PUCRS

ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO

VICTOR FLÁVIO DE ANDRADE ARAUJO

DIVERSITY IN VIRTUAL HUMANS: UNVEILING BIASES IN HUMAN CHARACTERISTICS REPRESENTATION

Porto Alegre
2024

PÓS-GRADUAÇÃO - *STRICTO SENSU*

AD VERVM DVCIT

Pontifícia Universidade Católica do Rio Grande do Sul

Pontifical Catholic University of Rio Grande do Sul
Faculty of Informatics
Computer Science Graduate Program

DIVERSITY IN VIRTUAL HUMANS: UNVEILING BIASES IN HUMAN CHARACTERISTICS REPRESENTATION

VICTOR FLÁVIO DE ANDRADE ARAUJO

Dissertation submitted to the Pontifical Catholic University of Rio Grande do Sul in partial fullfillment of the requirements for the degree of Ph. D. in Computer Science.

Advisor: Profa. Dra. Soraia Raupp Musse
Co-Advisor: Prof. Dr. Angelo Brandelli Costa

Porto Alegre
2024

Ficha Catalográfica

A553d Andrade Araujo, Victor Flávio de

Diversity in virtual humans : unveiling biases in human characteristics representation / Victor Flávio de Andrade Araujo. – 2024.
169.
Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Soraia Raupp Musse.
Coorientador: Prof. Dr. Angelo Brandelli Costa.

1. Percepção. 2. Humanos Virtuais. 3. Vieses. 4. Gênero. 5. Cor da Pele. I. Raupp Musse, Soraia. II. Brandelli Costa, Angelo. III. , . IV. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

Victor Flávio de Andrade Araujo

DIVERSITY IN VIRTUAL HUMANS: UNVEILING BIASES IN HUMAN CHARACTERISTICS REPRESENTATION

This Doctoral Thesis has been submitted in partial fulfillment of the requirements for the degree of Doctor of Computer Science, of the Graduate Program in Computer Science, School of Technology of the Pontifícia Universidade Católica do Rio Grande do Sul.

Sanctioned on March 26th, 2024.

COMMITTEE MEMBERS:

Profª. Drª. Birgit Lugrin (Institute of Computer Science - Media Informatics/University of Wuerzburg)

Prof. Dr. Marcelo Walter (PPGCC/UFRGS)

Prof. Dr. Avelino Francisco Zorzo (PPGCC/PUCRS)

Prof. Dr. Angelo Brandelli Costa (PPGP/PUCRS – Co-Advisor)

Profª. Drª. Soraia Raupp Musse (PPGCC/PUCRS - Advisor)

“You cannot hope to build a better world without improving the individuals. To that end each of us must work for his own improvement, and at the same time share a general responsibility for all humanity, our particular duty being to aid those to whom we think we can be most useful.”
(Marie Curie)

Acknowledgments

I would first like to express my deepest gratitude to my family. Mainly, due to the unwavering support of my wife Tamires. She really was my partner, and she always accompanied me during this journey (in addition to listening to my good and bad lamentations and crazy things). I would also like to thank all the friends and partnerships I made at VHLAB. Furthermore, I would like to extend my thanks to my advisor during the sandwich period, Catherine Pelachaud, for her patience, time and countless suggestions that enriched this work, and for welcoming me into her laboratory in Paris (it was a fantastic period, in terms of knowledge, social and personal). I would also like to thank my co-advisor, Angelo Costa, for having helped a lot with his knowledge in the field of Psychology, especially in sensitive subjects such as gender and skin color (certainly, his co-supervision was essential for each study present in this text). Last but not least, I would like to thank my advisor Soraia Musse, for supporting me throughout these last six years (between the master's degree and the doctorate), and for training me as a researcher (today, after that, for me research is as much a hobby as it is a job, that is, changing my perspective).

DIVERSIDADE EM HUMANOS VIRTUAIS: REVELANDO VIESES NA REPRESENTAÇÃO DAS CARACTERÍSTICAS HUMANAS

RESUMO

Nos últimos anos, a Computação Gráfica passou por avanços notáveis, particularmente evidentes no realismo aprimorado de animações e personagens apresentados em filmes, jogos e diversas mídias. No entanto, a teoria do Vale da Estranheza sugere que, à medida que os Humanos Virtuais se aproximam do realismo aumentado, há uma probabilidade maior de evocar desconforto em quem se envolve com eles. Este fenômeno representa um desafio significativo no domínio dos filmes e jogos, levando os profissionais da indústria e os investigadores acadêmicos a aumentarem o seu foco para o estudo da percepção humana. Reconhecendo que os seres humanos são utilizadores de entidades virtuais, as considerações vão além dos meros aspectos visuais e comportamentais para abranger um conjunto diversificado de características, incluindo gênero, cor da pele, características faciais, preconceitos do mundo real dentro do mundo virtual, etc. Portanto, este trabalho tem como objetivo realizar estudos sobre a percepção humana em relação a humanos virtuais, e quais sensações são causadas por suas diferentes características em uma abordagem que envolve vieses de gênero e cor da pele. Por exemplo, estudar a relação entre o gênero do personagem e o gênero do participante da pesquisa. Além disso, com base nos resultados dos estudos, este trabalho tenta encontrar soluções para reduzir tais vieses através da modelagem (design) de humanos virtuais. Os estudos deste trabalho apresentam fortes indícios de vantagem intragrupo dos participantes em relação aos humanos virtuais com gêneros binários, tanto para atribuição de gênero quanto para reconhecimento de emoções. Além disso, este trabalho fornece possíveis soluções para a desconstrução do gênero binário no uso de um Bebê Virtual sem gênero definido, e de um humano virtual adulto também sem gênero definido. Em relação à cor da pele, este trabalho apresenta discussões envolvendo Vale da Estranheza e humanos virtuais com diferentes cores de pele. Nesse caso, resultados apontam que humanos virtuais com pele escura talvez sejam considerados menos realistas do que humanos virtuais de pele branca devido ao fato de que os algoritmos de criação de cor da pele podem ser tendenciosos.

Palavras-Chave: Percepção, Humanos Virtuais, Vieses, Gênero, Cor da Pele.

