Investigating the Perceived Realism of the Other User's Look-Alike Avatars

Aisha Frampton-Clerk*
aishafclerk@gmail.com
Queensborough Community College
City University of New York, NY, USA

Oyewole Oyekoya
oyewole.oyekoya@hunter.cuny.edu
Hunter College
City University of New York, NY, USA

ABSTRACT

There are outstanding questions regarding the fidelity of realistic look-alike avatars that show that there is still substantial development to be done, especially as the virtual world plays a more vital role in our education, work and recreation. The use of look-alike avatars could completely change how we interact virtually. This paper investigates which features of other people's look-alike avatars influence our perceived realism. Four levels of avatar representations were assessed in this pilot study: a static avatar, a static avatar with lip sync corresponding to an audio recording, full face animation with audio and a full body animation. Results show that full-face and body animations are very important in increasing the perceived realism of avatars. More importantly, participants found the lip sync animation more unsettling (uncanny valley effect) than any of the other animations. The results have implications for the perception of other people's look-alike avatars in collaborative virtual environments.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; • **Computing methodologies** → *Perception*.

KEYWORDS

Look-alike Avatar, Virtual Character, Telepresence, Uncanny Valley

ACM Reference Format:
Aisha Frampton-Clerk and Oyewole Oyekoya. 2022. Investigating the Perceived Realism of the Other User's Look-Alike Avatars. In *28th ACM Symposium on Virtual Reality Software and Technology (VRST '22), November 29-December 1, 2022, Tsukuba, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3562939.3565636

1 INTRODUCTION

The use of avatars have become increasingly prevalent across a broad range of disciplines. Beyond gaming and film into social media, the use of avatars is becoming more mainstream. Assessing how true to life these avatars are perceived is the next step in making these immersive experiences more realistic for users. The technology that enables the creation of avatars that look and behave like the user (look-alike avatars) is becoming increasingly available. Look-alike avatars can be defined as digital models of people that look like the users they represent.

It has been hypothesized that the uncanny valley effect [13] will have adverse effects on user experience, changing the perception of the avatar and negatively affecting overall immersion. The uncanny valley theory suggests that as the appearance of a character gets closer in resemblance to traditional human features, its perception would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance. In this paper, the look-alike avatars likely fall within the uncanny valley as they are approaching but has not attained exact appearance. However, it is unclear which avatar features cause this unfamiliarity gap (uncanny valley effect) and why the familiarity suddenly decreases. This paper investigates this phenomenon further through the addition of lip syncing, full facial expressions and full body movement with the aim of identifying the features that add to the realism of avatars, as perceived by the other user. A recent study [14] has shown that users prefer the other user to be represented in a photorealistic full-body human avatar in both augmented reality and virtual reality due to its human-like representation.

This paper presents a pilot study that explores the perceived realism and uncanniness of realistic individualized avatars. The novelty of this approach is that participants judge the realism of the look-alike avatar of another person, not their own. This important research area remains unexplored, despite the increasingly collaborative and social uses of virtual environments and avatars.

2 RELATED WORK

Despite the advancement in technology, generating human-like virtual characters (avatars) is still a very challenging problem, as the virtual character has to look like and behave as the user it represents. Even if a virtual character is generated with exceptional likeness, believable motions and emotions can be difficult to generate. As such, the resulting characters tend to be close in appearance but does not faithfully represent the user. De Borst and de Gelder [4] argue that contrary to the uncanny valley hypothesis, it is not the most realistic looking virtual characters that evoke an eerie feeling, but rather those on the border between non-human and human categories, especially if they are combined with human-like motion. The paper also presents an in-depth discussion on the interaction between movement and appearance of a human-like stimulus.

There is evidence that social Virtual Reality (VR) users tend to construct their avatar's appearance consistently with their offline identity, even when they sometimes construct application-specific self-presentation [7]. The appearance and recognition of a digital self-avatar may be especially important for telepresence and social

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
VRST '22, November 29-December 1, 2022, Tsukuba, Japan
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9889-3/22/11…$15.00
https://doi.org/10.1145/3562939.3565636

VR [14] and is the next stage in immersing ourselves in the virtual world.

