Facial Emotion Recognition of Virtual Humans with Different Genders, Races, and Ages

Funda Durupinar
funda.durupinarbabur@umb.edu
Department of Computer Science
University of Massachusetts, Boston
Boston, MA, USA

JieHyun Kim
jiehyun.kim001@umb.edu
Department of Computer Science
University of Massachusetts, Boston
Boston, MA, USA

Figure 1: Virtual humans with facial expressions of varying emotions and intensities.

ABSTRACT

Research studies suggest that racial and gender stereotypes can influence emotion recognition accuracy both for adults and children. Stereotypical biases have severe consequences in social life but are especially critical in domains such as education and healthcare, where virtual humans have been extending their applications. In this work, we explore potential perceptual differences in the facial emotion recognition accuracy of virtual humans of different genders, races, and ages. We use realistic 3D models of male/female, Black/White, and child/adult characters. Using blendshapes and the Facial Action Coding System, we created videos of the models displaying facial expressions of six universal emotions with varying intensities. We ran an Amazon Mechanical Turk study to collect perceptual data. The results indicate statistically significant main effects of emotion type and intensity on emotion recognition accuracy. Although overall emotion recognition accuracy was similar across model race, gender, and age groups, there were some statistically significant effects across different groups for individual emotion types.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Computing methodologies → Procedural animation**.

KEYWORDS

facial expressions, emotion modeling, emotion recognition, virtual humans, perceptual bias

SAP '22, September 22–23, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-9455-0/22/09…$15.00
https://doi.org/10.1145/3548814.3551464

ACM Reference Format:
Funda Durupinar and JieHyun Kim. 2022. Facial Emotion Recognition of Virtual Humans with Different Genders, Races, and Ages. In *ACM Symposium on Applied Perception 2022 (SAP '22), September 22–23, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3548814.3551464

1 INTRODUCTION

Advances in immersive and collaborative technologies have expanded our social interactions, bringing virtual humans into influential roles. Multiple studies have shown that people show subjective feelings and similar behavioral and physiological reactions toward virtual humans as if they were real people [Bombari et al. 2015; de Borst and de Gelder 2015; Nass and Reeves 2003]. Even biases toward virtual humans imitate the real-world predispositions [Gamberini et al. 2015; Rossen et al. 2008; Wandner et al. 2014; Zipp et al. 2017]. Research has shown that racial and gender stereotypes can influence emotion judgments of others [Elfenbein and Ambady 2002; Halberstadt et al. 2020; Hess et al. 2004, 2010]. Stereotypical biases have severe implications in social life but are especially critical in domains such as education and healthcare, where virtual humans have been extending their applications. As the future of computer graphics entails increasingly diverse characters [Kim et al. 2021], identifying potential biases towards them can raise awareness, prevent amplification of existing inequalities, and encourage bias-mitigation strategies such as inter-group perspective-taking via VR embodiment [Beltran et al. 2022; Chen et al. 2021; Crone and Kallen 2022; Peck et al. 2013].

This paper explores the human perception of emotions displayed by virtual humans of different races (Black/White) and genders (Female/Male) for adult and child virtual characters. Facial expressions of six emotions (happiness, sadness, anger, fear, surprise, and disgust) have been commonly acknowledged as universal [Ekman and Cole 1972; Izard 1994]. Facial Action Coding System (FACS) is a standard for describing facial movement [Ekman et al. 2002], with well-defined parameters to encode these six emotions. 3D humanoid characters usually come with facial rigs that directly allow tuning Action Units (AU) specified in FACS. Therefore, differences in the perception of the six universal facial emotions likely result from factors other than their representation. As humans are particularly perceptive of faces and because of the precise encoding of facial expressions of six universal emotions, we focus on the faces of virtual humans to explore potential biases towards their characteristics. We address questions such as:

- Are there differences across race, gender, and age groups for overall emotion recognition accuracy?
- Are Black and male characters more likely to be incorrectly identified as expressing anger?
- Are female characters more likely to be incorrectly identified as expressing happiness?
- Do the participants' own races and genders affect their judgments?
- Do emotion intensities impact perceived emotions across groups?

To answer these questions, we conducted an Amazon Mechanical Turk (MTurk) study. We showed the participants video clips of each virtual human displaying facial expressions of six emotions with varying intensities and asked them to identify the emotion in each clip.

2 RELATED WORK

A vast amount of research has established the effect of facial characteristics on people's judgments about others' affective states and personalities [Carré et al. 2010; Hehman et al. 2015; Ormiston et al. 2017; Todorov et al. 2015]. Similarly, computer-generated faces and virtual characters are evaluated for their trustworthiness, dominance, aggressiveness, charisma, etc. based on their facial features and skin color [Araujo et al. 2021; Ferstl and McDonnell 2018; Ferstl et al. 2021; Hehman et al. 2015; Todorov et al. 2008].

Regarding emotion recognition, researchers have shown that race and gender can control classification accuracy [Halberstadt et al. 2020] and recognition speed [Hugenberg 2005]. For instance, Hugenberg presented White and Black computer-generated faces displaying happiness and anger to European American participants. The participants recognized happiness in White faces more quickly than in Black ones. In contrast, they focused on angry Black faces longer than on White ones. In a K-12 setting, Halberstadt et al. asked prospective teachers to identify the emotions in short video clips of Black and White children exhibiting facial expressions of the six universal emotions with varying intensity levels [Halberstadt et al. 2020]. They found that Black boys' emotions were more likely to be misinterpreted as angry than the emotions of White students. Inspired by this work, we followed a similar experimental setting and included child characters in our study. Differently, we collected judgments from a more general participant pool and used both adult and child characters.