DIVERSITY IN VIRTUAL HUMANS: UNVEILING BIASES IN HUMAN CHARACTERISTICS REPRESENTATION

ABSTRACT

In recent years, Computer Graphics (CG) has undergone notable advancements, particularly evident in the enhanced realism of animations and characters featured in movies, games, and various media. However, the Uncanny Valley (UV) theory suggests that as Virtual Humans (VHs) approach heightened realism, there is an increased likelihood of evoking discomfort in those who engage with them. This phenomenon poses a significant challenge in the realms of movies and games, prompting industry professionals and academic researchers to broaden their focus to study human perception. Recognizing that human beings are the users of virtual entities, considerations extend beyond mere visual and behavioral aspects to encompass a diverse array of characteristics, including gender, skin color, facial features, etc. This inclusive perspective raises concerns about the potential replication of real-world biases within the virtual realm. Therefore, this work carries out studies on human perception in relation to VHs, and what sensations are caused by their different characteristics in an approach involving gender and skin color biases. For example, studying the relationship between the character's gender and the research participant's gender. Furthermore, based on the results of the studies, this work attempts to find solutions to reduce such biases through the design of VHs. The studies in this work present strong indications of in-group advantage of participants in relation to VHs with binary genders, both for gender attribution and for emotion recognition. Furthermore, this work provides possible solutions for deconstructing the gender binary in the use of gerderless Virtual Baby (VB) and a genderless Adult VH. Regarding skin color, this work presents discussions involving UV and VHs with different skin colors, which point out that VHs with dark colored skin are perhaps considered less realistic than white colored skin VHs due to the fact that skin color creation algorithms skin color may be biased.

Keywords: Perception, Virtual Humans, Bias, Gender, Skin Color.

List of Figures

List of Tables

List of Acronyms

VH Virtual Human
UV Uncanny Valley
VB Virtual Baby
CG Computer Graphics
VR Virtual Reality

Contents

1. INTRODUCTION

In recent years, advances in Computer Graphics (CG) have allowed the entertainment industry to create very realistic Virtual Humans (VHs) [FCW21]. In some movies, real actors have been replaced by CG characters/VHs (such as Disney's 2016 Rogue One movie, 2019 Aladdin movie, and 2020 The Mandalorian series), and often this substitution is not even perceived by the public, however, there are still some perceived artifacts, such as the movement of the mouth and eyes [MM16]. Avoiding these oddities is important to generate a VH more similar to a human being, making the experience of watching a movie, playing a game, or interacting with a VH more fluid. Moreover, to understand the human sensations of discomfort related to VHs, we need to study human perception in relation to the assisted or interacted stimulus. In other words, we need to understand the reasons why people feel comfortable or not with VHs. Human perception is also a theme present in many pieces of research in CG [ZAJ+15, ZZM19], and it is considered very relevant when discussing the evolution of VHs and realistic faces.

One of the known theories in human perception regarding VHs is the Uncanny Valley (UV). According to Mori [MMK12], artificial beings made to appear too similar to real humans can fall into the UV, where sometimes a high degree of human realism evokes a strangeness feeling in the viewer, which can cause sensations of discomfort and negative emotional valences. According to Katsyri et al. [KFMT15], this sensation of discomfort is related to human identification, that is, the human being looks for human characteristics in the artificial being. Regarding human identification, according to Social Identity Theory [TT04], an individual tends to see their own group with a positive bias (in-group) compared to an out-group context. Furthermore, the theory also considers that, from a socio-psychological perspective, the individuals define themselves and are defined by others as members of a group. According to Brown [Bro20], Social Identity theory explains that people prefer to see themselves and their groups in a positive light, which implies that they look for positive distinction in their perceptions and relationships over other groups. In this case, some human characteristics that fit into these contexts are gender, skin color, race, etc. Both in the field of psychology and in computing, there are studies on in-group advantages [EA02, KST+15], that is, people in a group recognize and encode better the facial/body characteristics of real or VHs from their same group.

In addition, there are few studies with results focused on out-group advantages [SOS+22, HSS17] concerning VHs and UV, that is, people from a minority group evaluating VHs from a different group (out-group). In particular, creating CG characters/VHs for different media (movies, series, games, etc) should involve human representation that reaches all groups of human perception. For example, in work by Saneyoshi et al. [SOS+22], the authors evaluated the other-race in the UV effect concerning realism variations in a continuum of two VHs, one Japanese, and the other European. The authors concluded that in terms of perceived pleasantness, in-group participants did not tolerate slight differences.

As we can see in the previous paragraphs, human perception should be the main focus of VH design for both industry and academic research. Following these lines of thought, there are other questions in addition to the UV and Social Identity Theory (related to in-group and out-group contexts), that can be studied. For example, if the UV effect (perceptions of realism and discomfort) is impacted by the gender of VHs or the viewers, or if the human perception changed during the several years we are exposed to CG, and how this exposure impacts the gender diversity and skin color of groups of VHs. According to Tinwell et al. [TGW11], the technological advancements that help develop realistic VHs is accompanied over the years by people's discernment about this content. With that, Tinwell and colleagues believed that the UV would never be surpassed since discernment can help people to observe better the technical tricks.

The human perception is a complex phenomenon that is difficult to explain due to its multi-faceted nature. Gender is one of this dimensions, specifically in the VH context [SRLZ14, NY20, ZHRM13, AMRM19]. In recent years, the entertainment industry has made an effort to increase the representation of female characters in movies and video games[1,2]. For instance, between 2016 and 2020, the number of female protagonists in video games grew by 16%[3]. In a paper by Draude [Dra11], it is suggested that creating female characters in movies and games can make them more relatable and comfortable for viewers. However, despite this effort, the number of male VHs still outnumber females. This raises some questions about the reasons behind this gender imbalance: Is human perceived comfort influenced by gender, and how does this relate to realism? Are gender biases that occur in the real world also present in the virtual world? In simpler terms, does a VH's gender affect how they are perceived by humans?

[1] https://www.playstation.com/en-us/games/the-last-of-us-part-ii-ps4/
[2] https://www.playstation.com/en-us/games/horizon-zero-dawn-ps4/
[3] https://www.statista.com/statistics/871912/character-gender-share-video-games/

Another open question related to gender is how we perceive VHs when they can be visibly associated with a particular gender and whether that perception changes depending on the observer's gender. Laue [Lau17] states that we, as a species, tend to anthropomorphize and sympathize, even with the simplest and weirdest technologies. In addition, from the designer to the audience, we tend to attribute human characteristics, such as gender, to these technologies. Regarding the gender of a person who observes, according to Hall et al. [HHM10] and Cheetham et al. [CPJ+13], women more easily perceive tasks of face detection, discrimination of facial identity and facial expressions, in human beings. Regarding the faces of the VHs, according also to Chethaam et al. [CPJ+13], women recognize faces similarly to how they recognize real human beings, but more quickly than men. Still, according to the authors, regarding the categorization of less realistic faces, there is no significant difference between the perception of women and men.

One important human characteristic is skin color. In terms of biases related to skin color, for a long time, dark colored skin VHs were created and modeled from stereotypes, such as VHs presented without shirts and with lots of muscles in fighting games [4]. According to Kim et al. [KRD+22][5][6], multiple dimensions of diversity should be considered. The authors have pointed out that most techniques for creating skin color through CG are based on white skin color. They have also mentioned that the emphasis on lighting is exaggerated when applied to dark skin. This raises a question, whether dark colored skin conveys a lower level of realism to people compared to white colored skin. If this is true, then according to the UV theory and other studies that support it ([MC16, MGHK09]), reducing realism in VHs can change the perceived eeriness or discomfort. Bringing it to the context of our work, can VHs with dark colored skin cause discomfort or affect perceived realism?