Gaming applications are increasingly utilizing look-alike avatars because using avatars which are similar to the user in appearance and action is believed to provide users with a greater sense of reality and presence [16]. This presence and its effects may be mediated through self-identification and self-visualization [2, 10]. Increasingly, technology allows for exact features of a player to be represented, including characteristics such as facial features and specific movement characteristics.

Imagining oneself achieving certain activities has been found to improve performance in a variety of ways [5, 6]. Furthermore, visualizing the successful completion of a task primes the brain to actually perform the task. Simply watching another person engage in a task can activate a neurological action plan for the observed task [3, 11]. Studies have revealed that while simply observing an action being performed triggers an action plan for that activity, a simulation of the 'self' doing the task engages distinct neural structures surrounding the task [9, 15]. A 2012 study showed that by creating an avatar of an ideal self, subjects were better able to visualize and maintain an image of their ideal self [12] and other studies indicate that identification with an avatar can increase self-confidence and exercise behaviors, prompting the conclusion by the authors that virtual look-alike avatars can effectively instigate positive behavioral changes [1, 5]. Another study found that subjects behave differently toward and are more comfortable with an avatar whose appearance matches their own than toward one whose appearance is different from their own [1].

As noted above, several studies have investigated the effect of visualizing oneself in virtual environments. Additionally, studies on the uncanny valley have highlighted how people perceive the realism of robots and virtual characters. In prior studies, the robots and virtual characters are not necessarily modeled and evaluated on the likeness of the other user. As such, it is not clear how people perceive another user's look-alike avatar, which is addressed by this pilot study.

3 METHODS

3.1 Participants

A total of twenty-four participants completed the anonymous survey. Participants were recruited through emails to students, staffs and faculty at various institutions.

3.2 Materials

The virtual characters (avatars) used for this study were designed with the Reallusion Character Creator v3.44 software, using the Headshot plugin v1.11 to generate the avatar that looks like the users. Two subjects volunteered to have their look-alike avatar representations generated. The avatars were then transferred to iClone v8.02 software to generate the avatar animations.

This pilot study aimed to examine the more minute features that generate the uncanny valley effect in the resulting look-alike avatar. These features include the appearance of the avatar, lip sync, full facial expression and full body animation. Isolating these features and assessing participants' responses to each make it possible to separate how the reaction to each feature may differ. This was investigated through an anonymous survey format. Participants were exposed to videos of the same look-alike avatar with these features represented on the avatars in four ways: (i) a static rendering of the look-alike avatar without audio; (ii) the static avatar with lip sync corresponding to an audio recording; (ii) the static avatar with facial expressions as well as lip sync; and (iv) the avatar with full body movement including lip sync and facial expressions.

The static avatar representation served as a baseline for judging the appearance and similarity of the avatar to the user in the image without audio or tracking. The avatar representation with the lip sync had no eye or facial feature tracking. It simply had lip movement that matched the audio as accurately as possible using the Acculips lip sync feature of the iClone 8 software. The Acculips plugin is an automated program that creates lip movement for the virtual character from an uploaded audio file. The avatar representation with full facial expression consisted of face, eye, head and lip tracking. The LIVE FACE Profile for Motion LIVE Plugin v 1.0 was used to generate and stream the facial expressions, head and eye movement onto the avatars using the LIVE FACE app on the iPhone SE to capture real-time face data directly from the face of the users. The Acculips feature was used for lip movement. It mapped feature points on the faces of the two subjects that volunteered to have their look-alike avatar generated. The resulting avatars followed their human movement. The avatar representation with the full body consisted of face, eye, head and lip tracking, with audio and synthetic idle body animation. As such, the video displays the entire body of the avatar. The avatar uses the same audio introduction as in the previous videos.

3.3 Procedure and Measures

Video recordings of the avatar representations were generated and an anonymous online survey was distributed via Google Forms, which allowed for a diverse geographical distributed sample. A stationary image of the subject that is used to generate the look-alike avatar is presented in the corner of the screen (Figure 1), as a constant reminder of the original appearance of the subject.