Attribution of emotions to males and females also shows differences. For instance, women are stereotyped as more emotional and happy than men, whereas men are perceived to be angrier than women in general [Hess et al. 2004, 2010]. Hess et al. found women's expressions of happiness to be perceived as more intense and their expressions of anger and disgust to be less intense than expressions of the same intensity by men [Hess et al. 1997].

There is ample research on behavioral differences towards virtual humans of color. For instance, in an online survey, healthcare professionals rated male and African American virtual human patients as having higher pain and administered significantly more opioids than their demographic counterparts [Wandner et al. 2014]. In a virtual emergency training scenario, participants took more time to initiate help and made more errors while triaging agents with darker skin tones [Zipp et al. 2017]. Also in the healthcare domain, medical students' real-world skin-color biases were found to be correlated with their biases toward virtual humans [Rossen et al. 2008]. In an experimental setup involving a first-person shooter game, the African-American virtual agents were shot more often than the White ones [Correll et al. 2002]. Further evidence of shooter bias based on avatar race and socioeconomic status was found in a VR setting [Seitz et al. 2020]. Racial discrimination was also prevalent in an emergency scenario, where participants were expected to provide help to Black and White victims in a fire [Gamberini et al. 2015].

3 METHOD

3.1 Stimuli

We downloaded the four adult human models from Mixamo [Adobe 2022]. The male child models were purchased from TurboSquid [Shutterstock 2022], and the female child models were created by Reallusion Character Creator [Reallusion 2022]. All the models were selected and designed to have facial blendshapes as the state-of-the-art method for facial animation [Lewis et al. 2014]. We prepared facial expressions by tuning the blendshapes corresponding to the AUs in FACS. For each model, we animated six facial expressions in five levels ranging from the very beginning of an expression with *trace*-level intensity to *maximum* intensity [Ekman et al. 2002]. We created the animations and recorded the videos on Unity [Unity 2019], using its Universal Rendering Pipeline for realistic rendering. All the characters were displayed from the same distance and viewpoint, where the camera was angled straightly towards their faces. Each video clip was 1 second long. The Appendix shows images of the virtual characters captured at the apex of each intensity level per emotion expression.

3.2 Study Design

We performed an MTurk study to collect perceptual data. We presented the study description and a consent form at the beginning of the experiment. To prevent participant fatigue, we divided the study into 24 blocks, each consisting of 10 conditions. A block corresponds to a Human Intelligence Task (HIT) on MTurk. Participants were paid 0.1$ per HIT. We collected 36 data points per condition. Each condition consisted of the video of a virtual character displaying one of the six facial expressions corresponding to distinct emotions of happiness, sadness, anger, fear, surprise, and disgust. The participant was asked to identify the emotion in the video. In each video, a facial expression's strength varied in intensity, ranging from weak to full-strength on a scale of five. There were eight virtual characters for two races (Black/White), two genders (male/female), and two age groups (child/adult). Thus, there were a total of 240 videos (2 races $\times$ 2 genders $\times$ 2 age groups $\times$ 6 emotions $\times$ 5 intensity levels). Videos were displayed in random order.

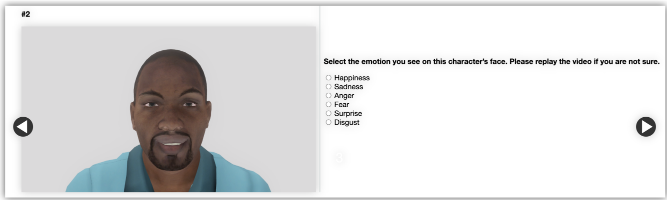


Figure 2: A screenshot from the MTurk user study.

Participants were allowed to participate in any number of blocks (HITs), but they could complete a particular HIT only once. Before the study, the participants were asked to provide optional demographic information, including their age, gender, and ethnicity. The University Institutional Review Board approved the experiment protocol. Figure 2 displays a sample question from the study.

3.3 Participants

We set participation qualifications as having an acceptance rate of $> 95\%$ and experience of more than 100 HITs. We restricted the participant locations to the United States of America in order to control for cultural differences. 208 (99F/109M) unique MTurk workers participated in the experiments. The average age was 38.79 and the ethnicity distribution was 72.6% White, 6.73% Black, 5.29% Hispanic/Latino, 8.17% Asian, and 4.81% Native American.

3.4 Results and Analysis

We first ran a multi-way, independent-subjects Analysis of Variance (ANOVA) to analyze the effect of the stimulus emotion type, emotion intensity, character race, gender, age, and participant gender on emotion recognition accuracy. Accuracy is the dependent variable, and the remaining factors are independent variables. We did not include participant race in the model due to the imbalanced number of participants across race groups. Only emotion type, gender, and intensity returned significant main effects. Interactions of more than two variables were not significant. Interaction of age with race yielded significant effects. Intensity had the highest effect factor ($F(1, 9407) = 8.08, p < 0.001$), followed by character race-age interaction ($F(1, 9407) = 5.12, p < 0.001$), stimulus emotion ($F(5, 9403) = 4.79, p < 0.001$), and gender ($F(1, 9407) = 4.72, p < 0.05$).