The purpose of this study is to examine how participants perceive various attributes such as comfort, realism, and emotions in relation to VHs with diverse gender and skin color. The study aims to determine whether biases related to gender and skin color, which may occur in the real world, also exist in the virtual world. Some of the specific topics discussed in this study include:

- Are people feeling more comfortable with newer CG characters than older ones? To answer this question, discussed in Chapter 3 - Section 3.1, we repro-

[4] https://www.pbs.org/independentlens/blog/leveling-up-representation-depictions-of-people-of-color-in-video-games/

[5] www.youtube.com/watch?v=ROuE8xYLpX8

[6] https://yaledailynews.com/blog/2022/03/09/yale-professors-confront-racial-bias-in-computer-graphics/

duced the work by Flach et al. [FdMM+12, DFH+12] which goal was to assess people's perception of comfort and realism in relation to VHs through the UV theory. Furthermore, as the work only had VHs until 2012, new VHs were included in our research, from 2012 until 2020. Similar to presented by Flach et al. [FdMM+12, DFH+12] and Kätsyri et al. [KMT17], which analyzed characters from different media, we were interested about the human perception of VHs in everyday situations, that is, without stylizing the characters, without controlling the intensity of emotion, lighting or shading. That is why characters from different media were used with images and videos obtained from the internet. The hypothesis is that perceptual comfort can be tested and measured as it is, in real life, even if there is no control over the tested data. While it is desirable for experiments to be highly controlled, e.g., lights and colors, we justify that is also interesting to study wild characters, as in games and movies, since they represent what people usually see and play with in real life. Therefore, we selected Flach et al. [FdMM+12, DFH+12] work because it contains CG characters from different media, and with different gender and skin color characteristics, and also have access to people's evaluations of perceived comfort and realism in relation to them, in 2012.

- Do people perceive comfort and realism in female and male CG characters in the same way? Does this change when participants are separated into women and men? These questions are also analyzed through the data collected in Chapter 3 - Section 3.1 and also discussed in Section 3.2.

- Are comfort and realism similarly perceived in a VH with different skin colors? To answer this question, a perceptive study was carried out on VHs, also using a questionnaire based on the same used in the work by Flach et al. [FdMM+12]. The question is analyzed in Section 3.3 of Chapter 3.

- Additionally, in terms of gender, we investigate about VHs thinking in a binary way (female and male), but how do people perceive genderless VHs? Does this change with respect to participant gender? Do we have a bias regarding gender attribution? To answer these questions, the work of Condry and Condry [CC76] has been reproduced in a virtual environment. In that work, the authors carried out a perceptual study in which they present, through videos, a real baby reacting to objects. The aim was to find out whether women and men perceived whether babies of different genders conveyed emotions differently. However, the baby was always the same, that is, only changing the textually gender identification and the name. In the case of the current work, a Virtual Baby (VB) was used. These questions are discussed in Chapter 4 - Section 4.1.

- Still regarding the VB study, would the perception results change if the participant could interact with the VB? That is, would the gender-based preference also emerge when interacting with a genderless VB? In some psychology studies [WSD76, CCH83, SKZ75, SL80, SL78, Gre89], in addition to just observing babies reacting to objects, participants could choose which objects a baby could play with, and they could also to react emotionally to the baby's actions. The objects had stereotypical visual characteristics that linked them to a specific gender, for example, a ball being related to a boy. The same baby was used, objects and virtual environments as in the previous item, and carried out a new perceptual experiment, in which participants could interact with the VB, choosing toys, reacting to the baby's actions with the toys, and approaching the baby with the mouse and keyboard. The question is analyzed in Chapter 4 - Section 4.2.

- Finally, would the gender-based preference also emerge when observing a genderless virtual adult instead of a VB? To try to answer this question, a nonbinary (genderless) virtual adult was modeled and served as stimuli for a new perceptual experiment. In this experiment, participants saw the virtual adult expressing emotions based on videos of real actors. To compare the results with the VB, the participants were divided into three groups. The first group received the virtual adult with a female name, the second group received the virtual adult with a male name, and the third group received the virtual adult without any name. The question is analyzed in Chapter 4 - Section 4.3.

1.1 Research Problem

The study of human perception in relation to VHs is beneficial for various industries and professionals that use them in design. These industries include cinema, gaming, designing, animation, development, teaching, and psychology, among others. People watch movies and play video games to feel good, and professionals who create these VHs must have a sense of what to do to make consumers feel good. Applications of VHs can be found in conversational agents used in industries or healthcare. For instance, psychologists or doctors may require VHs that do not cause negative feelings in their patients. With the last advances in hardware and CG software, much can be done, one main question is how we are going to model and animate next VHs in terms of their race, age, skin color, gender, in order to provide a good and fair experience to the users? We also need to think about developing a culture of bias reduction by modeling non-biased gender behaviors and modeling VHs with different skin colors. Therefore, one important finding can arise from this research area: are the algorithms enough advanced in order to animate and render realistic and diverse VHs?

1.2 Goals

The primary objective of this work is to conduct research that aids in designing VHs that consider gender and skin color diversities while taking into account human perception. For this, there are specific goals:

- **To study the influence of technological evolution, and its importance for gender and skin color diversity**. As technology is always evolving, it is necessary to know if this evolution influences human perception of VHs;
- **To study the influence of demographic characteristics**. It is important to analyze characteristics rooted in human culture such as gender, skin color, and others. Such analyses will enable us to understand if these characteristics affect human perception about VHs.
- **To study the in- and out- group influences in terms of gender and skin color of VHs**. From demographic studies, we can analyze, through psychological theories, the preferences, and effects of gender and skin color of VHs on people's perception.

Based on all these studies, this work intends to propose some recommendations for the design of VHs, always considering the objective of providing computing optimization (propose something that can realistically be done), diversity, and correct population representation to create VHs. Yet, for the audience, this work wants to improve the comfort, security, diversity, and usability experience.

1.3 Text Structure

The text is structured as follows: Chapter Related Work presents papers that influenced this present work; Chapter Perception of Virtual Humans from Different Media presents analyzes related to perceptions of VHs from different media in an UV theory perspective. In this case, evaluating perceived comfort and realism between different years, and how these differences impacted perceptions of gender and skin color of VHs; Chapter Gender Bias in Terms of Gender Attribution presents analyzes related to gender biases through attributes as gender attribution and perception of emotions in VHs with binary and nonbinary genders; Chapter Final Remarks presents final considerations of the analyzes presented during all the text, and the published papers. In addition, References are at the end of the text.

