Smart Mobile Virtual Humans: "Chat with Me!"

Sin-Hwa Kang, Andrew W. Feng, Anton Leuski, Dan Casas, and Ari Shapiro

Institute for Creative Technologies, University of Southern California, USA
{*kang*, *feng*, *leuski*, *casas*, *shapiro*}@ict.usc.edu

1 Introduction

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 [5], compared to non-animated or audio-only characters in everyday life. Based on previous studies [2, 7, 10], 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 also found that 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, 4]. Everyday life may bring the experience of awkwardness or uncomfortable sentiments in reaction to continuous mutual gaze. On the other hand, gaze aversion could also make a speaker think their partner is not listening. Our work further aims to address this question of what constitutes appropriate eye gaze in emotionally engaged interactions.

We developed 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 (see Figure 1). The virtual human displayed appropriate gaze that was either consisted of constant mutual gaze or gaze aversion based on a statistical model of saccadic eye movement [8] while listening. Both gaze patterns were accompanied by other forms of appropriate nonverbal feedback. 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.

2 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. The animation included facial expressions, head gestures, gaze, and other

upper body movements using our 3D chat-based virtual human (see Figure 1). 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. 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 [6]. Since smartphones were treated as an icon of emotionally engaged communication [7], 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.

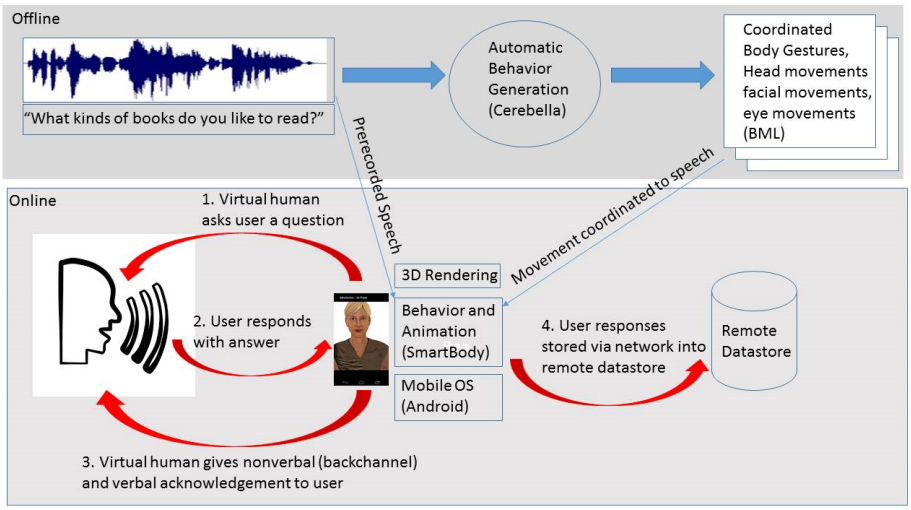

Fig. 1: [Top] Offline, a set of utterances are recorded and then processed by a nonverbal behavior generator (Cerebella [9]) and a lip sync process [11] The results are stored in a BML file for later use during runtime. [Bottom] Online, a user listens to a virtual human, then responds by holding the 'Press to Speak' button, causing the virtual human to backchannel. The user responses to questions are stored in a remote datastore (Amazon Web Services [3]). The system runs on an Android device using the SmartBody animation system.

For Study A, a total of 89 participants (35% men, 65% women; average 39 years old) 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. 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 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).

3 Preliminary Results and Discussion

For Study A, 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 ($M=17.26$, $SD=6.61$). 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 ($M=19.21$, $SD=5.93$). For Study B, we analyzed the objective data for the duration of users responses. 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 ($M=99.67$, $SD=86.39$). Regarding subjective measures, we did not find statistically significant difference for the conditions in the results of the study overall.

In general there was a trend that users interacted with a 3D animated virtual human with gaze aversion more, compared to communicating with a 3D

animated virtual human without gaze aversion, a virtual human with a static visage, or an audio-only interface.

This study successfully utilized a virtual humans nonverbal behavior when presented on smartphone devices to explore its effect on users responses. The results of our study go beyond the body of existing research by validating the previous findings in real world settings where the potential of such smartphone devices could be fully explored with no limitations. 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.

References

1. Andrist, S., Mutlu, B., Gleicher, M.: Conversational gaze aversion for virtual agents. In: Intelligent Virtual Agents. pp. 249–262. Springer (2013)
2. Bickmore, T., Mauer, D.: Modalities for building relationships with handheld computer agents. In: CHI'06 Extended Abstracts on Human Factors in Computing Systems. pp. 544–549. ACM (2006)
3. Cloud, A.E.C.: Amazon web services. Retrieved November 9, 2011 (2011)
4. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., Sasse, M.A.: 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. pp. 529–536. ACM (2003)
5. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. In: Intelligent Virtual Agents. pp. 125–138. Springer (2007)
6. Kang, S.H., Gratch, J.: 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, 202–207 (2012)
7. Kang, S.H., Watt, J.H., Ala, S.K.: Social copresence in anonymous social interactions using a mobile video telephone. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1535–1544. ACM (2008)
8. Lee, S.P., Badler, J.B., Badler, N.I.: Eyes alive. In: ACM Transactions on Graphics (TOG). vol. 21, pp. 637–644. ACM (2002)
9. Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., Shapiro, A.: Virtual character performance from speech. In: Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 25–35. ACM (2013)
10. Rincón-Nigro, M., Deng, Z.: A text-driven conversational avatar interface for instant messaging on mobile devices. Human-Machine Systems, IEEE Transactions on 43(3), 328–332 (2013)
11. Shapiro, A.: Building a character animation system. In: Motion in Games, pp. 98–109. Springer (2011)