Mapping and Recognition of Facial Expressions on Another Person's Look-Alike Avatar

Birate Sonia
University of Virginia
Virginia, USA
trj8ap@virginia.edu

Trinity Suma
Columbia University
New York, USA
trs2163@columbia.edu

Kwame Agyemang Baffour
Oyewole Oyekoya
City University of New York
New York, USA

(a) Actor A (Happy) (b) Actor B (Fear) (c) Actor C (Sad)

Figure 1: Avatar representations of the three actors

ABSTRACT

As Virtual Reality (VR) continues to advance and gain popularity, one of the persisting concerns revolves around the level of fidelity achievable in creating lifelike avatars. This study aims to explore the possibility, feasibility, and effects of controlling avatars in Virtual Reality using actual facial expressions and eye movements from three different actors mapped onto one actor's look-alike avatar. The objective is to explore the authenticity and appeal of avatars, particularly when they accurately portray facial expressions of their respective users compared to when they display facial expressions from other individuals. By properly mapping facial expressions and eye movements onto the avatar, we seek to aid with the development of a more realistic and captivating virtual experience that closely mirrors real-life interactions. Furthermore, we investigate whether mapping one's facial expressions to another person's look-alike avatar affects identification and recognition.

CCS CONCEPTS

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

KEYWORDS

look-alike avatar, virtual characters, facial expressions

ACM Reference Format:
Birate Sonia, Trinity Suma, Kwame Agyemang Baffour, and Oyewole Oyekoya. 2023. Mapping and Recognition of Facial Expressions on Another Person's Look-Alike Avatar. In *SIGGRAPH Asia 2023 Technical Communications (SA Technical Communications '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3610543.3626159

SA Technical Communications '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0314-0/23/12…$15.00
https://doi.org/10.1145/3610543.3626159

1 INTRODUCTION

Although technology has made significant progress, specifically in the Virtual Reality world, creating virtual characters (avatars) that resemble and behave like their users remains an exceedingly difficult task, considering the high level of similarity required between the character and the user it represents. It is often difficult to properly design an avatar that is totally pleasant, lifelike, and capable of eliciting emotions from its users. Skin tone, eye color, and hair texture, all have a large impact on an avatar's overall resemblance. Furthermore, other elements, such as the dynamic nature of human emotions and expressions, contribute to the challenge of capturing a very realistic look-alike avatar. Facial expressions are extremely complex and can indicate a wide range of emotions, including joy, excitement, sadness, and rage. To properly achieve this level of emotional realism, advanced scanning techniques, precise measurements, and complex algorithms are required. With all of these variables in play, it's no surprise that avatars aren't always complete and accurate representations of their users. Prior research on the mapping of face traits from one user to another has been limited. While the research in the related work section emphasize the potential benefits of virtual reality, they frequently don't touch on a possibly critical component of effectively capturing avatar facial expressions through facial mappings. As one of the many aspects in the production of an accurate representation, authenticity and plausibility are critical to generating a truly immersive and compelling virtual reality experience that closely resembles real-life interactions. With this study's focus on precise facial mapping techniques, we hope to improve the authenticity and engagement of virtual reality environments.

2 RELATED WORK

Frampton-Clerk and Oyekoya [Frampton-Clerk and Oyekoya 2022] investigated the features of other people's look-alike avatars and their influence on perceived realism. This study focused specifically on lip sync, full face animation with audio and full body animation. The authors found that results show a clear correlation between increased perceived realism and the presence of full facial features and body movements, and hence through this, there is a decrease of the uncanny valley effect which with the creation of highly realistic avatars, this becomes more of a worry. Seyama, and Nagayama [Seyama and Nagayama 2007] delved into the uncanny valley effect and measured observers' impressions of facial images whose degree of realism was manipulated by morphing between artificial and real human faces; the results were that the uncanny valley hypothesis was supported by the finding that highly realistic facial images evoked strong negative impressions. However, the uncanny valley phenomenon was only confirmed when morphed faces displayed abnormal features, such as peculiar eyes. These findings indicate that possessing an almost perfect realistic human appearance is necessary, but not enough, to trigger the uncanny valley. It is only when abnormal features are present that the uncanny valley effect emerges. Another study [Burleigh et al. 2013] on the uncanny valley effect provides the same conclusions, particularly that more realistic avatars caused discomfort between people, although this was only consistent when the faces had some unique or aberrant features. As a result, the findings imply that factors such as the category or context to which the face belongs influence people's negative sentiments.

Osman [Osman 2020] looked into the integration of tracked facial features for virtual reality users in virtual reality environments. The author investigated the rendering of virtual reality scenes using a head-mounted display that detects the user's gaze and captures images of their mouth to generate a virtual face. The virtual face, which includes eye and mouth movements, was then presented as an avatar in the virtual reality scene. The research centered on the technical issues and approaches for rendering virtual reality experiences with a head-mounted display (HMD).