2. RELATED WORK

With the aim of answering the questions presented in Chapter 1, this current chapter presents studies and research from the literature that involve human perception in relation to VHs, mainly in relation to gender and skin color. Therefore, the chapter is divided into the following: *i)* Section Human Perception, where concepts about perception are explained and some examples are presented; *ii)* Section Uncanny Valley Theory presents concepts and examples about UV theory; *iii)* Section In-group and Out-group Effects explains concepts and examples about in- and out-group theories; *iv)* In Section Skin Color Impact in VHs, based on the concepts explained in the previous sections (in- and out-group, UV, and perception), examples involving the skin color of VHs are presented; *v)* In Section Gender Impact in VHs, as well as in Section 2.4, examples involving gender in relation to VHs are presented; *vi)* and finally, Section The Importance of the Studies Presented in the Related Work explains the importance of this chapter for the progress of the work.

2.1 Human Perception

Perception comes from the Latin word with the same spelling that means "to apprehend" or "to understand". According to Zell et al. [ZZM19], the brain processes information from the senses and interprets its relevance to the organism. The first process of transforming information driven by the senses is called the ascending process. The second process is based on information acquired about the world through learning and provides a context for meaning information to be interpreted, known as the top-down process. Therefore, when looking at a person or a character, for example, a perception is created about what is being seen and represented by the face of the other.

Concerning the perception of VHs, many papers have evaluated the same variables (and others) that are used in this work. For example, the work of Zell et al. [ZZM19] was essential to understand the perception process (i.e., how to create a stimulus, how to measure and evaluate perceptual data, etc.). In other work of Zell et al. [ZAJ+15], the authors analyzed two traits of appearance: shape and material, and with the help of artists, they designed elaborate stimuli consisting of different stylization levels for both parameters. They analyzed how different combinations affect the perceived realism, appeal, eeriness, and familiarity. Also, the authors investigated how such combinations affect the perceived intensity of different facial expressions, and concluded that the shape of a character is relevant to its realism and expression. Chaminade et al. [CHK07] investigated how the appearance of animated characters can influence the perception of their actions. The authors presented different animated characters with movement data captured from human actors or by interpolation between poses and asked the participants to categorize movement as biological or artificial. The results showed that the more anthropomorphic, the less biological bias the character had. Wisessing et al. [WZC+20] designed perceptual studies to assess the effect of lighting on perception. The results were significant between lighting with character emotion, and with appeal.

2.2 Uncanny Valley Theory

The UV is a theory created by roboticist Masahiro Mori [MMK12] who analyzed the emotional reaction of humans to artificial beings, that is, it is a theory that takes into account the human perception of these artificial beings. According to his theory, if robots have a high degree of realism close to real humans, they may fall into the UV, which can cause an eerie impression on the viewer. From Mori's seminal work, several other researchers have used the UV theory to measure artificial characters' discomfort (robots, CG characters). For example, Katsyri et al. [KFMT15] reinterpreted the original UV hypothesis and revisited empirical evidence for the theoretically motivated UV hypotheses. One of their findings suggested that UV exists only under specific conditions, such as inconsistencies in realism (e.g., artificial eyes on a human-like face). In this work, the UV theory is used to measure comfort and perceived realism related to characters created using CG.

The effect of the UV theory on human perceptions of 3D models also has been investigated by the CG community. For example, in other work of Kätsyri et al. [KMT17], the authors investigated whether characters from semi-realistic animated films provoke negative human perceptions. In the work of MacDorman and Chattopadhyay [MC16], the main objective was to determine whether reducing the consistency of realism in visual characteristics increases the effect of the UV. The authors' hypotheses are based on the theory of the inconsistency of realism, which predicts that the effect of the UV is caused by an entity that has characteristics that are not all perceived as belonging to a real living anthropomorphic being. Through this hypothesis, the authors investigated the animacy for humans and non-human animals, and realism for humans, non-human animals, and non-human objects. Schwind [Sch18] conducted nine studies that examined the effects of UV on human perception, how it affects interaction with computer systems, what cognitive processes are involved, and the causes that may be responsible for the phenomenon. The work of Hyde et al. [HCKH16] and Hodgins et al. [HJO+10] reported how relevant realism is for characters created using CG. In the first, the authors performed two experiments to show how exaggerated facial movement can influence perceptions about cartoon characters and more realistic animated characters. In the second, the authors performed perceptual experiments, which explored how different anomalies have relative importance. For this, the authors used two methods: a questionnaire to determine the emotional perception about complete vignettes, with and without facial and auditory movement; and, through a task, they compared the performance of a virtual “actor" in short clips (extracted from the vignettes), which represented a series of different facial and body anomalies. Ruhland et al. [RPM17] used algorithms to synthesize real-time motion capture of human expressions with animation data created by designers. To validate synthesized animations, they conducted a perceptual study, and results indicated that the animations had an expressive similarity to animations made by hand.

Flach et al. [FdMM+12] investigated the UV theory to evaluate its effects on the perception of CG characters used in movies, animations, and computational simulations. The authors evaluated the human perceptions about these characters through a questionnaire containing images and videos of these characters. In the present work, the experiment is recreated with the same questionnaire, and also the same images and videos of CG characters. In addition, statistical analyses were performed to compare the results as opposed to Flach’s study, which did not include it. The work of Flacht et al. [FdMM+12] is an extensive work of the work of Dill et al. [DFH+12], wich the authors also investigated the UV theory to assess its effects on human perception of CG characters in various media.

2.3 In-group and Out-group Effects

Some studies can help us to understand in-group and out-group advantages. In Beaupre’s work [BH05], the author investigated cultural differences in facial recognition accuracy and in-group advantage for emotion recognition among sub-Saharan African, Chinese, and French Canadians living in Canada. The author found evidence that people can decode emotions better in groups they belong. According to the author, individuals of the same racial/ethnic group have advantages in facial decoding, attributing subtle differences in expressive style between members of different cultural groups, which could be more difficult to decode in the case of people outside their group. In addition, the author also mentioned that individuals from the same group could judge emotional expressions more easily due to factors such as familiarity with facial morphology.

In the same line of morphological questions, Balas and Nelson [BN10] evaluated facial shape and skin pigmentation (variation between dark and white skin) information in virtual humans through electrophysiological and behavioral measures of participants. The goal was to examine how the shape and color of facial skin modulate neural responses. In the results, the authors found that the two neural components evaluated (N170 and N250) were modulated by skin color, not depending on facial shape. However, the N250 component showed face sensitivity. In addition, participants overall reported that they did not have many exposures to faces with dark skin. This aspect of in-group advantage over own race is a natural effect of human beings, as explained in the work of Walker and Tanaka [WT03]. The authors also performed a perceptual study of the effect of in-group regarding race/ethnicity by generating a continuum of images by transforming an East Asian father's face into a Caucasian father's face. In the results, there was an interaction between the participating race/ethnicity and the race/ethnicity of the face. Thus, the authors concluded that these effects might occur at an early human developmental stage of face coding.