Participants watched a video of the first subject's four avatar representations (see top of Figure 1). Following this, participants rated the level of their perceived realism of each avatar representation on a 4-point Likert scale: (1) most realistic; (2) somewhat realistic; (3) somewhat unrealistic; and (4) least realistic. This is a forced Likert scale that forces the participants to form an opinion, as there is no neutral option. Participants are also asked an additional question on which of the four avatar representations they find to be the most unsettling. The same procedure is repeated for the second subject's avatar representations. (see bottom of Figure 1).

Providing the opportunity to expose the look-alike avatars of two subjects with varying appearances to the participants. The survey aimed to gain their initial reactions to the virtual characters as soon as they have finished watching the videos of the avatar representations. The survey and the videos are presented in one single form, enabling the participants to re-watch videos, compare representations and change their opinions as needed before submitting their final responses. These responses could not be changed after submission.

Figure 1: Avatar Representations of Virtual Characters A (top) and B (bottom) - (1) Static Avatar; (2) Lip Sync; (3) Full Facial Expression; (4) Full Body Animation

4 RESULTS

A non-parametric Friedman test of repeated measures was conducted to test between the responses for the avatar representations. There was a statistically significant difference in participants' perceived realism of the avatar representations ($\chi^2(3) = 31, p < 0.001$) with a median rank of 2 (static avatar), 3 (lip sync), 2 (full facial expression) and 2 (full body animation). Dunn-Bonferroni post hoc tests found statistically significant differences between the lip synced avatar and each of the other avatar representations ($p < 0.01$).

Non-parametric chi-square goodness of fit tests were used to assess the frequencies of participants' choices for the most unsettling avatar representation. There was a statistically significant difference in participants' choices of the most unsettling avatar ($\chi^2(3) = 43.833, p < 0.001$) with a frequency of 2 (static avatar), 29 (lip sync), 16 (full facial expression) and 1 (full body animation). In addition, there was no statistically significant association between the choices selected for the virtual characters and the avatar representations ($\chi^2(3) = 4.112, p = 0.250$) i.e. the choices selected for the most unsettling avatar representation was similar for both virtual characters, A and B.

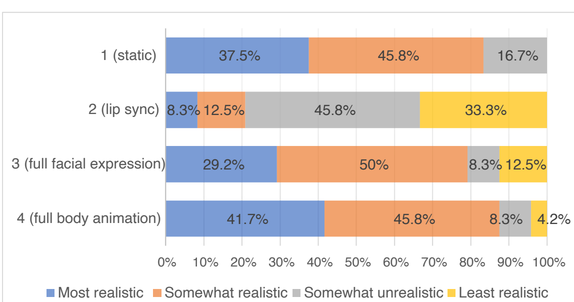

Figure 2: Percentage of participants' perceived level of realism for Character A

Figures 2 and 3 shows the distribution of the percentage of participants' perceived realism of the four avatar representations for the two look-alike avatars (Characters A and B). For Character A, 83% of participants stated that the static avatar representation was either most or somewhat realistic. Similar results are seen for character B where 79.2% of participants stated that the static avatar

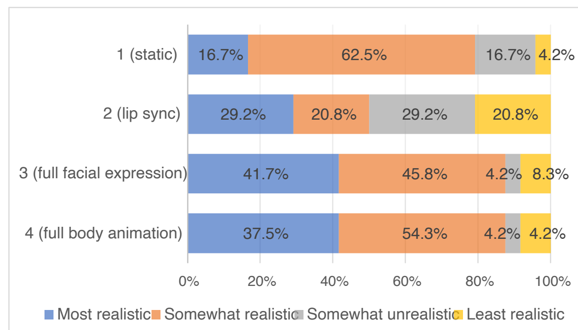

Figure 3: Percentage of participants' perceived level of realism for Character B

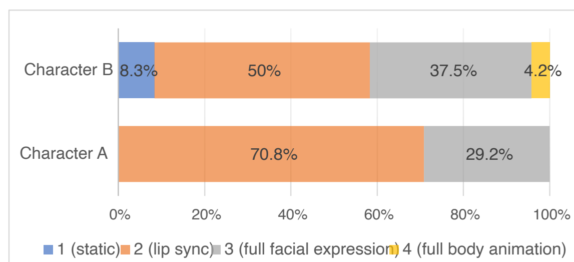

Figure 4: Percentage of participants' choices of most unsettling avatar representation

