Study comparing video-based characters and 3D-based characters on mobile devices for chat

Sin-Hwa Kang[*1], Andrew W. Feng[†1], Mike Seymour[‡2], and Ari Shapiro[§1]

[1]Institute for Creative Technologies, University of Southern California
[2]University of Sydney

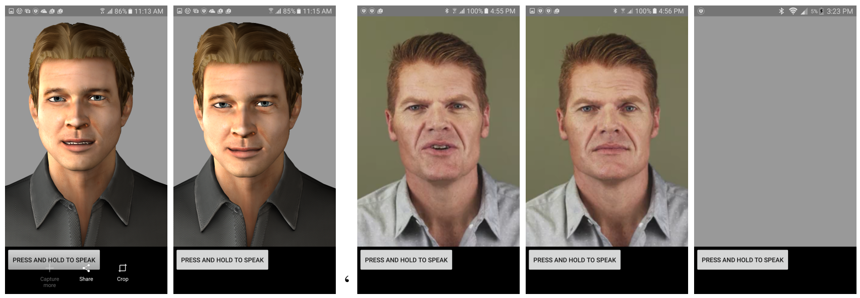


Figure 1: *An 'in the wild' study comparing a chat interaction between a 3D-based character, a video-based character, and an audio-only interface.*

Abstract

This study explores presentation techniques for a 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 an animated virtual character as opposed to a real human video character capable of displaying realistic backchannel behaviors. An audio-only interface is compared additionally with the two types of characters. The findings of our study suggest that people are socially attracted to a 3D animated character that does not display backchannel behaviors more than a real human video character that presents realistic backchannel behaviors. People engage in conversation more by talking for a longer amount of time when they interact with a 3D animated virtual human that exhibits realistic backchannel behaviors, compared to communicating with a real human video character that does not display backchannel behaviors.

Keywords: virtual humans, agents, smartphones, chat applications, nonverbal behavior, self-disclosure, rapport, reciprocity, facial expressions

Concepts: •**Computing methodologies** → **Perception;**

1 Introduction

Interactive 3D characters and agents are an important part of many games, simulations and other interactive, digital experiences. A traditional method of generating such characters involves generating the 3D models and control algorithms using 3D production tools and real-time 3D simulation platforms such as game engines. Recently, mobile devices have become important platforms for communication and entertainment. Modern mobile platforms are typically powerful enough to run real-time 3D simulations like their desktop counterparts. The advantages to using 3D for characters are numerous, including the ability to change the appearance of the 3D character, create novel responses and interactions, change the camera angle, lighting and so forth. Some disadvantages to using 3D for an interactive character include the complexity of creating such an entity and the difficulty in creating a high fidelity appearance. To clarify, we use the term *3D* to refer to the type of media and process that is used to generate the final content, although the final result is display in 2D on a screen. This is in contrast to a virtual reality environment or holographic display environment where the content is seen in 3D directly. An alternative method to creating an interactive 3D character that can respond to a user is to use a series of 2D video clips captured from a 2D camera trained on an actor and to arrange the resulting video clips in a directed graph structure that determines which clip will be played depending on the state of the character and the conversation. For example, a single video clip can be played when the character asks a question, then another to respond to a user, then another for the character to wait for a response. By playing a video clip that is appropriate to the interaction using consistent camera framing, a set of video clips can be dynamically arranged to yield the perception of interactivity. We term this style of interaction *interactive video graphs*, as the state of the simulation in combination with the connectivity of the graph dynamically determines the next video to be played, yielding a dynamically changing presentation for the purpose of interactivity. Figure 2 shows the similarities and differences between a 3D-generated state-based architecture, and the *interactive video graph* approach we are presenting. Researchers in computer graphics and human modeling have been attempting to generate 3D animated characters that can match the behavioral and visual fidelity of a real person. Often the result can elicit the uncanny valley effect where the appearance and behavior of the virtual characters can not match those of a real person to a sufficient degree, causing a repulsion to such characters. Therefore, we aimed to discover a novel way to respond to the issue. We investigated how users felt when interacting with a real human video character via *interactive video graphs* that matched their appearance and behavior, compared to a human-looking but not photorealistic 3D animated character. We

*kang@ict.usc.edu
†feng@ict.usc.edu
‡mike.seymour@sydney.edu.au
§shapiro@ict.usc.edu