In the work of Kang and Lau [KL13], the authors conducted studies to explore out-group advantages. In one of the results, one of the groups of participants was European Americans, and the other group was Asian Americans, and both groups were asked to identify imitated emotions and spontaneous emotions from each of the two groups, that is, in-group and out-group. One of the results revealed that an in-group advantage was observed in the spontaneous expressions scenario but not in the mimicked expression condition. According to Kang and Lau, individuals from ethnic minorities who live in multicultural settings may be more accurate in recognizing the emotions of members of the majority group. The authors also mentioned that when two groups are conditioned to the same national border, members of minority groups can more accurately recognize the emotions of members of a majority group than vice versa. In work by Hehman et al. [HMG10], the authors investigated whether the out-group question would result from social categorization and not perceptual experience. Eight dark and white-skin faces were shown to participants, which were spatially organized by race/ethnicity or university affiliation. Results showed that participants had related memory for own-race faces when grouped by race/ethnicity. While, when grouped by the university, participants had positive results for faces of the university itself.

Still on out-group advantage, the work of Elfenbein and Ambady [EA03b] showed that the results suggested that emotional recognition can be characterized by subtle differences in style across cultures [EA03a], which become more familiar with greater cultural contact. The greater the cultural contact with, for example, VHs with dark colored skins, the more people become familiar with these characters. The out-group advantage explains the result of the work of Saneyoshi et al. [SOS+22], where results showed that Japanese participants reacted similarly to Asian and European VH groups.

2.4 Skin Color Impact in VHs

Several studies regarding VHs' perception correlate with issues of race/ethnicity and skin color. For example, in work by Banakou et al.[BBN+20], the authors performed a perceptual study with two conditions, BodyType (their virtual body had a White or Black-skin) and the surrounding crowd was Negative, Neutral, or Positive towards the participant. The authors argued that a negative effect prevents the formation of new positive associations with black people, and anguish leads to the dispossession of the virtual body. The work of Bedder et al. [BBB+19] explains implicit bias towards the negative evaluation of people in the out-group when using a virtual body, for example, a white person using a black virtual body. In another work by Banakou et al. [BHS16], the authors mention that incorporating white people into a black virtual body is associated with an immediate decrease in their implicit racial prejudice against black people. With this, the authors tested whether the reduction in implicit bias lasts for one week and whether multiple exposures can increase it. In one of the results, the authors showed that the implicit bias decreased more for those with the black virtual body than for white ones. That is, showing that exposure can also increase the out-group advantage. In work by Hasler et al. [HSS17], the results showed that white participants who used the virtual dark skin color expressed greater mimicry when they saw virtual groups of dark skin colors. The authors mentioned that these white participants treated groups of black virtual humans from the environment as if they were their in-group, while the group of white virtual humans was treated as an out-group. Again, compared with the work of Saneyoshi et al., this case also reports an out-group advantage. Obremski et al. [OAL+23] presented an approach to reduce implicit and explicit racial bias using virtual agents in VR. Their results showed that both the implicit racial bias and the explicit bias of the participants were reduced after interacting with virtual agents. In this case, the results of Obremski and colleagues work show the importance of VHs in reducing bias. In the work of Zhao et al. [ZMAV+23], for five types of race/ethnicity of virtual agents, participants were more accurate in identifying the race/ethnicity of White and Black virtual agents than Asian, Hispanic and Indian virtual agents. In other studies, we also can see perception of blendshapes in characters with black and white skin colors (McDonnell et al. [MZCD21]), perception of faces of various ethnic groups (Krumhuber et al. [KST+15]), among others.

2.5 Gender Impact in VHs

This section is divided into two parts: *i)* Section 2.5.1 presents attributes that may be important for the perception and attribution of gender in relation to adults VHs; *ii)* Based on studying gender non-binarism (female and male) in relation to genderless VHs, Section 2.5.2 presents studies involving gender since childhood; *iii)*, and Section 2.5.3 presents studies involving interactions with VBs, and also gender bias in VBs.

2.5.1 Adults Observing and Interacting with Virtual Adults of Different Assigned Genders

According to Johansson [Joh73], humans can generally recognize and categorize human behavioral movements from little information. Emotions are also part of these human behaviors ([EF78, Ekm92, Bas78]), and emotional perception is also a topic related to VHs ([EHEM13, AWdS+21, TNC13]). Regarding gender, results from the literature show that the facial expressions of women and men can influence the attribution of gender ([MGCPV07, AMRM19]). The same happens when we discuss emotions ([ZHRM13, ZHRM15, DK22]) and gender ([MJH+07, ADM21, GFGW+23]) of VHs.

The gender perception of VHs is present in several CG studies. For example, in two studies, Zibrek et al. [ZHRM13, ZHRM15] investigated the perception of gender of virtual characters (obtained by real actors, being women and men) under different emotional conditions and attributed emotional feelings to the characters and measured the effect on people's perception. The results indicate that gender classifications are affected by the emotion displayed by the character, being more evident when the gender was stereotyped than by the character's walking movement. The participants' gender can also impact their interaction with VHs. For instance, studies conducted by Zibrek et al. [ZNO+22, ZNO+20b, ZNO+20a, ZHRM15, ZHRM13] have demonstrated that women and men tend to prefer different distances when interacting with adult VHs of various genders. In the work of Durupinar and Kim [DK22], results showed that female VHs are better perceived in convincing the expressed emotion when compared with male VHs. Bailey and Blackmore [BB17] assessed gender differences in the perception of avatars. The results indicated that gender is important in the perception of emotions. Tinwell et al. [TNC13] investigated whether inadequate superior facial animation in virtual human-like characters evoked a sense of strangeness due to a perception of psychopathic traits in these characters. In one of their results, the authors found that virtual male characters with no superior facial expression were classified as stranger than female characters, and psychopathic ratings were a strong predictor of this perceived strangeness. In general, the results showed that in general male characters were stranger than female ones. Seyama and Nagayama [SN07] investigated the UV by measuring human perceptions of facial images whose degree of realism was manipulated by the transformation between artificial and real human faces. The authors evaluated whether the gender classification of artificial and real humans could be an influencing factor for the UV. However, found no significant results. McDonnell et al. [MBB12] investigated the effect of render style on the perception of VHs. The authors used several rendering styles and analyzed perceptual features, such as appeal, familiarity, realism, and others. One of the results showed that people were more correct when they saw female virtual humans than male.

