Generating Look-Alike Avatars: Perception of Head Shape, Texture Fidelity and Head Orientation of Another User's Look-Alike Avatar

Kwame Agyemang Baffour *
CUNY Graduate Center, University of Energy & Natural Resources

Oyewole Oyekoya†
Hunter College, City University of New York

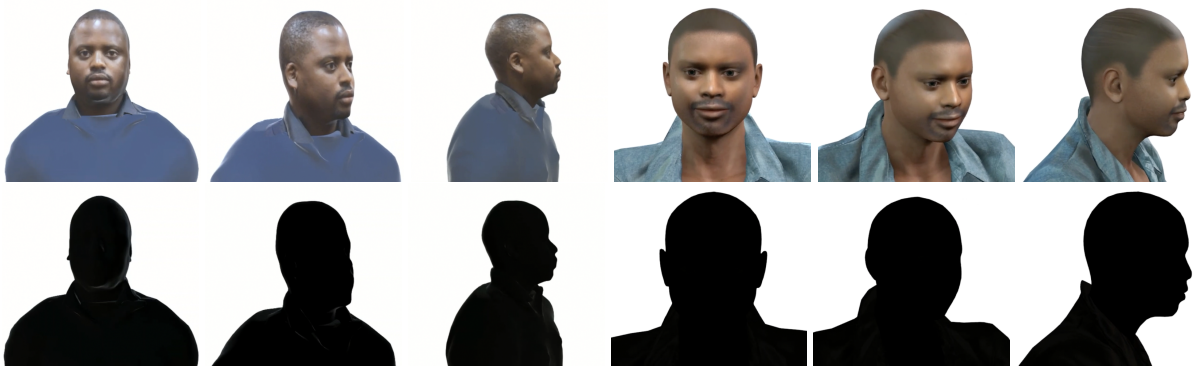

Figure 1: Avatar representations showing the textured (top row) and silhouette (bottom row) renderings for different head orientations (0°, 45°and 90°orientations). The left three columns show the photogrammetry-generated avatars and the right three columns show the single-image reconstructed avatars.

Abstract

Using photorealistic look-alike avatars may enhance the likeability and realism of avatars in collaborative virtual environments. This research seeks to determine the influence of head shape, texture fidelity and head orientation of a look-alike avatar on perception of likeability and visual realism, especially when judged by other people. Two textured look-alike avatars were generated using: (i) three-dimensional (3D) stereophotogrammetry; and (ii) 3D face reconstruction from a single full-face image. Participants compared three different head orientations (0°, 45°, 90°) of the look-alike avatars' textured heads to their corresponding head silhouettes, to emphasize the differences in head shapes. Results suggest that participants prefer geometrically accurate photorealistic avatars of other people due to the accuracy of the head shape and texture fidelity. Participants ranked the likeability and realism of the look-alike avatars similarly regardless of the head orientation.

Index Terms: Computing methodologies—Perception—;——Computing methodologies—Virtual reality

1 Introduction

Beat-em-up genre [4] Look-alike avatars are known to potentially have a substantial impact on the user's experience with immersive content and can also affect how users interact with one another in a shared virtual environment [2, 5].

A variety of factors, including the avatar's characteristics, have a role in the design or choice of an avatar. Molina et al [3] argued that making avatars look as humanly as possible by adding photo-realistic textures to these avatars, will enhance the plausibility of the avatars and virtual environments.

In this study, we conducted an empirical evaluation of visual realism and perception relating to the head shape, texture fidelity and head orientation of two look-alike avatars of the same individual generated using two different methods.

*e-mail: kbaffour@gradcenter.cuny.edu
†e-mail: oyewole.oyekoya@hunter.cuny.edu

2 Methods

A total of ninety-one (37 females, 52 males and 2 non-binaries) participants completed the anonymous online survey. The age ranges were: 18-24 (N=38); 25-34 (N=45); 35-44 (N=7) and 45-54 (N=1). Participants were recruited through emails to students, staff and faculty at various institutions.