MIG 2016, October 10-12, 2016, San Francisco, CA
ISBN: 978-1-4503-ABCD-E/16/07
DOI: http://doi.acm.org/10.1145/9999997.9999999

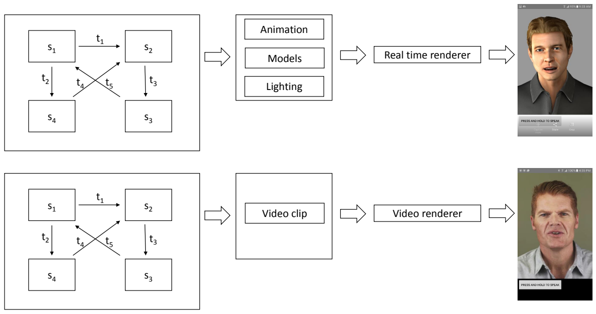


Figure 2: *A state-transition graph for 3D characters (above) and for an interactive video graph for video-based characters (below). The s represent states, while the t represent transitions from those states. In the case of the interactive video graph, the video clip shown is directly related to the current state. In contrast, the 3D character's appearance is rendered by combining several inputs.*

conjecture that people would expect a video character to respond better without any discrepancy between appearance and behavior, compared to a 3D animated character. In addition, we compared the two types of characters that are described above and an audio-only character. We were motivated to investigate this comparison as we expect our findings might be able to provide other answers about whether Siri, the consumer-level Apple companion using natural language technology, would need a human-like face that users can identify or connect with [Kang et al. 2015].

The ultimate goal of our study is to explore the best practice of implementing feasible characters in social interactions using a smartphone as the medium of communication.

2 Background and Challenges

2.1 Background on the Uncanny Valley

The *uncanny valley* phenomenon [Mori et al. 2012] reported in robotics research, in which anthropomorphic robots whose behavior (kinetic realism) is slightly inappropriate, but whose visual realism is high are seen negatively as somewhat *creepy.* To better understand the uncanny valley effect, [McGloin et al. 2009] have presented some evidence that expectancy violation may produce the uncanny valley effect. Users expect more realistic behavior from realistic anthropomorphic characters, and if this expectancy is violated, evaluate the characters more negatively than less realistic ones. There are numerous studies that explored relationships between appearance and behavior of a virtual human whose outputs demonstrate the uncanny valley effect. [McDonnell et al. 2012] investigated the uncanny valley by varying the rendering quality of a character appearance, and discovered that photo-realistic characters and very cartoonish ones were rated with the highest scores on the social aspects of user perceptions of the characters. However, lack of motion in a photo-realistic character was rated lower, compared to the lack of animation in a cartoonish character. In another study [Kokkinara and McDonnell 2015], researchers explored a photo-realistic character versus a cartoonish character when the characters displayed animation using a real-time tracking of users' own faces. The researchers found that the photo-realistic face appealed equally to the cartoon face regardless of different levels of animation realism. This is contradictory to the uncanny valley effect in which the findings of previous studies that demonstrated negative feelings of lower animation realism for highly realistic virtual humans, compared to the realism for cartoonish characters. The researchers argued that the interesting findings might be owing to users' agency and ownership of the real-time tracked faces. However, the decreased animation realism elicited the decrease of character appeal although the animation realism did not affect the perceived agency and ownership of the character. The conclusions of the studies suggest that there is no difference for different levels of appearance realism, but different levels of behavioral realism. Other studies further investigated whether the decrease of behavioral realism of virtual characters would affect user social responses. Gratch and colleagues [Gratch et al. 2007; Wang and Gratch 2010] compared realistic 3D animated characters with appropriate nonverbal responses and characters without appropriate nonverbal responses and found people felt greater rapport with the characters with appropriate nonverbal responses. [Hodgins et al. 2010] also investigated how the decrease of behavioral realism of a photo-realistic character created using a motion-capture system would affect users' emotional reactions and discovered that removal of facial expressions and/or voice decreased users' emotional richness in their reactions to video scenes in which the photo-realistic characters were acting. Furthermore, [Kokkinara and McDonnell 2015] examined the effect of full or partial facial movements on users' perception of virtual characters when the users experienced real-time tracking data of themselves over their own characters. The character appearance displayed two different forms of visual realism: realistic or cartoonish. They found that users liked higher degrees of behavioral realism obtained via real-time tracking. These studies demonstrate that the decrease of behavioral realism in virtual characters would affect user perceptions of the virtual characters.