In terms of gender attribution, in the work of McDonnell et al. [MJH+07], the authors measured gender perception/and gender attribution of virtual models based on movement and appearance. Regarding appearance, the authors used female and male models and models without an assigned gender (androgynous and skeleton representation). Furthermore, in a second experiment, the authors changed the shape of the female and male models. Results showed that gendered features in the appearance of a VH could change the gender attribution. In the work of Zhao et al. [ZMAV+23], for binary gender types of virtual agents, participants were more accurate in identifying masculinity in male virtual agents, and identifying femininity in female virtual agents. In this line of gender attribution, Draude [Dra11] stated humans tend to assign gender to VHs in the same way they tend to assign gender to objects themselves, even without any indication of gender [MJH+07, MJH+09]. The studies of Nag and Yalçın [NY20], Ghosh et al. [GFGW+23] and Hess et al. [HAJK04] discussed that when there are some visual features (e.g., hair, facial features), people still show gender bias in gender attribution of a VH's gender and also about emotion recognition. This can be explained by Social Identity theory, in which people prefer to see themselves and their peers (social groups) on a positive spectrum [Bro20], which may be related to what was said in the previous paragraphs about in-group advantages. These cases shows that people in a group can have certain advantages and preferences in categorizing people in the same group [GBBM07, EA02, KST+15]. For instance, in terms of emotions and gender, as shown in the work of McDuff et al. [MKKL17], women tend to be more expressive, positively and negatively, than men. In the work of Abbruzzese et al. [AMRM19], the results showed that women might be better at recognizing emotions expressed by other women. At the same time, people in an out-group can fall into the case of the Dehumanization Theory ([Bro20]), i.e., people always try to humanize their peers but dehumanize those who do not resemble them. In an analogy, when this current work talk about dehumanization, for example, it is talking about reducing the anthropomorphic level in the perception of VHs if they are not part of the in-group.

In cases where there are no defined gender groups, gender anonymity can change the perspectives of the results. For example, the work of Morales-Martinez et al. [MMLD20] shows that gender non-identification can generate different perceptions and actions compared to when people know another individual's gender. The authors explored how anonymity/identifiability in user profile design impacts student interactions in a large multicultural classroom in two geographic locations. Regarding gender, the results showed that when students interacted with their real identities, there were gender effects (both significant and negative) that were absent when students were anonymous. In work by Kao et al. [KMJ+24], the authors investigated how gender-anonymous voice avatars influence women's performance in online computing group work. The results showed that when all avatars used masked voices, female participants spoke longer, spoke more words, and scored higher on computer problems compared to all avatars with unmasked voices.

2.5.2 Adults Observing and Interacting with Real Babies of Different Assigned Genders

According to [Sco07, Wal10], gender is a social and cultural construction built on stereotypes. For example, a pink shirt is assigned to a girl, while a blue shirt is assigned to a boy. So, a woman's self-identify as female had the construction of her feminine self based on a social standard of femininity. According to Condry and Condry [CC76], babies are considered genderless until a certain age. In other words, before that, the gender assignment is given to the baby by adults. In this sense, in the work of [WSD76], the authors carried out a perceptual experiment with two groups of mothers, one group receiving a male baby with a male name and wearing an outfit considered male (wearing a blue outfit), and the other group receiving the male baby with a female name and dressed in clothing considered feminine (wearing a pink outfit). The authors informed the participants that they would be in a room with the baby, three toys assumed sex-appropriateness (a male, a female, and a neutral toy), and asked them to play with the baby for eight minutes. The authors measure the interactions through observations made during the experiment, being divided into participant actions and baby actions. Regarding the participating actions (which are the ones of interest for our work), the authors observed toy handling (for example, preference for the female toy), auditory, visual (for example, smiling), and tactile stimulation. In the work of Smith and Lloyd [SL78], the authors revisited the work of Will et al., and in addition to the analyses carried out in the first study, the authors evaluated the duration of use of the toy. Regarding the results, in both studies, there were trends that mothers presented more female toys to a baby with a female name and presented more male toys to a baby with a male name. The time of use of one of the male toys was greater when it was related to a baby with a male name. In addition, mothers tended to smile more at the baby with a female name.

Unlike the studies presented in the previous paragraph, the work by Culp et al. [CCH83] carried out a study to measure interactions between parents and babies, that is, measuring whether the participant's gender also influenced the results. In a 10-minute experiment, the actions of the adults recorded in relation to the interaction with the baby were vocalization-verbalization, gaze direction, facial expression (smile, frown, animated, "ooo" face, neutral), physical contact, and toy used (having toys considered female, male, and neutral). In some results, overall, the baby with a female name received more verbal interaction, less gaze direction, and more neutral faces. Women chose the feminine toy more than men, especially when the baby had a feminine name. The men smiled more towards the babies than the women. In the studies proposed by Seavey et al. [SKZ75] and Sidorowicz and Lunney [SL80], the authors presented a baby without visual indications of gender to three groups of participants, one group receiving a baby with a female name, a baby with a male name, and an unnamed baby. The studies lasted around three minutes, and the authors evaluated the use of the toy and the physical interaction between the participant and the baby. Participants were not limited to just being parents as in other studies. In the work of Seavey et al., one of the results showed that the female toy was used more than the others. In addition, both women and men used the female toy more when the baby had a female name, while most participants thought the unnamed baby was a boy. In one of the results of the study by Sidorowicz and Lunney, most women and men chose more the male toy for the baby with a male name and chose more the female toy for the baby with a female name. Furthermore, it was also found a study that measured whether gender influenced the closeness between babies and adults [Gre89]. The study of Greeno suggested that the genders of the adult and the baby are factors that can influence the proximity between the adult and the baby, for example, women getting closer to girls.

2.5.3 Adults Observing and Interacting with Virtual Babies of Different Assigned Genders

There are some studies that evaluate interactions between humans and VBs. These studies involved baby feeding training [TGG22, TGG21, PCH+10, PCC+10], haptic interaction [HMA+17, MHA+17], and speculative scenarios for parent training [SCD+22]. In fact, these studies show us that it is important to measure interaction between humans and VBs for training future scenarios. Until the writing of this present work, there is no research carried out on human-VB observation or interaction that involves evaluating gender biases in the attitudes of humans with different genders towards VBs with different genders. This current work aims to have participants observing and interacting with a VB as in studies conducted with real human adults and babies ([CC76, WSD76, CCH83, SKZ75, SL80]). So, our goal in the present work is to compare real and virtual observations/interactions of adults with babies (real and virtual ones). To do that, as mentioned in Chapter 1, this current work recreated the experiment by Condry and Condry [CC76] using virtual babies and toys, animations, and virtual environments.