For post-hoc analysis, we performed Tukey Honestly Significance Detection (HSD) test with an alpha value of 0.05 (Table 1). The post-hoc analysis yielded no significant difference between female and male characters. There were significant differences between pairs of emotion types except for anger-surprise and fear-disgust. Differences between intensity levels were also statistically significant, except when intensities were more pronounced (above level 3) when overall accuracy increased. HSD didn't return any significant differences between races when we analyzed the accuracy of child and adult models separately for the interaction effects. Although ANOVA did not reject the null hypothesis for participant gender, HSD returned significantly higher emotion recognition accuracy ($p < 0.01$) for female participants than for males.

Table 1: Multiple Comparison of Means - Tukey HSD for accuracy comparison for stimulus emotions, intensity, and participant gender ($\alpha = 0.05$). Significant factors are highlighted in bold.

Emotion 1	Emotion 2	Mean Diff.	p
Hap	**Sad**	**-0.15**	**-0.0**
Hap	**Ang**	**-0.36**	**0.0**
Hap	**Fear**	**-0.49**	**0.0**
Hap	**Surp**	**-0.38**	**0.0**
Hap	**Dis**	**-0.53**	**0.0**
Sad	**Ang**	**-0.20**	**0.0**
Sad	**Fear**	**-0.34**	**0.0**
Sad	**Surp**	**-0.22**	**0.0**
Sad	**Dis**	**-0.37**	**0.0**
Ang	**Fear**	**-0.149**	**0.0**
Ang	Surp	-0.02	0.76
Ang	**Dis**	**-0.17**	**-0.0**
Fear	**Surp**	**0.12**	**0.0**
Fear	Dis	-0.03	0.30
Surp	**Dis**	**-0.15**	**0.0**
Intensity 1	**Intensity 2**	**Mean Diff.**	**p**
1	**2**	**0.17**	**0.0**
1	**3**	**0.27**	**0.0**
1	**4**	**0.32**	**0.0**
1	**5**	**0.31**	**0.0**
2	**3**	**0.10**	**0.0**
2	**4**	**0.15**	**0.0**
2	**5**	**0.14**	**0.0**
3	4	0.05	0.02
3	5	0.04	0.13
4	5	-0.01	0.97
P. Gender 1	**P. Gender 2**	**Mean Diff.**	**p**
Female	**Male**	**-0.0307**	**0.0025**

To understand the perception of each emotion type better, we display the confusion matrices per race, gender, and age group (Figure 3). The results indicate that happiness was the best-recognized emotion, followed by sadness. The lowest recognition accuracy was for disgust, as it was generally confused with anger.

For a deeper analysis of emotion type differences, we compared the emotion recognition biases across models of different genders, races, and ages by two-tailed independent-subjects t-tests with Bonferroni correction (Figure 4(a)). Bias for a given emotion type e refers to the erroneous attribution of a different emotion when the stimulus was e. Consistent with the literature, we found that happiness bias was significantly higher for females than males ($p < 0.001$). We didn't find any anger bias across genders or races. However, male characters were more likely to be misidentified as sad than female characters ($p < 0.05$), and children were more likely to be misidentified as angry than adult models ($p < 0.001$). We also found a higher disgust bias for White characters ($p < 0.05$).

We also computed the sensitivity (recall) values for each emotion (Figure 4(b)). Sensitivity differences were highly significant ($p < 0.001$) for sadness and disgust between races; sadness and anger between age groups. We observed slightly significant ($p < 0.05$) differences in disgust between genders and surprise between age groups. Recognition of sadness was more accurate in White and child characters than in Black and adult characters. Anger was more accurate for adult characters than children. Recognition of disgust