representation was either most or somewhat realistic. For both characters, the full body avatar representation was perceived as either the most or somewhat realistic by over 85% of participants. However, the lip synced avatar representation of Character A, only 20.8% of participants believed that this avatar was the most or somewhat realistic. Slightly more participants believed that character B's lip synced avatar was realistic. Here 50% of participants rated it the most or somewhat realistic.

This same sentiment about how different avatar representations affects perceived realism is seen when participants are asked to specify the avatar representation they find most unsettling (Figure 4. 70% of participants found the lip synced avatar representation to be the most unsettling and only 30% found the full facial expression avatar representation to be the most unsettling. Similar Results are seen for Character B, where 50% of participants found the lip synced avatar representation to be the most unsettling. A further 37% of participant specified that the full facial expression avatar representation was the most unsettling.

5 DISCUSSIONS

Prior studies have mainly focused on how users want to be represented as avatars [7]. In an increasingly collaborative and social virtual world, this paper focuses on how users perceive other people's look-alike avatar. Next-generation look-alike avatars should take into consideration, not just the self-perception of the user but also the perception of the person viewing and interacting with the user's avatar.

Four avatar representations were evaluated and each representation builds upon the last. The static avatar representation served as a baseline for judging the appearance and similarity of the avatar to the user in the image. Lip sync is added next, followed by the addition of facial expressions and eye movements and finally a full body representation. Aside from the static avatar representation, which focused on appearance only, the other three representations utilised the same audio so that the focal point of the evaluation is the perception of the added features. The original pictures of the subjects also remains a constant, as participants evaluate the videos so they can compare the perceived level of realism of the avatar representations. Separating the different features in this way enables us to investigate discrepancies between how each avatar representation is perceived. Changing slight features from one avatar to the next gives the opportunity to see if small changes can impact how the virtual characters are viewed by another user.

When participants were asked to rank the perceived realism of characters A and B, the distinction between the avatars with more pronounced expressions and animations was clear. The static avatar and the full body avatar representations were often perceived as most or somewhat realistic. These findings show that several factors contribute to the uncanny valley effect and make virtual characters appear realistic. Facial expression is one of the most important features and has shown to have a major effect on perceived realism. Our eyes are first drawn to people's faces, and especially their eyes, so even the slightest alterations are detected. This explains why the lip movement without any facial expression or eye movement was perceived as the most unsettling avatar representation. As more features were added, participants found the representations less unsettling. The full body avatar representations was found to be the least unsettling and most realistic. From this, it is possible to infer that the addition of features and movements is crucial in increasing the perceived realism of look-alike avatars in order to avoid the unfamiliarity gap (uncanny valley). Interestingly, the static avatar representation was generally perceived as realistic and not unsettling, which shows that the still avatar's appearances were likeable. This is in line with De Borst and de Gelder's argument [4] that the uncanny valley hypothesis, that focuses on behavioral effects, suggests that adding movement increases the familiarity for stimuli that were rated as likeable when still. That paper further argues that movement decreases familiarity even further for human-like images that were rated as unlikeable when still. In other words, the full body animation of look-alike avatars that are rated as likeable when static did result in an increase in their perceived realism, as we found.

5.1 Limitations and Future Work

While the mainstream use of virtual reality remains an exciting prospect, there are still outstanding questions about how to create the most immersive and realistic virtual characters. Future research will investigate the influence of other individual facial features (such as eyes, eyelids, chin, brows and emotions) on avatar realism as well as isolate and identify which other features play an important role in making the most realistic avatar.

