was a trend that shows users answered more questions when communicating with animated character with gaze aversion, compared to communicating with animated character with no gaze aversion (M=19.95, SD=5.91) or no image at all (M=19.21, SD=5.93).

Study B.
We analyzed the objective data for the duration of users' responses (see the graph (d) in Figure 3). The users in the animation condition with gaze aversion (149.5 seconds) tended to talk longer than users in the other conditions (animation without gaze aversion: 99.7 seconds, static: 128.6 seconds, no image: 125.4 seconds). There was no statistically significant difference among the 4 conditions. However, for only gaze related conditions, the results of an Independent-Samples T-Test analysis show that there was a strong trend [t(107.22)=2.297, p=.024] that users talked for a longer time with an animated character with gaze aversion (M=149.47, SD=148.54) than an animated character without gaze aversion (M=99.67, SD=86.39).

Results for the subjective measures (Study A)

Study A.
We performed a Two-Way Between- Subjects ANOVA with two independent variables: condition and users' gender. We had a female version of a virtual human only, thus we wanted to further explore how users' perception of and responses to the female virtual human could have been driven by gender differences. We did not find a statistically significant

difference for the 4 conditions overall. However, Between-Subject ANOVA results [F(1, 81)=4.85, p=.030] demonstrate that males (M=3.40, SD=1.17) were socially attracted to the virtual human in our applications overall more than females (M=2.87, SD=1.01). The results also [F(1, 81)=4.97, p=.029] demonstrate that males (M=3.81, SD=1.17) reported wanting to have another conversation with the virtual human in our applications more than females (M=3.07, SD=1.01).

These outcomes indicate that male users might have felt more socially engaged with our virtual human and wanted to have another interaction compared to female users. This may have been because the virtual human was portrayed as a female.

DISCUSSION AND FUTURE WORK

This study successfully utilized a virtual human's nonverbal behavior when presented on smartphone devices to explore its effect on users' responses. We achieved this by conducting research outside the restrictions of the lab where the potential of such devices could be fully explored. Namely, we incorporated our state of the art application on smartphones that were used in real world settings with no limitation. Specifically, we investigated whether the nonverbal behavior of a virtual human would encourage people to talk for a longer amount of time when displayed on the small screen of a smartphone. We further explored whether gaze aversion as a listening behavior exhibited by 3D virtual humans would affect the interaction.

We found that the users continued to talk more with a 3D animated virtual human exhibiting gaze aversion while listening, compared to a static virtual human. We also discovered that the users inclined to talk for a longer amount of time with a 3D animated virtual human that displayed gaze aversion while listening, compared to interacting with a 3D animated virtual human that did not present gaze aversion while listening. These results mirror the existing findings that people are more likely to engage in conversations on a smartphone with a virtual human that displays nonverbal behavior including gaze aversion in a timely manner. Therefore, the results of our study go beyond the body of existing research by validating the previous findings in real world settings.

Our results further demonstrate a trend toward people responding to animated characters (with or without gaze aversion) and faceless voice-only characters alike regarding their interaction time with the characters, while people were more likely to cite negative reasons for declining to complete their conversation with the voice-only characters. This implies that although facial cues are known to help communicators avoid or resolve conversational misunderstandings [15] and deliver emotional signals most proficiently in social interactions [6, 8], one cannot anticipate a difference in engagement time with either of the two prototypes. We offer that this result may be due to the way that people interact with their smartphones. Namely, we are likely to be engaged in multiple tasks at the same time when using a smartphone in natural settings. Such behavior leaves little room for the user to focus their visual attention on the smartphone screen. If this is indeed the case, then the visualization of our virtual character might not have affected users' interaction to the extent that we anticipated. Another reason for the result might be related to the incongruity of a 3D animated character coupled with the voice of a real human, violating user expectations in sort of an "uncanny valley" effect [21]. We will further investigate the effect of other factors such as multi-tasking on users' engagement in their interaction with characters in future work.

With regard to gaze, the results of our study revealed that users interacted for a longer period of time with an animated virtual human that averted its gaze while listening, compared to an animated virtual human that did not avert its gaze. Based on this observed trend, we suggest that a virtual human should avert its gaze while listening in interactions in order to elicit greater engagement from human users. Our findings do not affirm whether gaze aversion significantly provokes higher levels of engagement in human users concerning a type of gaze aversion other than the listening behavior. Our future work will explore a different type of gaze aversion such as a function of turn-taking and discourse structure [23] in various social contexts.

It has been known that visible verbal behavior enhances the comprehension of verbal cues more, compared to listening to voice only [13] report. If people view a virtual human's appropriately-timed nonverbal behavior in general, they will be better able to comprehend the content of the agent's speech rather than merely listening. Gaze is also well known to play a prominent role in rapport-building and effective communication in human-to-virtual human interactions [24, 1]. Thus, we aim to continue improving the nonverbal behavior of our virtual humans including eye gaze to generate more salient, and true to nature emotional signals. This will be specifically pertinent when interacting with virtual humans on smartphones, which are a medium through which people tend to share more personal information with others compared to any other communication media [10].