As we have discovered in existing studies, discrepancy between appearance and behavior would elicit the uncanny valley effect. We questioned whether there would be a difference in users' perceptions of virtual characters between a real human video character that displays the highest behavioral realism in facial expressions, nodding, eye movements, and head orientation and a human-like 3D animated character that presents greatly realistic animation. There are no studies that have compared a real human video character versus a human-like 3D animated character to explore the uncanny valley effect. If the real human video character could not respond to user behaviors properly, would the character be rated lower for social aspects than the 3D animated character that could not respond to user behaviors properly or vice versa? Per the uncanny valley theory, people might feel the uncanny valley more with the real human video character if the character does not respond properly, compared to the 3D animated character when it does not since the real human video is expected to display the highest behavioral realism.

2.2 Background and Motivation for Video-Based Interactive Characters

There is a large body of research in computer graphics and animation that covers the manipulation and editing of video- or image-based data. By contrast, an interactive video graph approach does not manipulate the imagery within each clip, but rather determines the order in which the clips are played. The use of prepared videos for the purpose of character interaction is not a new concept. [Marinelli and Stevens 1998] terms this kind of interaction for the purpose of dyadic interaction as the *Synthetic Interview*. [Traum et al. 2015a] describes this as *time offset interaction* describing the collection of video data at a different time from the synthetic interaction, and used it to interact with a Holocaust survivor using large database of answers [Traum et al. 2015b]. Demonstrations that use video for the purpose of interactions has been used for historical information in *Ben Franklin's Ghost* [Sloss and Watzman 2005] by capturing video of a historical figure look-alike. Likewise, an interactive man in a chicken costume (*Subserviant Chicken* used videos to respond to user requests as part of a marketing campaign [Ochman 2004]. Fundamentally, the approach differs from a traditional 3D approach in that the data has been pre-rendered and the collection of rendered videos is used to simulate interaction. Similarly, [Rizzo et al. 2011] used prerendered videos generated from 3D content to construct a virtual online coach where idling, talking and responsive actions were streamed to a web-based client based on the interaction state.

There are some advantages to using an *interactive video graph* over a traditional 3D character interaction. For one, it is difficult to generate a photorealistic controllable character using 3D methods, as they often require specialized capture equipment as well as complex processing and digital artistry. In addition, many near-photorealistic 3D characters can trigger the uncanny valley effect as explained above. By contrast, capturing 2D video capture is well understood, and the tools and equipment needed to do so, such as cameras, microphones and lights, are ubiquitous and inexpensive. There are several disadvantages to using an interactive video graph approach versus a traditional interactive 3D character approach. Since videos are captured, editing and changing the content (and thus the video clip) is difficult, and typically requires an entirely new capture clip to be produced. In addition, transitions between video clips in the interactive video graph will cause a visual pop or jitter between last frame of the first clip and the first frame of the second clip since it is unlikely that the actor will maintain an identical position in the camera frame between clips. Such transition effects can be modified by using 2D camera effects, such as blurring or cross fading, or other methods of 2D transitions by creating models that represent the character which are then reprojected back into 2D. Thus the length of the video clip must be long enough so as to not become a distraction to the user. Thus, a video clip from an *interactive video graph* is practically limited in the clips that can be played by the amount the transition will be distracting. By contrast, a 3D approach would allow a per-frame change to the 3D character given the proper instructions, including the possibility of layering animation for the purpose of controlling different parts of the body. Also, in contrast to an interactive 3D approach, an *interactive video graph* approach requires that the videos have already been rendered and can either be displayed on the mobile device directly or streamed from a remote server with a minimum of latency. Using the interactive 3D approach, the character is typically rendered in real-time. Another disadvantage is that a video clip will typically also require more storage space and greater network bandwidth than a set of instructions to animate a 3D character would. To display a 3D character on a mobile device, the 3D assets and rendering capability needs to exist on the device. In order to animate such a character, the cost is dependent on the number of degrees of freedom (DOF) of the character per frame. In contrast, in order to animate a video, the cost is the size of the video per frame.