Four avatar representations were used in this study. One look-alike avatar representation of the instructor was generated by scanning the head and upper body of the instructor using the 3dMDhead System [1]. The lower body was scanned and generated with a Structure Sensor [2]. The 3dMDhead system was chosen for the face because of the high render quality.The generated 3D human model was then automatically rigged using Adobe Mixamo [3]. The other look-alike avatar of the same instructor was designed with the Reallusion Character Creator v3.44 software, using the Headshot plugin v1.11. The avatars were then transferred to iClone v8.02 software for animation. The other two avatar representations were derived from darkening the textures of the above described avatars to create silhouettes.

Three videos each, of the four avatars were generated at three different head orientations; 0°(front), 45°(3/4) and 90°(side/profile).

The experiment used a 3-factor repeated measures (within-subjects) design with three independent variables: (i) avatar generation (photogrammetry and single-image); (ii) rendering mode (textured and silhouette); and (iii) head orientation (0°, 45°, 90°). Using a counterbalanced measures design, the avatar representations were mixed to reduce any confounding influence on the order and sequence effects such as learning effects or fatigue. Three dependent variables were measured: (i) avatar likeability and (ii) level of realism and (iii) avatar preference.

Video recordings of the avatar representations were generated and an anonymous online survey was distributed via Qualtrics, which allowed for a diverse geographically distributed sample. Stationary images of the subject upon which the look-alike avatar is based were presented in the corner of the screen.

Participants watched randomly selected videos of the four avatar

[1] https://3dmd.com/products/
[2] https://structure.io
[3] https://mixamo.com

representations after which they rated their likeability of the avatar representation on a 5-point Likert scale: (1) Dislike a great deal; (2) dislike somewhat; (3) neither like nor dislike; (4) like somewhat; (5) like a great deal. Participants were also asked to rate their perceived level of realism of the avatar's head shape to the reference image on a 5-point Likert scale: (1) least realistic; (2) somewhat unrealistic; (3) neutral; (4) somewhat realistic; (5) most realistic. Finally, participants were asked an additional question on which of the four avatar representations they found to be the most realistic based on the head shape.

3 Results

An ordinal logistic regression of the perceived likeability and realism was conducted on the independent variables (using SPSS Statistics 25 software [4]). We treat the photogrammetry and textured conditions as the reference categories for the independent variables, avatar generation and rendering mode respectively.

Likelihood-Ratio test showed statistically significant main effects of avatar generation ($\chi^2(1) = 20.050, p < 0.001$) and rendering mode ($\chi^2(1) = 26.322, p < 0.001$) on perceived likeability. There was no statistically significant main effect of head orientation ($\chi^2(1) = 0.014, p = 0.993$) on perceived likeability. Wald test shows that the independent variables, avatar generation ($p < 0.001$) and rendering mode ($p = 0.002$), were significant predictors of perceived likeability. The odds of the single-image generated avatar being perceived as more likeable was 0.430 (95% CI, 0.250 to 0.737) times that of the stereophotogrammetry-generated avatar, a statistically significant effect, Wald $\chi^2(1) = 9.413, p = 0.002$. In other words, participants were likely to perceive the stereophotogrammetry-generated avatar as more likable than the single-image generated avatar. The odds of the silhouetted rendering mode being perceived as more likeable was 0.394 (95% CI, 0.233 to 0.668) times that of the textured rendering mode, a statistically significant effect, Wald $\chi^2(1) = 11.985, p < 0.001$. In other words, participants were likely to perceive the textured rendering mode as more likeable than the silhouetted rendering mode.

Likelihood-Ratio test showed significant main effects of avatar generation ($\chi^2(1) = 58.955, p < 0.001$) and rendering mode ($\chi^2(1) = 39.393, p < 0.001$) for perceived realism. There was no significant main effect of head orientation ($\chi^2(2) = 5.265, p = 0.072$) on perceived realism. Wald test shows that the independent variables, avatar generation and rendering mode, were significant predictors ($p < 0.001$) of perceived realism. The odds of the single-image generated avatar being perceived as more realistic was 0.293 (95% CI, 0.173 to 0.497) times that of the stereophotogrammetry-generated avatar, a statistically significant effect, Wald $\chi^2(1) = 20.756, p < 0.001$. In other words, participants were more likely to perceive the stereophotogrammetry-generated avatar as more realistic than the single-image generated avatar. The odds of the silhouetted rendering mode being perceived as more realistic was 0.337 (95% CI, 0.200 to 0.566) times that of the textured rendering mode, a statistically significant effect, Wald $\chi^2(1) = 16.879, p < 0.001$. In other words, participants were more likely to perceive the textured rendering mode as more realistic than the silhouetted rendering mode.