Although the literature has not reached a consensus regarding the ideal gaze patterns for a virtual human, one thing researchers agree on is that inappropriate gaze could negatively impact conversations at times, even worse than receiving no visual feedback at all [1]. Everyday life may bring the experience of awkwardness or uncomfortable sentiments in reaction to continuous mutual gaze. On the other hand, total gaze aversion could also make a speaker think their partner is not listening. Previous work further states that appropriate gaze behavior can be interpreted in different ways based on the nature of the social context [19]. These concerns are valid, and we keep them in mind while continually striving to develop the appropriate nonverbal behaviors in virtual humans displayed on smartphones.

In addition, this study found that male users were more socially attracted to and wanted to have an interaction with a female character more so than female users. In our future work, we will conduct an investigation using both a male and a female virtual human to further examine the role of gender effect on users' responses.

We are currently upgrading the behavior of our virtual human to improve some of its nonverbal and verbal behavior, such as delivering better empathetic feedback in an appropriate manner that the current study is lacking. We are also planning on implementing more spontaneous conversation between a user and a virtual human by applying further natural language processing algorithm. Once we complete upgrading our virtual human for use on a smartphone, we will fully carry out our experiment for a comparison between virtual humans that embody different verbal and nonverbal behavior by releasing our application via the Google Play Store online.

REFERENCES

1. Sean Andrist, Bilge Mutlu, and Michael Gleicher. 2013. Conversational gaze aversion for virtual agents. In *Intelligent Virtual Agents*. Springer, 249–262.
2. Timothy Bickmore and Daniel Mauer. 2006. Modalities for building relationships with handheld computer agents. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM, 544–549.
3. Fabio Brugnara, Daniele Falavigna, and Maurizio Omologo. 1993. Automatic segmentation and labeling of speech based on Hidden Markov Models. *Speech Communication* 12, 4 (1993), 357–370.
4. Amazon Elastic Compute Cloud. 2011. Amazon web services. *Retrieved November* 9 (2011), 2011.
5. Paul Ekman and Wallace V Friesen. 1977. Facial action coding system. (1977).
6. Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 529–536.
7. Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Intelligent Virtual Agents*. Springer, 125–138.
8. Carroll E Izard. 1997. Emotions and facial expressions: A perspective from Differential Emotions Theory. *The psychology of facial expression* (1997), 57–77.
9. Sin-Hwa Kang and Jonathan Gratch. 2012. Socially anxious people reveal more personal information with virtual counselors that talk about themselves using intimate human back stories. *Annual Review of Cybertherapy and Telemedicine* 181 (2012), 202–207.
10. Sin-Hwa Kang, James H Watt, and Sasi Kanth Ala. 2008. Social copresence in anonymous social interactions using a mobile video telephone. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1535–1544.
11. Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R Thórisson, and Hannes Vilhjálmsson. 2006. Towards a common framework for multimodal generation: The behavior markup language. In *Intelligent virtual agents*. Springer, 205–217.
12. Sooha Park Lee, Jeremy B Badler, and Norman I Badler. 2002. Eyes alive. In *ACM Transactions on Graphics (TOG)*, Vol. 21. ACM, 637–644.
13. Anton Leuski, Rasiga Gowrisankar, Todd Richmond, Ari Shapiro, Yuyu Xu, and Andrew Feng. 2014. Mobile personal healthcare mediated by virtual humans. In *Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces*. ACM, 21–24.
14. Martin Lindstrom. 2011. You love your iPhone. Literally. *New York Times* 1 (2011), 21A.
15. Stephen W Littlejohn and Karen A Foss. 2010. *Theories of human communication*. Waveland Press.
16. Stacy Marsella, Yuyu Xu, Margaux Lhommet, Andrew Feng, Stefan Scherer, and Ari Shapiro. 2013. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 25–35.
17. Catherine Pelachaud. 2009. Studies on gesture expressivity for a virtual agent. *Speech Communication* 51, 7 (2009), 630–639.
18. Isabella Poggi, Catherine Pelachaud, and Fiorella De Rosis. 2000. Eye communication in a conversational 3D synthetic agent. *AI communications* 13, 3 (2000), 169–181.
19. Mario Rincón-Nigro and Zhigang Deng. 2013. A text-driven conversational avatar interface for instant messaging on mobile devices. *Human-Machine Systems, IEEE Transactions on* 43, 3 (2013), 328–332.
20. Kerstin Ruhland, Sean Andrist, Jeremy Badler, Christopher Peters, Norman Badler, Michael Gleicher, Bilge Mutlu, and Rachel Mcdonnell. 2014. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics State-of-the-Art Report*. 69–91.
21. Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *PRESENCE: Teleoperators and Virtual Environments* 16, 4 (2007), 337–351.
22. Ari Shapiro. 2011. Building a character animation system. In *Motion in Games*. Springer, 98–109.
23. Obed Torres, Justine Cassell, and Scott Prevost. 1997. Modeling gaze behavior as a function of discourse structure. In *First International Workshop on Human-Computer Conversation*.
24. Ning Wang and Jonathan Gratch. 2010. Don'T Just Stare at Me!. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1241–1250.
25. Yuyu Xu, Andrew W Feng, Stacy Marsella, and Ari Shapiro. 2013. A practical and configurable lip sync method for games. In *Proceedings of Motion on Games*. ACM, 131–140.