Despite the disadvantages of using an *interactive video graph*, there are numerous instances where such a mechanism would be an advantage, such as one where the visual fidelity of the character is important, or where the appearance of the character must match to a high degree a particular person, or when authenticity is desired. In addition, applications where the camera and lighting are fixed, and where the character generally remains in the same position in the field of view lessen the disadvantages of the interactive video graph approach versus the traditional 3D character approach. We have identified one such instance; to use an agent to converse with a user for the purpose of dialogue or information intake. The agent can be displayed with a non-moving camera trained on the character's face and upper body, similar to the view one would see during a video chat. In this circumstance, the lighting and cameras are fixed,

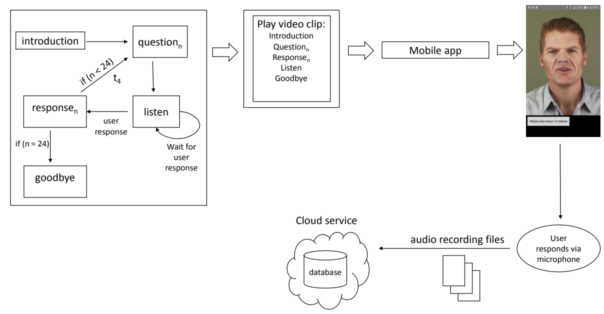


Figure 3: *Architecture for video-based interactive character using an interactive video graph. Users are asked questions by the character, which they respond via mobile device microphone. Recordings are then stored on a cloud server for later processing.*

the character generally does not move into or out of frame, and the position of the character generally remains the same. In addition, a turn-based dialogue application will require the character and user to hand off the dialogue turn for a period of time necessary to play a video clip.

3 Architecture

The character used two types of behaviors: 1) speaking behaviors and 2) listening behaviors with backchannel feedback. For a video-based character, we recorded numerous videos of a human asking questions and responding to speech. The video-based character's listening behaviors included a real human's generic listening behaviors such as facial expressions, head nodding, eyebrow raises, and slight smiles. For the 3D animated character, the speaking behaviors were generated by using the recorded voice of the real human, then nonverbal behaviors such as head movements, lip synch to speech and gestures were generated automatically from a automatic behavior generating system [Marsella et al. 2013] during runtime executed on a real-time 3D virtual character system [Feng et al. 2015]. The listening behaviors, which included head nods, slight smiles and eyebrow raises, were generated while the user was speaking. 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 a cloud file storage (Amazon Web Services) for offline transcription. Figure 3 describes the offline and online application process flow. We describe how we designed and conducted an experimental evaluation using our virtual characters in the following section.

4 Experimental Evaluation

This study evaluates a more flexible and advanced stand-alone application that was available on Google Nexus (any versions greater than 5) or Samsung Galaxy (any versions greater than S5) and released via the Google Play Store. 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.

4.1 Study Design

Our study examined users' perceptions and reactions to a virtual human based on various presentation types: (1) video with nonverbal backchannels, (2) video without nonverbal backchannels, (3) animation with nonverbal backchannels, (4) animation without nonverbal backchannels, and (5) audio-only. To test a comparison between the presence and absence of characters of any type, an audio-only condition was added to the design. In this study, we manipulated realistic nonverbal backchannels to investigate the effect of behavioral realism on the interaction between users and virtual characters. The nonverbal backchannels included facial expressions, head gestures, gaze, and other upper body movements. Because users were asked to use the button 'Click and Hold to Speak' when they answered each question, we designed nonverbal backchannels as a way to intentionally increase users' self- disclosure and comfort, rather than other functions such as turn-taking. We did not gauge users' nonverbal responses as we could not control the users' smartphone in its native setting. Users answered a total of twenty four questions with 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 et al. 2015]. Since smartphones were treated as an icon of emotionally engaged communication [Kang et al. 2008], 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.

4.2 Participants and Procedure

A total of 95 participants (25% men, 75% women; average 38 years old) were randomly assigned to one of 5 conditions: video with nonverbal backchannels (N=20; 25% men, 75% women), video without nonverbal backchannels (N=19; 26% men, 74% women), animation with nonverbal backchannels (N=20; 30% men, 70% women), animation without nonverbal backchannels (N=18; 11% men, 89% women), and audio-only (N=18; 35% men, 65% women). The participants were given $10 compensation when they completed the study. Participation required a total of 35 minutes on an individual basis. Participants were first asked to fill out a pre-questionnaire. The pre-questionnaire included questions pertaining to users demographics. Participants were then asked to download and install our mobile application. Participants were then asked to answer 12 questions asked by a virtual human and fill out the 1st post-questionnaire. Participants were then asked if they would like to continue, and if so, answer up to 12 additional questions and fill out the 2nd post-questionnaire. 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.