2.6 The Importance of the Studies Presented in the Related Work

This section illustrates how the studies presented in this chapter relate to the present work.

The study by Flach et al. [FdMM+12] is the basis of the studies involving UV carried out in Chapter 3. Firstly, in Section 3.1, the work of Flach and colleagues was recreated with the same VHs used, in addition to more recent VHs. With this, as the intention was to use VHs from different media, that is, VHs considered ready for the audience (characters from films, games, etc.), it was possible to compare the perception of comfort and realism (UV theory) between different years. Furthermore, it was possible to have a dataset with a considerable and similar number of female and male VHs to perform the gender comparison in Section 3.2. Finally, in Section 3.3, to make comparisons involving VHs' skin color, the questionnaire design used in the work of Flach et al. was also used to capture people's perception of VHs with different skin colors. Furthermore, the studies presented in Section 2.4 were important for the basis of the perceptual study on skin color discussed in Section 3.3.

As mentioned in Chapter 1, in this present work, more specifically in Chapter 4, the study by Condry and Condry [CC76] is replicated, which presented videos of a real baby interacting with objects (toys) to two groups of participants. In order to assess gender attribution and bias, the same baby was presented with a name and different gender for each group of participants. The questionnaire design of Condry and Condry's study is important to all of the studies conducted in Chapter 4. First, Section 4.1 replicates Condry and Condry's study by using a virtual baby. Such baby was sent with an assigned gender and name for two groups, while a third group received the same baby without gender and name specification. In Section 4.2, based on the experiments of interactions between babies and adults presented in Section 2.5.2, an interactive experiment was carried out with the same virtual baby and the same Condry and Condry questionnaire design. In Section 4.3, using the same questionnaire design of Condry and Condry, to assess the influence of gender on the perception of emotions and gender attribution in adult VHs, the present work was also based on experiments by Zibrek et al. [ZHRM13, ZHRM15]. These experiments used animations based on real actors (male and female) behind female and male adult VHs. In addition, Section 2.5.1 also helped to understand human perception regarding adult virtual humans, and not baby, which involved gender and emotions.

3. PERCEPTION OF VIRTUAL HUMANS FROM DIFFERENT MEDIA

This chapter aims to to present some analyzes regarding the perception of VHs/CG Characters from different media (movies, series, games, simulations, etc) in terms of gender and skin color biases. In this sense, this chapter is divided into four parts: *i)* Section 3.1 presents a perceptive study about CG characters/VHs designed in different periods of time in order to compare the perceptions of 2012 vs. 2020. Furthermore, this work was important to create a dataset containing characters with different skin colors and genders; *ii)* Section 3.2, discusses some comparisons that were studied between women's and men's perceptual data perceptual data about female and male characters; *iii)* Section 3.3, discusses some comparisons that were studied between dark and white colored skin characters from different media; *iv)* and finally Section 3.4, which presents considerations and discussions for each section.

3.1 Perception of CG Characters in Relation to the Advance of Technology

This section aims to try to answer the following question:

- Are people feeling more comfortable with newer CG characters than older ones?

To answer this question, the experiment by Flach et al. [FdMM+12, DFH+12] was recreated using the same questionnaire containing the same images and videos, as presented in 2012. Flach's work was chosen because the images and videos are still available and they present a varied origin (games, movies, internet), and because had diversity in terms of gender and skin color characteristics in relation to the CG characters/VHs. With this, a comparison analysis between the perception of ten years ago and the current perception data regarding the UV effect was made. Because of this, the questionnaire created in 2012 [FdMM+12] was used to allow a fair comparison, but also including images and videos of more recent characters. The main goal is to observe the UV effect with the new characters compared to the ones from the previous work [FdMM+12].

To facilitate the interpretation, it was necessary to separate the perceptual data into three groups: *i)* old characters (O), represented by images and videos that were used in Flach's work [FdMM+12] in 2012; *ii)* new characters (N), which are also comprised by images and videos included in this work, but from the last six years before 2020 (year that data collection was performed); and *iii)* all characters (A), i.e., the full set with old and new characters. To make this text easier to read, O and N were used to represent the perceived stimuli in 2012 and 2020, respectively. Therefore, during the text, tuples were used to refer to the evaluated group of characters and the period of analysis; for example, (O, O) represents old characters and stimuli evaluated in 2012. With this, five hypotheses were raised:

- $H0_1$ defining that $(O, O) = (O, N)$, i.e., the perceptual data obtained in 2012 indicate similar comfort with data obtained in 2020, about the characters from 2012;
- $H0_2$ defining that $(O, O) = (A, N)$, i.e., the perceptual data obtained in 2012, w.r.t. 2012 data, indicate similar comfort with data obtained in 2020, concerning all characters (from 2012 and 2020);
- $H0_3$ defining that $(O, O) = (N, N)$, i.e., the perceptual data obtained in 2012, w.r.t. 2012 data, indicate similar comfort with data obtained in 2020, about new characters;
- $H0_4$ defining that $(O, N) = (N, N)$, i.e., the perceptual data obtained in 2020, w.r.t. 2012 data, indicate similar comfort with data obtained in 2020, about new characters. Also, this hypothesis was performed for realism;
- $H0_5$ defining that variation of comfort (difference) $|(O_{image}, O) - (O_{video}, O)|$ is similar to $|(N_{image}, N) - (N_{video}, N)|$ and also similar with $|(A_{image}, N) - (A_{video}, N)|$, where *image* and *video* refer to specific domain instead of global data.

This section is divided into Section 3.1.1, which presents the methodology, and Section 3.1.2, which presents the results.

3.1.1 Methodology

The Section 3.1.1 has three parts: Section The CG Characters of 2012 and 2020 presents all characters used in this part of the work. Section The Questionnaire presents the proposed questionnaire related to comfort, realism, and familiarity with the CG character, and finally, Section Creating the Comfort Chart defines how the Comfort chart is modeled. Indeed, the Y-axis has several interpretations (comfort, eeriness, creepiness, strangeness, familiarity, likability) [KFMT15, MGHK09], in the case of this work, the term **Comfort** is considered. The X-axis states for **Human Likeness** and it does not have an universally accepted way for its categorization. In Comfort chart, artificial beings are categorized by their levels of realism, or human likeness (X-axis), being, from left to right, the least similar to the most similar. The Y-axis, on the other hand, presents the perceptive values of the comfort of people about the evaluated artificial beings.