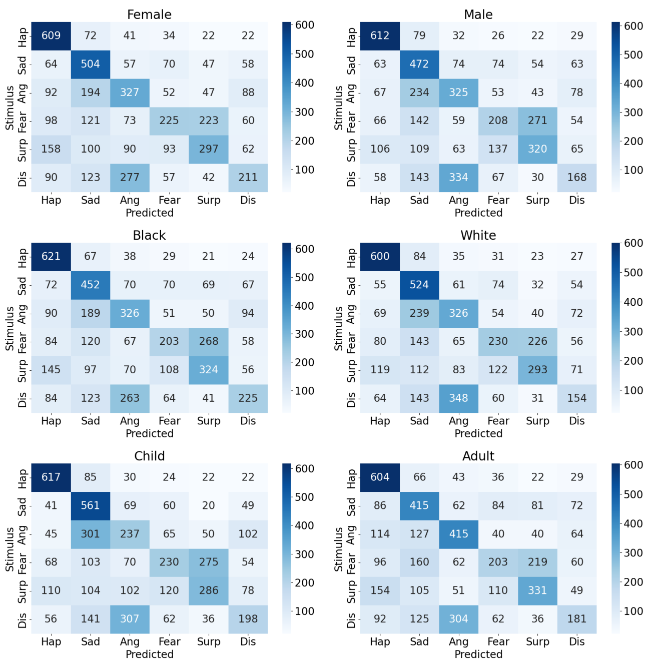
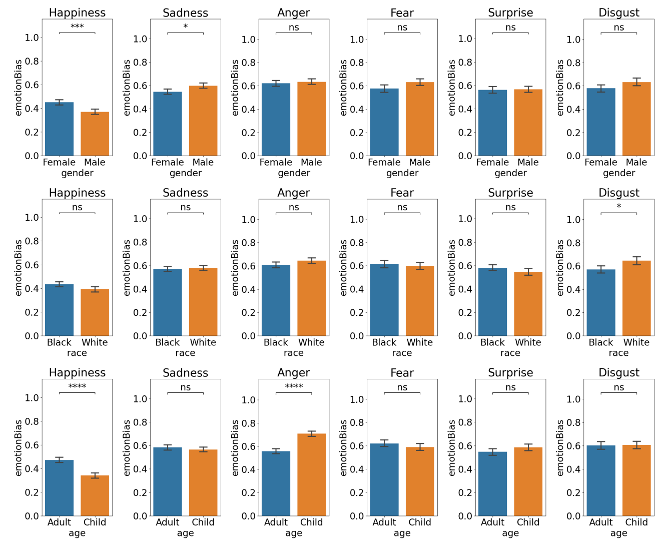
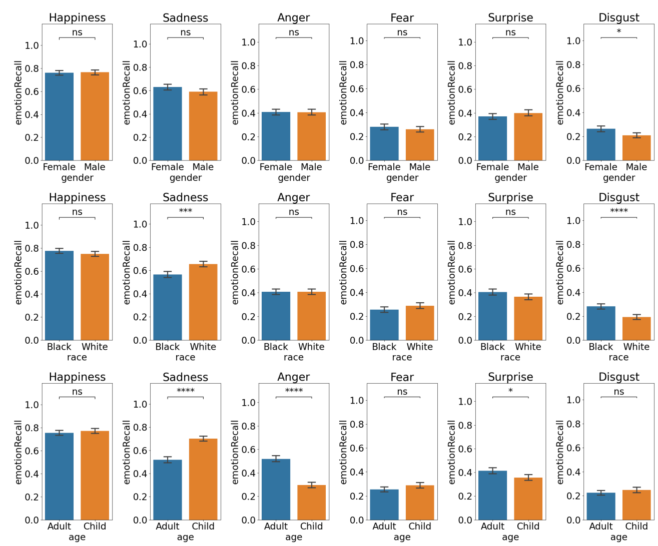


Figure 3: Confusion matrices for true and recognized emotions grouped by gender, race, and age in videos.



(a)



(b)

Figure 4: Emotion classification bias (a) and (b) sensitivity per gender, race, and age for each emotion in videos.

was slightly more accurate for females and surprise more accurate for adults.

To understand the effect of intensity, we evaluated emotion recognition biases across groups for each intensity level by two-tailed independent-subjects t-tests with Bonferroni correction. Figure 5 shows a trend for misidentification decrease with an intensity increase. Slightly significant ($p < 0.05$) differences in happiness bias were observed only for low intensities for female and adult characters. Highly significant differences in anger bias were found for high intensities in child characters ($p < 0.01$ to $p < 0.0001$) .

3.5 Discussion and Future Work

This work is an initial step towards assessing the effects of race, gender, and age on the facial emotion perception of virtual humans. We found happiness to be consistently the best-recognized emotion of all. This is in line with reports from the literature on emotion categorization, where happiness has been the emotion to be recognized the fastest [Hugenberg 2005]. The lowest accuracy was for disgust, which was highly confused with anger. Overall emotion recognition accuracy results were similar for all groups, which indicates that the traits we tested do not have any broad effects on how well emotions are perceived. However, there were some differences across groups for individual emotions. Consistent with the literature [Hess et al. 2004], female characters were perceived as happy significantly more often than male characters, especially when the expressions were ambiguous in lower intensities. Contrary to the literature, we haven't identified any difference in anger bias towards male or Black characters compared with female or White characters. However, higher disgust bias in the comparison groups (i.e., female and White) indicates a trend as disgust and anger were mostly confused. Age was a discriminating factor in the perception of sadness, perhaps due to a sensitivity for children's sadness..

Literature suggests that females outperform males at facial emotion recognition [Hall 1978; Wingenbach et al. 2018]. Our findings support this on virtual human faces. However, we didn't find any effects of participant gender on emotion classification accuracy or biases.

To keep the experimental requirements manageable, we only included binary characteristics per group, which is a simplification of