4.3 Measurements

For subjective measures, in the first post-questionnaire, we utilized Social Attraction to measure users' feelings of attraction toward a virtual human. The measure consisted of 6 items whose examples included "I would like to have a friendly chat with a virtual human." We also measured Virtual Rapport to assess users' feelings of rapport with a virtual human. The measure consisted of 17 items whose examples included "I felt I had a connection 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 = .873, Virtual Rapport: Cronbach's alpha = .940). For objective measures, we analyzed users' feedback derived from their voice input. 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) negative reasons for quitting the conversation, and 3) the average number of words in answers (total number of words in each user's answers divided by the number of questions answered).

5 Results

5.1 Results for the subjective measures

We performed a Between-Subjects ANOVA. The results [$F(4, 90)=2.75$, $p=.033$] with a Tukey HSD Test demonstrate that users reported more social attraction to a 3D animated character without nonverbal backchannels ($M=3.91$, $SD=.98$) significantly more than a video character with nonverbal backchannels ($M=2.93$, $SD=1.03$), see Figure 4 (a). We found a trend that is similar to the results for social attraction in users' feeling of rapport, although the result is not statistically significant. Users reported their feeling of rapport more when they interacted with the 3D animated character without nonverbal backchannels, compared to interacting with the video character with nonverbal backchannels that is reported in Figure 4 (b). We did not find statistically significant difference for the additional question related to likelihood to converse with the virtual human in the future.

5.2 Results for the objective measures

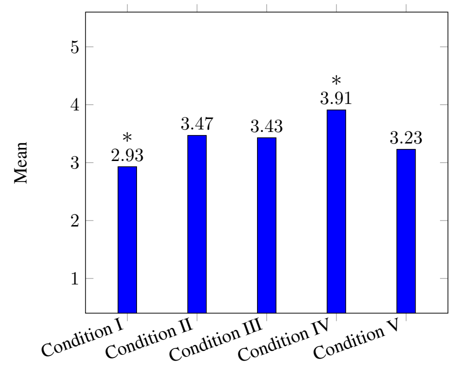
To measure the length of the conversation, we used the number of the additional questions that the user answered after the first 12 mandatory ones before stopping. We performed a Between-Subjects ANOVA with a Tukey HSD Test. Our results [$F(4, 90)=2.19$, $p=.076$] demonstrate that users tended to answer more questions when they interacted with a 3D animated character that demonstrated nonverbal backchannels ($M=23.35$, $SD=2.68$), compared to interacting with a video character that did not exhibit nonverbal backchannels ($M=18.58$, $SD=5.98$). A Tukey HSD Test shows the difference is statistically significant. Results are reported in Figure 4, (c). We also examined 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: "Feel a little uncomfortable talking to him," "I did not really want to interact anymore," and "I am 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 4, (d). The results show that users were more likely to cite negative reasons for not completing all 24 questions when interacting with the video character that did not present nonverbal backchannels while not citing any negative reasons at all when interacting with the 3D animated character that displayed nonverbal backchannels. There was no statistical significance for the average number of words in answers.

6 Discussion and Future Work

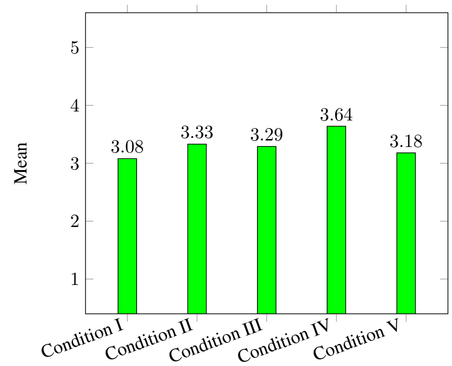
This study successfully utilized a virtual human's fidelity when presented on smartphone devices to explore its effect on users responses. We found that users were more attracted to a 3D character in social situations when they did not present backchannel behaviors, compared to a video character that presented backchannels. We also discovered in general that users tended to feel greater rapport with a 3D animated character that did not display backchannels than a video character that presented backchannels. Furthermore, users communicated for a longer period of time with a 3D animated character that displayed backchannels, compared to interacting with a video character that did not present any backchannels. The results also demonstrated a trend that users did not cite any negative reasons for quitting the communication when they interacted with a 3D animated character with backchannels whereas they cited the highest number of negative reasons when they interacted with a video character without any backchannels.