Non-parametric Pearson's chi-square test was performed to assess the frequencies of participants' choices for the most preferred avatar across the two rendering modes. There was a statistically significant difference in participants' choices of the most preferred look-alike avatar ($\chi^2(1) = 23.674, p < 0.001$) with a frequency of 125 (stereophotogrammetry-generated avatar) and 59 (single-image generated avatar). There was a statistically significant association between the avatar generation and rendering mode ($\chi^2(1) = 13.198, p < 0.001$) i.e. the choices selected for the most preferred generated avatar significantly differed for both the textured and silhouetted rendering modes. Bonferroni-adjusted comparisons of proportions showed that there was a statistically significant difference ($p < .05$) in the proportion of participants who preferred the stereophotogrammetry-generated avatar ($N = 51$) and the single-image generated avatar ($N = 41$) in the silhouetted rendering mode. For the textured rendering mode, there was a statistically significant (higher) difference in the proportion of participants ($p < .05$) who preferred the stereophotogrammetry-generated avatar ($N = 74$) and the single-image generated avatar ($N = 18$).

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

4 Discussions and Conclusions

There was a clear preference for the look-alike avatar generated with 3D stereophotogrammetry over the single-image generated avatar in both the textured and silhouette conditions. To put simply, participants preferred the avatar of other people to look more closely like those users than an approximate estimation. In our study, stereophotogrammetry-generated avatar was perceived as more likeable than single-image generated avatar. The same finding applied to the silhouetted condition, where the shape of the avatar was more emphasized. It can also be seen that perceived likeability and realism of a look-alike avatar are directly influenced by how the avatar was generated and the rendering mode.

Interestingly, perceived likeability and realism were rated similarly for the three different head orientations. although prior research has shown that 90°(profile) and 0°(full-face) views result in poorer recognition performance than a 45°(3/4) view [6].

Majority of users found the textured avatar from the photogrammetry technique to be more realistic based on its head shape compared to the textured avatar from the single image technique. In the comparison of the silhouettes for both avatars however, the percentage of users that found the single-image avatar silhouette more realistic based on the head shape as against the photogrammetry avatar silhouette was not too distinguished.

Although participants clearly preferred the photogrammetry generated avatar over the single-image generated avatar due to the fact that the latter was not a very accurate representation of the individual, there was still a good level of likeability and perceived realism towards it. This makes a case that avatars generated using relatively inexpensive technologies may still be viable options for telepresence and collaborative virtual environments as indicated in prior studies [1].

Acknowledgments

This work was supported by a grant from PSC-CUNY (Award No. TRADB-51-248).

References

[1] J. N. Bailenson, A. C. Beall, J. Blascovich, and C. Rex. Examining virtual busts: Are photogrammetrically generated head models effective for person identification? *Presence*, 13(4):416–427, 2004.

[2] A. Frampton-Clerk and O. Oyekoya. 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*, pp. 1–2. ACM, 2022.

[3] E. Molina, A. R. Jerez, and N. P. Gómez. Avatars rendering and its effect on perceived realism in virtual reality. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 222–225. IEEE, 2020.

[4] D. Roeder. The adaptation of 3d video game production techniques to the 2d game play mechanics of the beat-em-up genre. 2012.

[5] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp. An 'in the wild'experiment on presence and embodiment using consumer virtual reality equipment. *IEEE transactions on visualization and computer graphics*, 22(4):1406–1414, 2016.

[6] N. F. Troje and H. H. Bülthoff. Face recognition under varying poses: The role of texture and shape. *Vision research*, 36(12):1761–1771, 1996.