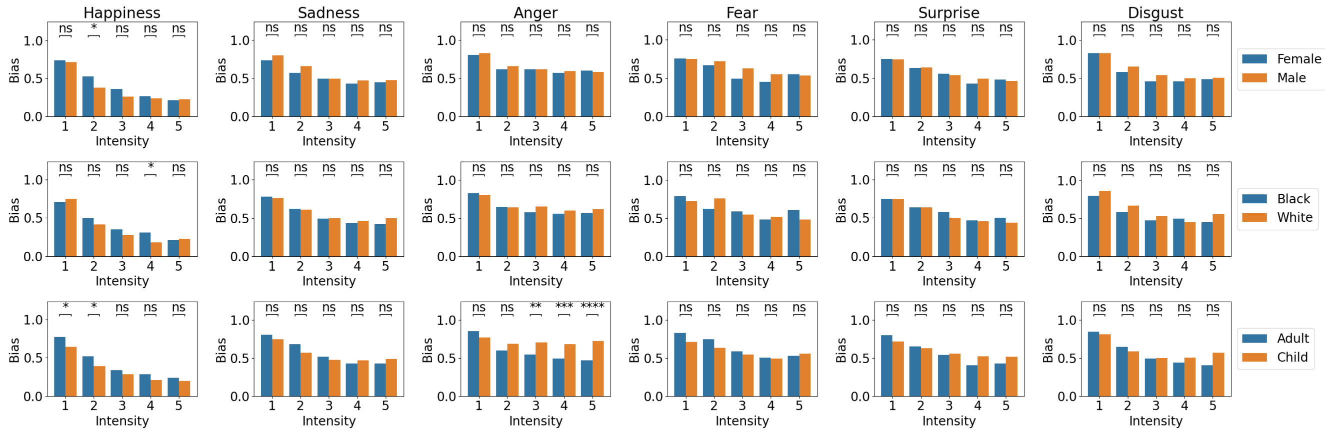


Figure 5: Emotion classification bias per gender, race, and age for each intensity level.

the diversity of human beings. Another limitation was the number of samples per stimulus group. Although there were three representative models per race, gender, and age group, each combination had only one sample. For instance, there was only a single White, adult, female model. We preferred to use off-the-shelf models whenever possible, as we wanted to understand the perceptual effects of easily accessible and prevalent models. However, ready-made models are not available for all the race/gender/age combinations. In the future, we plan to get better insights with a more diverse set of categories and a larger number of representative models per category. Future work also includes exploring the effects of model realism (e.g., cartoon vs. hyper-realistic models), rendering styles [McDonnell et al. 2012], and brightness/shadow levels [Wisessing et al. 2020].

REFERENCES

Victor Flávio Araujo, Diogo Hartmann Muller Schaffer, Angelo Brandelli Costa, and Soraia Raupp Musse. 2021. Perception of Computer Graphics Characters in Groups with Skin Color Diversity. In *Anais Estendidos do XX Simpósio Brasileiro de Jogos e Entretenimento Digital*. SBC, 975–978.

Kevin Beltran, Cody Rowland, Nicki Hashemi, Anh Nguyen, Lane Harrison, Sophie J Engle, and Beste F Yuksel. 2022. Using a Virtual Workplace Environment to Reduce Implicit Gender Bias. *International Journal of Human–Computer Interaction* (2022), 1–17.

Dario Bombari, Marianne Schmid Mast, Elena Canadas, and Manuel Bachmann. 2015. Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. *Frontiers in psychology* 6 (2015), 869.

Justin M Carré, Mark D Morrissey, Catherine J Mondloch, and Cheryl M McCormick. 2010. Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception* 39, 3 (2010), 356–377.

Vivian Hsueh Hua Chen, Gabrielle C Ibasco, Vetra Jing Xuan Leow, and Juline Yun Yee Lew. 2021. The Effect of VR Avatar Embodiment on Improving Attitudes and Closeness Toward Immigrants. *Frontiers in psychology* (2021), 4722.

Joshua Correll, Bernadette Park, Charles M Judd, and Bernd Wittenbrink. 2002. The police officer’s dilemma: using ethnicity to disambiguate potentially threatening individuals. *Journal of personality and social psychology* 83, 6 (2002), 1314.

Cassandra L Crone and Rachel W Kallen. 2022. Interview with an avatar: Comparing online and virtual reality perspective taking for gender bias in STEM hiring decisions. *PloS one* 17, 6 (2022), e0269430.

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.

Paul Ekman and James Cole. 1972. Universal and cultural differences in facial expressions of emotions. In *Nebraska symposium on motivation*, Vol. 19. 207–283.

Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. 2002. *Facial Action Coding System*. Research Nexus division of Network Information Research Corporation.

Hillary Anger Elfenbein and Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin* 128, 2 (2002), 203.

Ylva Ferstl and Rachel McDonnell. 2018. A perceptual study on the manipulation of facial features for trait portrayal in virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 281–288.

Ylva Ferstl, Michael McKay, and Rachel McDonnell. 2021. Facial Feature Manipulation for Trait Portrayal in Realistic and Cartoon-Rendered Characters. *ACM Trans. Appl. Percept.* 18, 4, Article 22 (oct 2021), 8 pages. https://doi.org/10.1145/3486579

Luciano Gamberini, Luca Chittaro, Anna Spagnolli, and Claudio Carlesso. 2015. Psychological response to an emergency in virtual reality: Effects of victim ethnicity and emergency type on helping behavior and navigation. *Computers in Human Behavior* 48 (July 2015), 104–113.

Amy G Halberstadt, Alison N Cooke, Pamela W Garner, Sherick A Hughes, Dejah Oertwig, and Shevaun D Neupert. 2020. Racialized emotion recognition accuracy and anger bias of children’s faces. *Emotion* (2020).

Judith A Hall. 1978. Gender effects in decoding nonverbal cues. *Psychological bulletin* 85, 4 (1978), 845.

Eric Hehman, Jessica K Flake, and Jonathan B Freeman. 2015. Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin* 41, 8 (2015), 1123–1134.

Ursula Hess, Reginald B Adams Jr, and Robert E Kleck. 2004. Facial appearance, gender, and emotion expression. *Emotion* 4, 4 (2004), 378.