While the focus of this research was primarily on avatar facial expressions, other studies have explored additional factors that contribute to user realism in virtual reality experiences. For instance, Ying et al. [Ying et al. 2022] examined the influence of telepresence and social presence on users' reactions to virtual reality commercials. Their findings revealed that higher levels of telepresence in these experiences enhanced users' cognitive processing and emotional connection. Another study by Freeman and Maloney [Freeman and Maloney 2021] delved into the presentation and perception of self in Social Virtual Reality. Their research highlighted the importance of avatars accurately representing users' identities and appearances, as this contributed to a greater sense of presence and authenticity in virtual social interactions. Furthermore, the visual quality of the virtual environment itself plays a significant role in creating a realistic experience. Studies, such as the work by Slater [Slater et al. 2009], have shown that incorporating advanced visual techniques like real-time recursive ray tracing can greatly enhance the realism and immersive nature of virtual environments when compared to simpler graphics lacking features such as shadows and reflections (raycasting). Other research, utilizes the use of avatars, androids, and virtual reality stimuli in neuroimaging research which allows for the exploration of human cognition and social processes in more ecologically valid contexts [de Borst and de Gelder 2015].

Suh et al. [Suh et al. 2011] theorized how users form attitudes and intentions regarding avatars in realistic, task-focused virtual worlds using a conceptual framework based on elf-congruity (the degree to which the avatar resembles its user) and functional congruity (the degree to which the avatar executes its intended functions). The study's findings imply that how much an avatar resembles the user (self-congruity) influences the user's identification with the avatar. This identification, in turn, leads to more favorable opinions toward the avatar and a stronger desire to utilize it. The above literature summarizes the importance of realism in virtual reality and indicate how an increase in the realism of avatar facial expressions through facial expression mapping can prove useful. Prior research emphasizes the possibility of four categories of users when building avatars, which are: realistics, ideals, fantasies, and role players [Neustaedter and Fedorovskaya 2009]. However, the findings of facial mapping research hold potential benefits for all four user categories, as facial mappings can be tested and applied to users across the spectrum of realistics, ideals, fantasies, and role players.

3 METHODS

3.1 Participants

A total of 24 participants completed the anonymous study. Participants were recruited through emails to students, staffs and faculty at various institutions.

3.2 Materials

The virtual character (avatar) utilized in this study was created with the Reallusion Character Creator v3.44 software, which includes the headshot plugin v1.11. This enabled us to create an avatar bearing an almost perfect resemblance to the actor. The process involved importing a high-resolution picture of the actor into the software, which facilitated the generation of an avatar identical to the actor. After the avatar was created with Reallusion, it was transferred to the iClone software v7.93 for facial expression mapping. The avatar expression mappings included face, eye, head, and lip tracking. The Motion LIVE Plugin v1.1, a package in the IClone software suite was used in collaboration with LIVE FACE Profile (downloaded from the Apple App Store), to easily develop and map realistic facial expressions, head movements, and eye animations to the avatars.

The Apple iPhone SE, on which the LIVE FACE app was downloaded and installed was used for capturing real-time facial data from users. The phone remained stationary in a specific spot during the capture process, resulting in a very immersive and almost lifelike experience. Three amateur actors volunteered to have their facial expressions captured and animated for the study. The principal actor, Actor A (Figure 1a) volunteered to have their look-alike avatar generated.

There was one 'faithful' avatar representation where facial expressions from Actor A were mapped onto their corresponding avatar using seven emotions (neutral, happy, sad, fear, anger, disgust, and surprise). Using these seven emotions, the same procedure was repeated for the two 'unfaithful' avatar representations, where the facial expressions of two other actors, Actors B and C (see Figures 1b and c respectively), who volunteered for the study, were projected onto the principal actor's look-alike avatar.

In order to gather data and assess the effectiveness of the avatar representations, a series of video recordings featuring the avatars and participants were meticulously generated. The avatar representations were recorded, and an anonymous online study was issued via Google Forms, allowing for a diverse geographically spread sample.

The form had seven sections, with each section containing video representations of the actors and their corresponding avatar representations. Each section consisted of a 3x3 multiple-choice grid with each avatar representation presented in the column and the actors presented in the rows. Participants were limited to one response per column and the row order was shuffled randomly to reduce any confounding influence of the order and sequence effects such as learning or fatigue. A total of 42 videos were generated for the study. Through each section, three videos featured the three participants expressing a specific expression, while the subsequent three videos showcased the faithful and unfaithful representations for the avatar corresponding to the same expression.