In our study, the nonverbal backchannels were realistic, but less synchronized to the intended message of the speaker since speech recognition technology was not used in this interaction. The backchanel behavior that our characters exhibit are triggered during the listening phase of the dialogue turn. In other words, while the user is speaking, the character will nod or otherwise respond nonverbally with a fixed set of responses, Our system does not detect the presence of a speech signal, or respond incrementally to dialogue [DeVault et al. 2009]. Rather, the backchannel behaviors go through a fixed set of responses, without consideration for the specific content of a user during their speaking turn. The backchannel response is triggered when the user assumes the speaking turn by pressing the 'Click and hold to speak' button. The interactive video graph approach has limitations to interactivity that do not affect the traditional 3D approach. A video-based character could not easily be designed to dynamically respond to a user often and quickly as could be done with an interactive 3D character, as frequently swapping out 2D videos would lead to an exaggerated popping or jumping effect. It is notable that nonverbal backchannels are important to socialization and communication, but only when they are appropriate and synchronized to the intended message of the speaker [Gratch et al. 2007; Wang and Gratch 2010]. However, our findings demonstrate that people are more inclined toward 3D animated characters in social interactions regardless of exhibiting backchannels than video characters presenting less synchronized backchannel behaviors. [MacDorman et al. 2009] notes that slight imperfections in realistic characters could yield uncanny valley effects, and people might evaluate the realistic characters more negatively than cartoonish characters. In the same vein, we argue people would feel the uncanny valley more with a video character than a 3D character if the video character do not convey synchronized backchannels. We contend that people grant more leniency with respect to behavioral anomalies in the 3D animated character as opposed to the video character. We further argue that people expect a real human video to exhibit synchronized backchannels as they might perceive the video character like a real human.

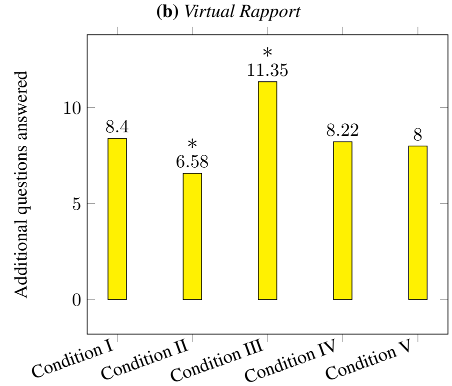
Our results further demonstrate a trend toward people responding to video or 3D animated characters and faceless voice-only characters alike regarding users' rapport and engagement time with the characters. This repeats the findings of [Kang et al. 2015] that proposed that the nature of smartphone use leaves little room for the user to focus their visual attention on the smartphone screen by do-



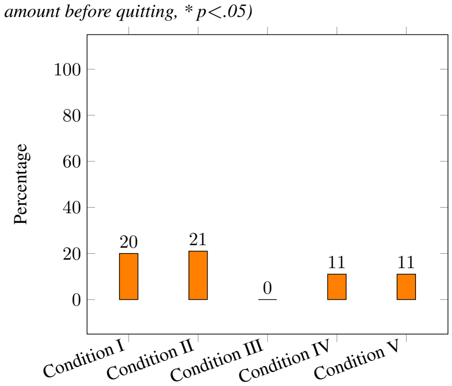
(a) *Social Attraction, * p<.05)*



(b) *Virtual Rapport*



(c) *Number of questions answered beyond the required amount before quitting, * p<.05)*



(d) *Negative Quit Reasons*

Figure 4: *Results for subjective (a and b) and objective (c and d) measures in each of 4 conditions: (Condition I: video with backchannels, Condition 2: video without backchannels, Condition III: 3D with backchannels, Condition IV: 3D without backchannels, Condition V: audio-only.*

ing multiple tasks concurrently. 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. This finding needs further research on attention of mobile users. We could implement eye tracking in future rounds of this study to address this lack of rich information which could help inform just how important the fidelity of a virtual human is to use.

Based on the results of our current study, it is hard to confirm what form of a virtual character would elicit smartphone users' engagement in communication with a virtual character more. Our long term research goals include determining the effectiveness of using interactive video graph-based characters against using entirely 3D-based characters. While the final appearance of the two methods can be differentiated to the user as a matter of photorealism, the method of generating such content is vastly different. A finding that one method is more or less effective than another would be of interest to those developing such user interfaces and applications. Likewise, a finding that they are both equally or similarly effective would also be of interest, making them interchangeable as a means of interaction. In our study, we did not find definitive answers to those questions, and for future work we intend to perform larger study that can better differentiate the effectiveness of such practices.