Ursula Hess, Sylvie Blairy, and Robert E Kleck. 1997. The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior* 21, 4 (1997), 241–257.

Ursula Hess, Pascal Thibault, Reginal B Adams Jr, and Robert E Kleck. 2010. The influence of gender, social roles, and facial appearance on perceived emotionality. *European Journal of Social Psychology* 40, 7 (2010), 1310–1317.

Kurt Hugenberg. 2005. Social categorization and the perception of facial affect: target race moderates the response latency advantage for happy faces. *Emotion* 5, 3 (2005), 267.

Carroll E Izard. 1994. Innate and universal facial expressions: evidence from developmental and cross-cultural research. (1994).

Theodore Kim, Holly Rushmeier, Julie Dorsey, Derek Nowrouzezahrai, Raqi Syed, Wojciech Jarosz, and AM Darke. 2021. Countering Racial Bias in Computer Graphics Research. *arXiv preprint arXiv:2103.15163* (2021).

John P Lewis, Ken Anjyo, Taehyun Rhee, Mengjie Zhang, Frederic H Pighin, and Zhigang Deng. 2014. Practice and theory of blendshape facial models. *Eurographics (State of the Art Reports)* 1, 8 (2014), 2.

Adobe. 2022. Mixamo. https://www.mixamo.com/. Accessed: 2022-06-13.

Reallusion. 2022. Reallusion Character Creator. https://www.reallusion.com/character-creator/. Accessed: 2022-06-13.

Shutterstock. 2022. TurboSquid. https://www.turbosquid.com/. Accessed: 2022-06-13.

Unity. 2019. Unity. https://unity.com/. Accessed: 2019-08-14.

Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–11.

Clifford Nass and Byron Reeves. 2003. *The media equation: how people treat computers, television, and new media like real people and places.* CSLI Lecture Notes.

Margaret E Ormiston, Elaine M Wong, and Michael P Haselhuhn. 2017. Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual*

Differences 105 (2017), 40–42.

Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition* 22, 3 (2013), 779–787.

Brent Rossen, Kyle Johnsen, Adeline Deladisma, Scott Lind, and Benjamin Lok. 2008. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In *Proceedings of Intelligent Virtual Agents*. Springer Berlin Heidelberg, 237–244.

Katharina R Seitz, Jessica J Good, and Tabitha C Peck. 2020. Shooter bias in virtual reality: the effect of avatar race and socioeconomic status on shooting decisions. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 606–607.

Alexander Todorov, Christopher Y Olivola, Ron Dotsch, Peter Mende-Siedlecki, et al. 2015. Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual review of psychology* 66, 1 (2015), 519–545.

Alexander Todorov, Chris P Said, Andrew D Engell, and Nikolaas N Oosterhof. 2008. Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences* 12, 12 (2008), 455–460.

Laura D Wandner, Marc W Heft, Benjamin C Lok, Adam T Hirsh, Steven Z George, Anne L Horgas, James W Atchison, Calia A Torres, and Michael E Robinson. 2014. The impact of patients' gender, race, and age on health care professionals' pain management decisions: an online survey using virtual human technology. *International Journal of Nursing Studies* 51, 5 (2014), 726–733.

T. Wingenbach, C. Ashwin, and M. Brosnan. 2018. Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PloS ONE* 13, 1 (2018). https://doi.org/10.1371/journal.pone.0190634

Pisut Wisessing, Katja Zibrek, Douglas W Cunningham, John Dingliana, and Rachel McDonnell. 2020. Enlighten me: Importance of brightness and shadow for character emotion and appeal. *ACM Transactions on Graphics (TOG)* 39, 3 (2020), 1–12.

Sarah A Zipp, Tyler Krause, and Scotty D Craig*. 2017. The impact of user biases toward a virtual human's skin tone on triage errors within a virtual world for emergency management training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 2057–2061.

APPENDIX

Figure 1: Facial expressions with increasing intensity per row for Black female adult model.