Several limitations should be considered when generalizing the findings of this pilot study. First is the exploration of only two male avatars. Future work will focus on generating look-alike avatars with a broader variety of genders, races, ages and other demographics as this may influence perceived realism and yield a richer set of results. Another limitation is the sample size of twenty-four participants, given that this was an online survey. Future work will concentrate on obtaining a larger sample size and evaluating more than two look-alike avatars. The online survey did not obtain demographics information such as gender, age, race and prior experiences with using avatars. The lack of demographic information meant it was not possible to draw conclusions about why particular participants perceived avatars differently as the survey was completed anonymously. Having knowledge of other external influences that may affect participants' opinions of the look-alike avatars (such as prior exposure to avatars) would have allowed for more precise conclusions. Using the survey format also presents its own limitations, as this format may have led to selection bias. In future, this will be mitigated by comparing results with in-person study using VR systems. However, anonymity may have also helped with obtaining honest feedback. Another limitation was the way the characters were separated into only four avatar representations, three of which focused on the face and one focused on the full body. In future, we'll keep the point of view constant and investigate the effect on perceived realism (if any) of viewing the avatar's upper body versus the full body.

Another limitation of the anonymity is the lack of information on the participants' familiarity of the two subjects that was used to generate the look-alike avatars. It is not clear how many of the participants know or recognize the faces used in the study and this will likely affect processing [8]. Future studies will focus on investigating the effects of familiarity of the look-alike avatars being generated.

6 CONCLUSIONS

In this paper, we conducted a pilot study to investigate how people perceive the realism of other people's look-alike avatars and which features and motions contribute most heavily to the overall realism. The results show a clear correlation between increased perceived realism and the presence of full facial features and body movements. This suggests that the facial characteristics and body motion of a look-alike virtual character may be crucial features needed to avoid the uncanny valley.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Science Foundation, Research Experience for Undergraduates program (Award No. 2050532).

REFERENCES

[1] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (2006), 359–372.

[2] Elizabeth Behm-Morawitz. 2013. Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in human Behavior* 29, 1 (2013), 119–128.

[3] Beatriz Calvo-Merino, Daniel E Glaser, Julie Grèzes, Richard E Passingham, and Patrick Haggard. 2005. Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex* 15, 8 (2005), 1243–1249.

[4] Aline W. de Borst and Beatrice de Gelder. 2015. Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00576

[5] Jesse Fox and Jeremy N Bailenson. 2009. Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology* 12, 1 (2009), 1–25.

[6] Jesse Fox and Jeremy N Bailenson. 2009. Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology* 12, 1 (2009), 1–25.

[7] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.

[8] Mar Gonzalez-Franco, Anna I. Bellido, Kristopher J. Blom, Mel Slater, and Antoni Rodriguez-Fornells. 2016. The Neurological Traces of Look-Alike Avatars. *Frontiers in Human Neuroscience* 10 (2016). https://doi.org/10.3389/fnhum.2016.00392

[9] Julie Grezes and Jean Decety. 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human brain mapping* 12, 1 (2001), 1–19.

[10] Jillian Hamilton. 2009. Identifying with an avatar: a multidisciplinary perspective. In *Proceedings of the Cumulus 38 South 2009 Conference: Hemispheric Shifts Across Learning, Teaching and Research*. Swinburne University of Technology and RMIT University, 1–14.

[11] Marc Jeannerod. 1994. The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain sciences* 17, 2 (1994), 187–202.

[12] Youjeong Kim and S Shyam Sundar. 2012. Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior* 28, 1 (2012), 241–250.

[13] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.

[14] Minna Pakanen, Paula Alavesa, Niels van Berkel, Timo Koskela, and Timo Ojala. 2022. "Nice to see you virtually": Thoughtful design and evaluation of virtual avatar of the other user in AR and VR based telexistence systems. *Entertainment Computing* 40 (2022), 100457.

[15] Perrine Ruby and Jean Decety. 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature neuroscience* 4, 5 (2001), 546–550.

[16] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications* 29, 3 (2009), 76–84.

Links to the YouTube videos that participants of the study viewed.

https://www.youtube.com/watch?v=8pVJyPogZcQ&ab_channel=AishaFrampton-Clerk

https://www.youtube.com/watch?v=njViZf5UKXY&ab_channel=AishaFrampton-Clerk

Link to paper publication

https://dl.acm.org/doi/10.1145/3562939.3565636