In our future work, we will test these results with both a male and a female character to further examine the role of gender effect on users' responses. We will also explore synchronized backchannel behaviors of virtual characters with users' behaviors in various social contexts by implementing speech processing systems for greater fidelity and communicative coherence in the interactions.

7 Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant (No. R0184-15-1030, MR Avatar World Service and Platform Development using Structured Light Sensor) funded by the Korea government (MSIP).

References

DEVAULT, D., SAGAE, K., AND TRAUM, D. 2009. Can i finish?: Learning when to respond to incremental interpretation results in interactive dialogue. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Association for Computational Linguistics, Stroudsburg, PA, USA, SIGDIAL '09, 11–20.

FENG, A. W., LEUSKI, A., MARSELLA, S., CASAS, D., KANG, S.-H., AND SHAPIRO, A. 2015. A platform for building mobile virtual humans. In *Intelligent Virtual Agents*. Springer International Publishing, 310–319.

GRATCH, J., WANG, N., GERTEN, J., FAST, E., AND DUFFY, R. 2007. Creating rapport with virtual agents. In *Intelligent Virtual Agents*, Springer, 125–138.

HODGINS, J., JÖRG, S., O'SULLIVAN, C., PARK, S. I., AND MAHLER, M. 2010. The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP) 7*, 4, 22.

KANG, S.-H., WATT, J. H., AND ALA, S. K. 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.

KANG, S.-H., FENG, A. W., LEUSKI, A., CASAS, D., AND SHAPIRO, A. 2015. The effect of an animated virtual character on mobile chat interactions. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, ACM, 105–112.

KOKKINARA, E., AND MCDONNELL, R. 2015. Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, ACM, 221–226.

MACDORMAN, K. F., GREEN, R. D., HO, C.-C., AND KOCH, C. T. 2009. Too real for comfort? uncanny responses to computer generated faces. *Computers in human behavior 25*, 3, 695–710.

MARINELLI, D., AND STEVENS, S. 1998. Synthetic interviews: the art of creating a'dyad'between humans and machine-based characters. In *Interactive Voice Technology for Telecommunications Applications*.

MARSELLA, S., XU, Y., LHOMMET, M., FENG, A., SCHERER, S., AND SHAPIRO, A. 2013. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, 25–35.

MCDONNELL, R., BREIDT, M., AND BÜLTHOFF, H. H. 2012. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG) 31*, 4, 91.

MCGLOIN, R., NOWAK, K. L., STIFFANO, S. C., AND FLYNN, G. M. 2009. The effect of avatar perception on attributions of source and text credibility. In *Proceedings of ISPR 2009 The International Society for Presence Research Annual Conference. Philadelphia: Temple University Press*.

MORI, M., MACDORMAN, K. F., AND KAGEKI, N. 2012. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE 19*, 2, 98–100.

OCHMAN, B., 2004. Burger king has fun with subservient chicken viral campaign.

RIZZO, A., LANGE, B., BUCKWALTER, J. G., FORBELL, E., KIM, J., SAGAE, K., WILLIAMS, J., DIFEDE, J., ROTHBAUM, B. O., REGER, G., ET AL. 2011. Simcoach: An intelligent virtual human system for providing healthcare information and support. *International Journal on Disability and Human Development 10*, 4, 277–281.

SLOSS, E., AND WATZMAN, A., 2005. Carnegie mellons entertainment technology center conjures up benjamin franklins ghost. press release.

TRAUM, D., GEORGILA, K., ARTSTEIN, R., AND LEUSKI, A. 2015. Evaluating spoken dialogue processing for time-offset interaction. In *Proceedings of SIGDIAL*, 199–208.

TRAUM, D., JONES, A., HAYS, K., MAIO, H., ALEXANDER, O., ARTSTEIN, R., DEBEVEC, P., GAINER, A., GEORGILA, K., HAASE, K., ET AL. 2015. New dimensions in testimony: Digitally preserving a holocaust survivors interactive storytelling. In *Interactive Storytelling*. Springer, 269–281.

WANG, N., AND GRATCH, J. 2010. Don't just stare at me! In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1241–1250.