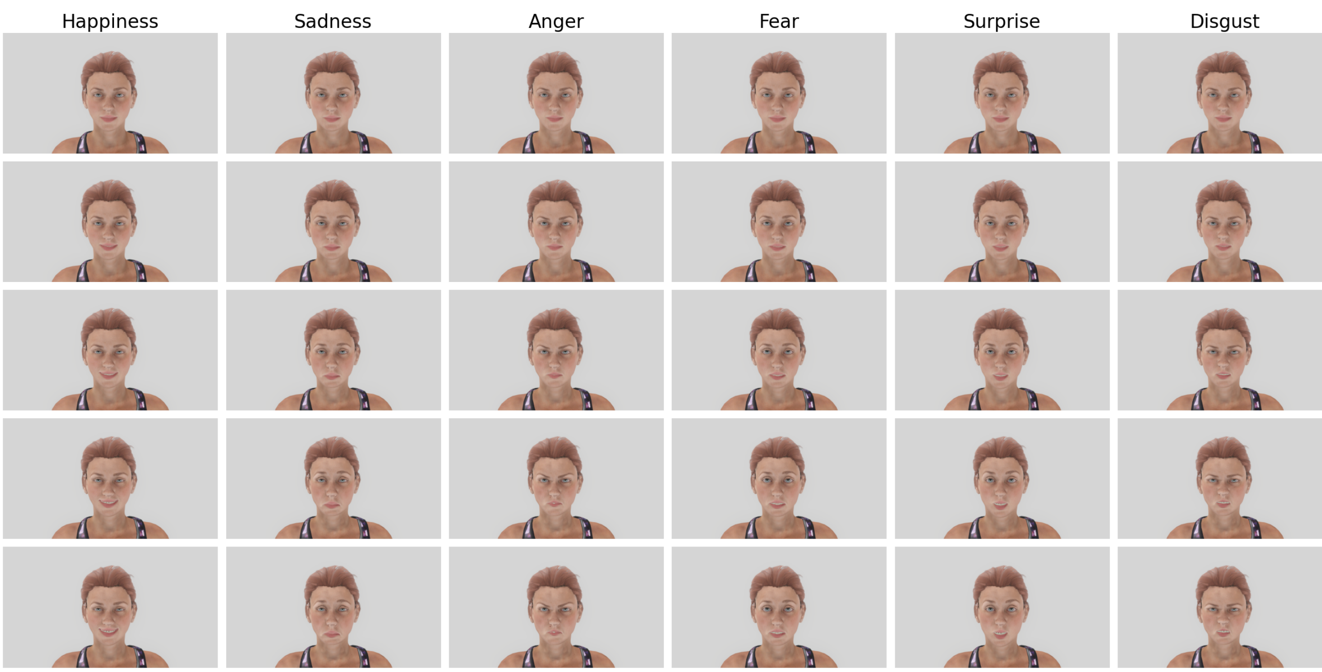


Figure 2: Facial expressions with increasing intensity per row for White female adult model.

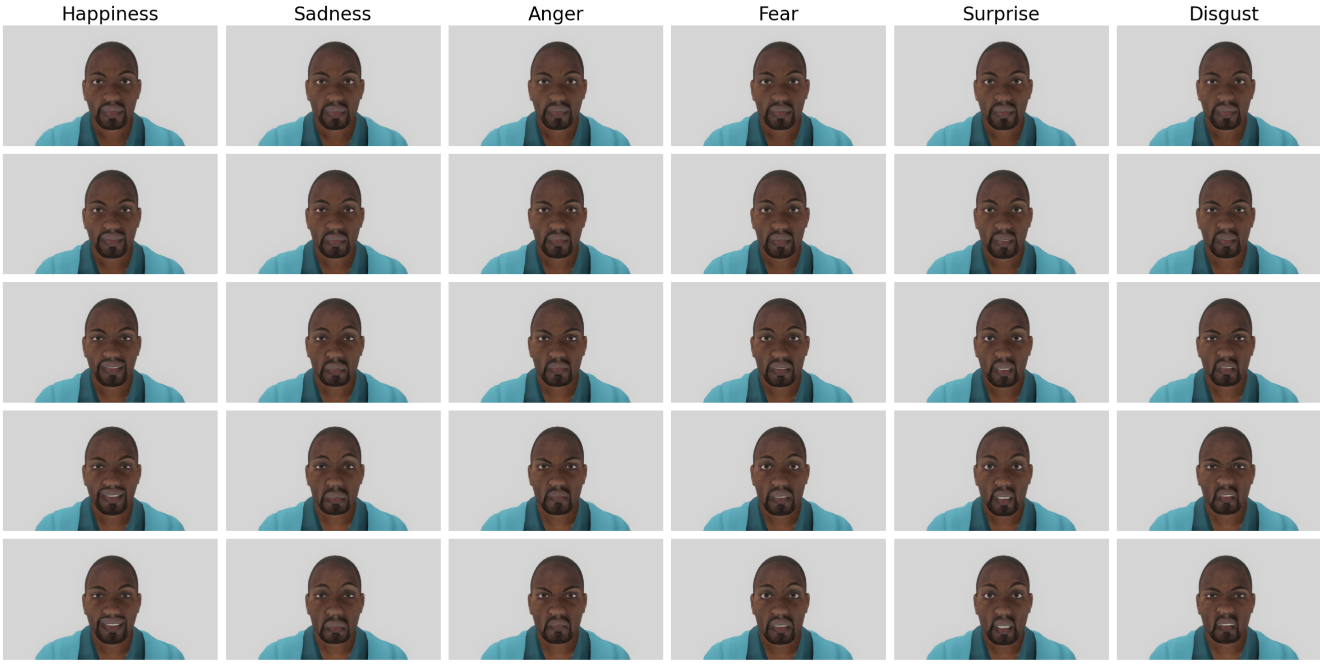


Figure 3: Facial expressions with increasing intensity per row for Black male adult model.

Figure 4: Facial expressions with increasing intensity per row for White male adult model.