3.3 Procedures and Measures

The study's primary goal was to assess participants' ability to correctly match each avatar with its corresponding represented individual. Pre-experiment questions required users to provide their age and gender identification. Participants were able 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. The study employed forced answer sections where participants were required to match each individual with the corresponding avatar representation, so one individual could not be matched with more than one avatar representation. For example, if a participant matched Actor A with Avatar C, then only Avatars A and B can be selected for Actors B and C. The last section of the form included two important questions, one question asked about familiarity i.e. if participants knew any of the actors. This question attempted to gauge whether familiarity affects the matching accuracy. A final question asked for general feedback from the study.

4 RESULTS AND ANALYSIS

The study used a repeated measures (within-subjects) design with two independent variables: (i) actors (A, B and C); and (ii) facial expressions (neutral, happy, sad, fear, anger, disgust and surprise).

We treat actors and facial expression as categorical independent variables (factors), and accuracy as a dichotomous dependent variables and conducted a binomial logistic regression of the accuracy on the actors and expressions.

A logistic regression was performed (using SPSS Statistics 25 software [1]) to ascertain the effects of the actors and facial expressions on the likelihood that participants will accurately match the actors to their facial expressions. Binomial logistic regression enables us to estimate the probability of participants being able to accurately match the avatar's facial expression to the actor's real face. If the estimated probability is greater than or equal to 0.5 (better than even chance), it is likely that cases can be correctly classified (i.e., predicted) from the independent variables. The logistic regression model was statistically significant, ($\chi^2(8) = 32.728, p < 0.0001$) and the model correctly classified 66.5% of cases. In other words, participants were generally able to match an actor's facial expressions to the correct avatar. Furthermore, the data fit the model well based on the Hosmer-Lemeshow tests ($\chi^2(8) = 8.541, p = 0.382$).

We treat the neutral expression and Actor A as the reference categories for the independent variables, facial expression and actor respectively. For the facial expressions, statistical test based on Wald test results show that fear ($p < 0.0001$), happy ($p = 0.001$), sad ($p = 0.003$) and surprise ($p = 0.004$) added significantly to the model/prediction, but anger ($p = 0.402$) and disgust ($p = 0.402$) did not add significantly to the model. The odds of accurately matching an actor's facial expression to its corresponding avatar representation are greater for the fear (3.80 times), happy (3.29 times), sad (2.88 times) and surprise (2.69 times) as opposed to the neutral facial expressions. Anger, disgust and neutral expressions were more likely to be inaccurately matched than the other expressions, in line with Figure 2, which shows the percentage of facial expressions matched accurately to the actors.

For the actors, statistical test results show that neither Actor B ($p > 0.05$) nor Actor C ($p > 0.05$) added significantly to the model/prediction, but did not add significantly to the model. The odds of accurately matching any of the actors' facial expressions to its corresponding avatar representations are similar based on statistical tests i.e. there was no statistically significant difference in odds between all actors. In other words, facial expressions were similarly identifiable regardless of the actor whose facial expressions was mapped to the the look-alike avatar.

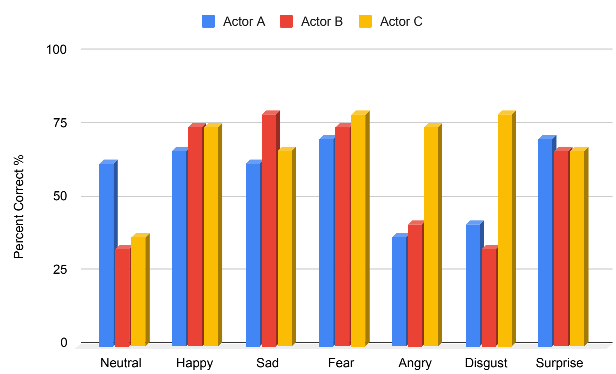

Figure 2: Percentage of expressions matched accurately

5 DISCUSSIONS

We used the neutral expression as the reference category for the facial expressions because it served as a baseline for measuring the matching accuracy for the other expressions, especially as the neutral involves no facial movement. As such, it was expected that the matching accuracy of the neutral expression will have the lowest values. Fear, happy, sad, and surprise were accurately and consistently matched. This suggests that there was enough variation in facial muscle cues of the three different actors that participants used as identifying information in matching facial expressions of the actors to their corresponding avatars. Like the neutral facial expression, anger and disgust were found to be more difficult to match than other facial expressions.

[1]https://www.ibm.com/spss

We used Actor A as the reference category because it served as a baseline for measuring the matching accuracy for the other actors, especially as the look-alike avatar was modelled based on the features of Actor A. The unfaithful representations (Actors B and C) had similar levels of matching accuracy as the faithful representation (Actor A). Figure 2 shows that Actors B and C had a low overall match ratio for neutral, with just about 30% of participants being able to fully identify the users. Actor C, the male actor, on the other hand, was well identified, with 70% of participants identifying the actor's neutral expression. Overall, neutral had the lowest matching accuracy and this would have been due to the fact that there were little to no facial expressions, as expected in the study.

It was a surprising finding that anger and disgust were just as difficult to match as neutral, however, a closer scrutiny of the study stimuli and the results may offer some explanation. We note that Actors A and B had a low overall matching accuracy for anger and disgust, but Actor C was highly identified in both categories. In both cases (anger and disgust), it was evident from the video that there were facial movements that enabled participants to identify Actor C more easily, as the actor may have over-emphasized the facial expressions. Overall, the findings do confirm that participants did find angry and disgust facial expressions harder to discern. Comparing Actor C's facial expressions to Actors A and B, they had less facial movement cues than Actor C. We argue that these unique facial movement cues are very important for avatar identification and recognition.

Comments and feedback from the participants and the actors provided further insights into these findings. Actor C had commented that he was not good at acting and may have overcompensated his depiction of the facial expressions. This was also evident to some participants. We argue that this should not be treated as a confounding influence on the results, as humans typically display a range of differing emotions in social situations anyway.

Participants also scrutinized the facial expressions, head and eye movements to assist in judging and matching the actors with their facial expressions. The more facially expressive, and the more facial muscles an individual seemed to express, the more cues the participants seemed to utilize. With Actor A being the faithful representation of the look-alike avatar, it is quite interesting that this wasn't the highest matching accuracy. These findings suggest that participants may be putting more weighting on facial expressions and movements when trying to identify and recognize the person behind the avatar. On the other hand, Actor C, the unfaithful male representation, was more easily recognized for the anger and disgust facial expressions, most likely because participants paid close attention to the actor's facial muscle cues and expressiveness. This highlights that visual fidelity alone might not be enough for accurate portrayal of the other user in social virtual environments.

5.1 Limitations and Future Work

Several limitations should be considered when generalizing the results of this study. Firstly, the principal look-alike avatar was of a specific gender and race. Therefore, it remains uncertain how the diverse gender and race representation of actors may have influenced the survey participant selection process. Future work will also investigate the role of tracked eye gaze movement[Oyekoya et al. 2009] on how look-alike avatars are perceived.

The software that was used to generate the look-alike avatar works by mapping a single picture of the actor to a preset 3D model of a virtual character allowing for the 3D artist to manually adjust head shape and other features. While this workflow is convenient, it produces a look-alike avatar that is similar but not an exact photorealistic look-alike avatar. Future work should focus on whether the results are similar for a photorealistic look-alike avatar.

Another limitation is the sample size of 24 participants could be higher given that the study was conducted online. Future work will aim to run this study on online platform (like MTurk). Additionally, only two participants had never met at least one of the actors. As majority of the participants knew at least one actor, it is possible that familiarity may have assisted with matching. Finally, all actors used in this study were amateurs so it is not known whether this results will apply to professional actors.

6 CONCLUSION

In this study, we explored the ability of participants to accurately match seven facial expressions of three actors to their corresponding avatar representations. The look-alike avatar of Actor A was used as the baseline avatar onto which all actors' facial expressions were mapped. In effect, participants were judging whether they could identify a person based on their facial expressions. Taken together, these results suggest that in order to achieve better emotion recognition and evoke emotions in users interacting with avatars in a virtual environments, a high level of visual detail that includes facial muscle cues and strong facial expressiveness are essential in avoiding the uncanniness of look-alike virtual characters.

ACKNOWLEDGMENTS

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

REFERENCES

Tyler J Burleigh, Jordan R Schoenherr, and Guy L Lacroix. 2013. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in human behavior* 29, 3 (2013), 759–771.

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), 576.

Aisha Frampton-Clerk and Oyewole Oyekoya. 2022. Investigating the Perceived Realism of the Other User's Look-Alike Avatars. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. 1–5.

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.

Carman Neustaedter and Elena A Fedorovskaya. 2009. Presenting identity in a virtual world through avatar appearances.. In *Graphics Interface*. 183–190.

Steven Osman. 2020. Integration of tracked facial features for VR users in virtual reality environments. US Patent 10,636,217.

Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2009. A saliency-based method of simulating visual attention in virtual scenes. In *Proceedings of the 16th ACM symposium on virtual reality software and technology*. 199–206.

Jun'ichiro Seyama and Ruth S Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16, 4 (2007), 337–351.

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.

Kil-Soo Suh, Hongki Kim, and Eung Kyo Suh. 2011. What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIs Quarterly* (2011), 711–729.

Tianyu Ying, Jingyi Tang, Shun Ye, Xiaoyuan Tan, and Wei Wei. 2022. Virtual reality in destination marketing: telepresence, social presence, and tourists' visit intentions. *Journal of Travel Research* 61, 8 (2022), 1738–1756.