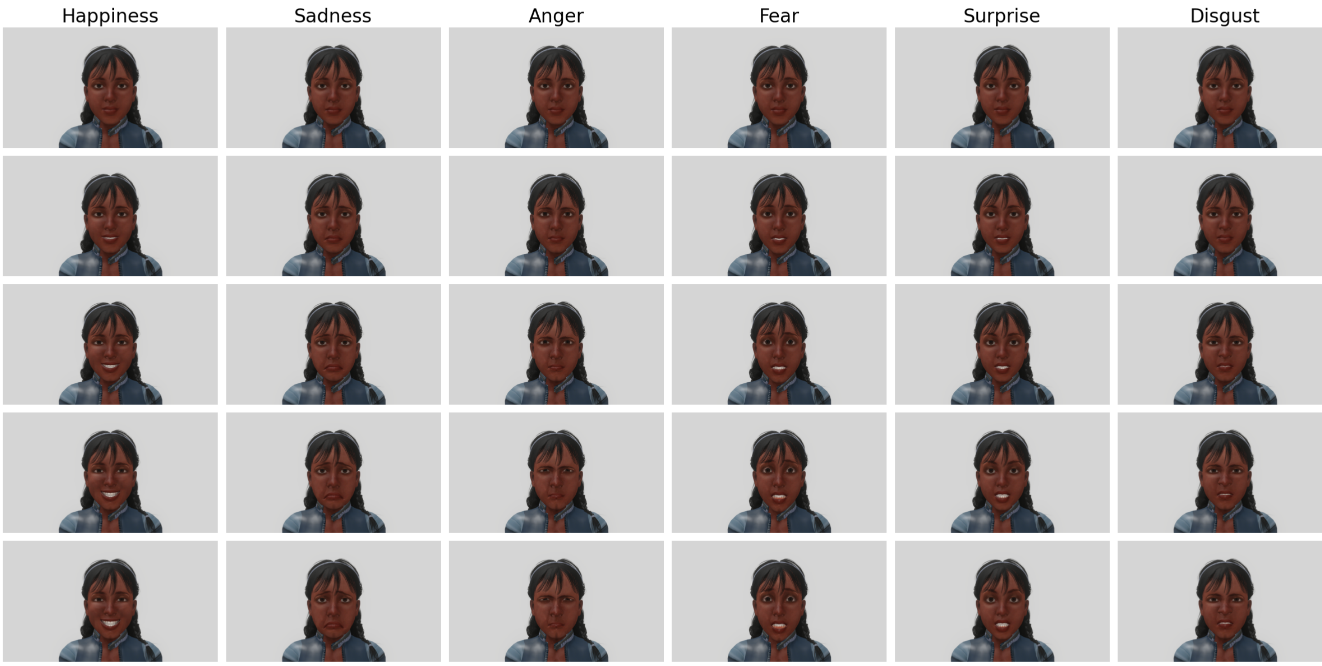


Figure 5: Facial expressions with increasing intensity per row for Black female child model.

Figure 6: Facial expressions with increasing intensity per row for White female child model.

Figure 7: Facial expressions with increasing intensity per row for Black male child model.

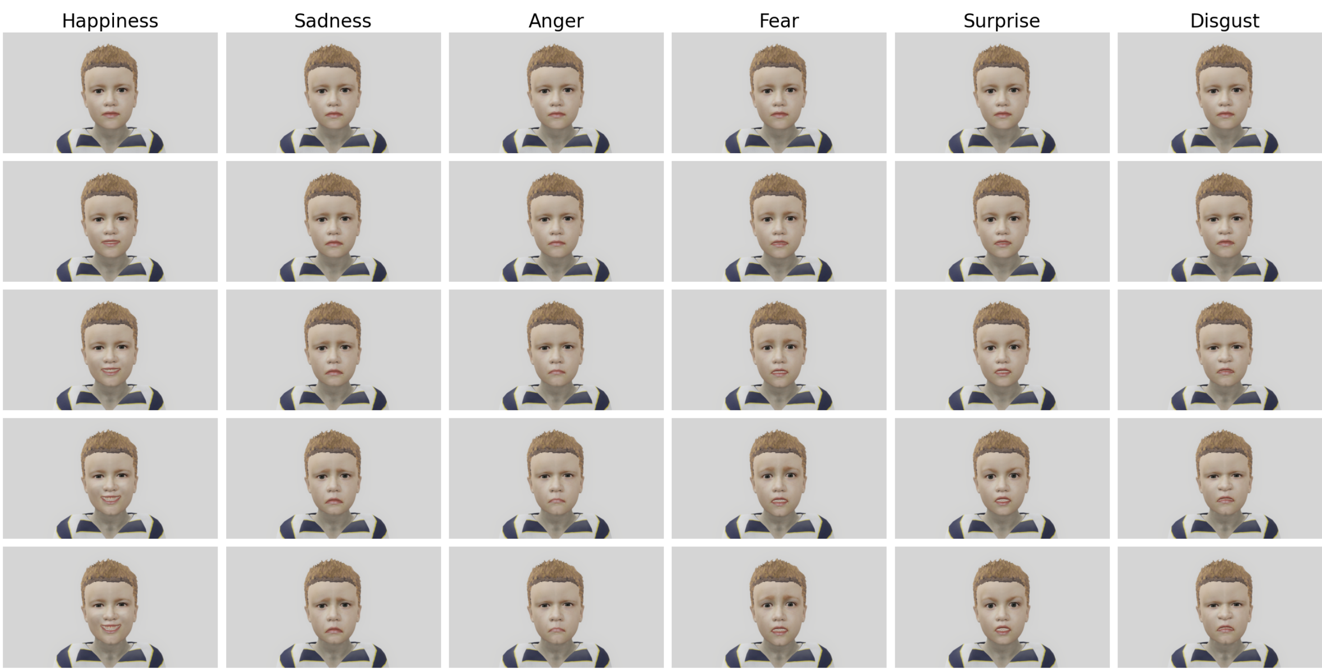


Figure 8: Facial expressions with increasing intensity per row for White male child model.