The CG Characters of 2012 and 2020

To try to answer the question "Are people feeling more comfortable with newer CG characters than older ones?", the work of Flach et al. [FdMM+12] was reproduced. The same set of ten characters was used, comprised by images and videos, which are listed in Figure 3.1 from (a) to (j). In addition to the CG characters from 2012, 12 CG characters were added, which were from 2013 to 2020, shown from (k) to (v). With these data, as proposed in [FdMM+12], the human likeness criterion was evaluated, which contributes to the order the characters are placed in the horizontal axis of the UV chart (detailed in Section Creating the Comfort Chart). To ensure the variation of human likeness, some characters, which could represent a human being more realistically, were chosen, as shown in the cases (k), (p), (t), and (v) in Figure 3.1. Therefore, characters that escape from realism were needed (i.e., stylized, unrealistic, badly designed, etc.), counteracting the others cited earlier, such as the cartoon characters shown in Figure 3.1 (m), (q), (s), and (u). This counterpoint is necessary to form the horizontal axis of the UV chart. All pictures and movies were obtained from the YouTube platform. It was necessary to limit the search to videos with a large number of views and descriptions that contain copyright to avoid amateur videos. [1]

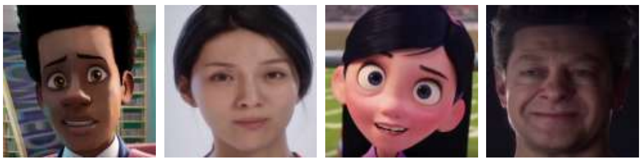
In addition to the realism factor, other restrictions were followed as proposed in [FdMM+12] concerning the choice of each character: *i)* the character has to represent a human being (i.e., avoiding animals, for instance); *ii)* it should not be placed in an unreal place; *iii)* the character should wear normal (and not) minimal clothes to avoid distortions in perceptions; and finally *iv)* the scene should be fo-

[1] Copyrighted images reproduced under "fair use policy".

(a) Internet (b) Internet (c) Internet (d) Movie: The Incredibles 1 (e) Internet: Obama's Cartoon (f) Internet

(g) Movie: Cloudy with a Chance of Meatballs (h) Movie: Beowulf (i) Game: Heavy Rain (j) Movie: Rango (k) Internet: Unreal 4 Engine (l) Movie: Alita

(m) Movie: How to train your dragon 2 (n) Movie: Thor Ragnarok (o) Movie: Rogue One (p) Series: Love, Death and Robots (q) Movie: Moana (r) Game: Overkill's The Walking Dead



(s) Movie: Spider-Verse (t) Internet: Unreal 4 Engine (u) Movie: The Incredibles 2 (v) Internet: Unreal 4 Engine

Figure 3.1: All characters used in this work. From (a) to (j) there are the characters used in the work of Flach et al. [FdMM+12], and from (k) to (v) are the most recent characters added in this paper (all the characters' pictures and short sequences have been taken from Youtube videos).

cused on the character's face so that the participants could catch the movement of the mouth, the eyes, among other expressions. All of these restrictions were used to avoid possible negative influences on human perceptions. Additionally, comparing with the base dataset, this work tried to include more diversity in terms of female gender (characters k, l, o, p, q, t, u) and dark colored skin (q, r, s).

Regarding people's familiarity with the tested characters, this work tried to choose characters that could be considered known to the general public (e.g., from movies) as characters (l, m, n, o, q, s, u), as done in [FdMM+12]. Furthermore, little-known characters were also included, e.g., k, t, and v, as in [FdMM+12].

The Questionnaire

First, just as it was necessary to compare human perceptions of CG characters with the results obtained by Flach et al. [FdMM+12], it was used the same structure of five questions of the reproduced study, as presented in Table 3.1. Following this structure, questions with categorical answers were used, where Q2 and Q4 are explicitly used to build the comfort chart (based on UV chart). Q2 was the only one with three possible answers (based on Likert Scales with three scores), and it was used to measure the participants' perception of realism regarding the characters. Q2 answer options were used to define the order the characters in the Comfort chart (X-axis). Q4 aimed to measure the perceived comfort quantitatively and indicates values in the chart (Y-axis). The entire questionnaire was assembled using Google Forms, and before answering the survey, participants received no explanation about the original intent of the research. This was done to avoid any type of influence on the participants' responses. Furthermore, it was necessary to use demographic and CG familiarity questions to try to avoid bias. Also, the ethical guidelines were followed for the application of the questionnaire, where all participants were asked if they agreed to grant access to their answers and personal information regarding age, gender, educational level, and familiarity with CG.

The process was divided into two steps, in which the presented characters (from Figure 3.1) were selected randomly. In the first step, an image of each character was shown before all questions, as referred to in Table 3.1. In the second step, performed just after the first step, a video of each character was shown before asking the same five questions. These steps evaluate people's comfort level when observing characters in the pictures, where they are static, and also in the videos,

Table 3.1: Questions regarding human perception applied to the participants.

Question	Possible answers
Q1: Do you think that the character in the picture/video above is:	*a*) A real person *b*) Created using CG
Q2: If created using CG, how realistic does it seem?	*a*) Very realistic *b*) Moderately realistic *c*) Unrealistic
Q3: Do you know this character?	*a*) Yes *b*) No
Q4: Do you feel some discomfort (strangeness) looking to this character?	*a*) Yes *b*) No

where characters are moving (animations). The level of comfort is asked in question Q4, as shown in Table 3.1.

Creating the Comfort Chart

As mentioned before, the Comfort chart is represented by a 2D plot, where the X-axis indicates the level of character realism from less to more realistic (from left to right, having higher values for realism on the right). The Y-axis defines the perceived comfort (%) of people when watching the characters. It goes from less to more comfortable, where less comfortable is associated with small values in the Y-axis. Only positive values were used in both axes.

To define the order of the characters in the horizontal axis, the averages of scores of Q2 answers were used. Thus, each character presents an average value of realism. As a result, the Human Likeness axis is shaped by the increasing order of each character's realism value. It was tried to follow this same methodology of ordering CG characters on the horizontal axis. In addition to such analyses, a comparison was made between the characters by levels of realism. According to Katsyri et al. [KFMT15], at least three levels of human likeness are necessary for comparisons between them regarding the levels of realism. The characters were divided into three levels of realism based on the three possible answers from Q2 (i.e., "Unrealistic", "Moderately Realistic", and "Very Realistic") according to the following rule: *i)* characters defined as unrealistic when having realism values ≤ 1.5; *ii)* characters defined as moderately realistic when having values of realism ≤ 2.5; and *iii)* characters defined as very realistic when presenting realism values > 2.5.

Figure 3.2 shows the final order of the 22 images in X-axis, and the average percentages of answers “Unrealistic” (blue line), “Moderately Realistic” (red line), and “Very Realistic” (yellow line) of all characters on Y-axis. The green line presents the percentages of answers “Created using CG” (question Q1 in the form) of all characters. Therefore, it is interesting to see that on the right side of Figure 3.2, subjects seem to be more confused when asked to define whether or not the very realistic characters were created using CG. Especially if we look at the character (r), who was the only one with dark colored skin among the group of very realistic CG characters.

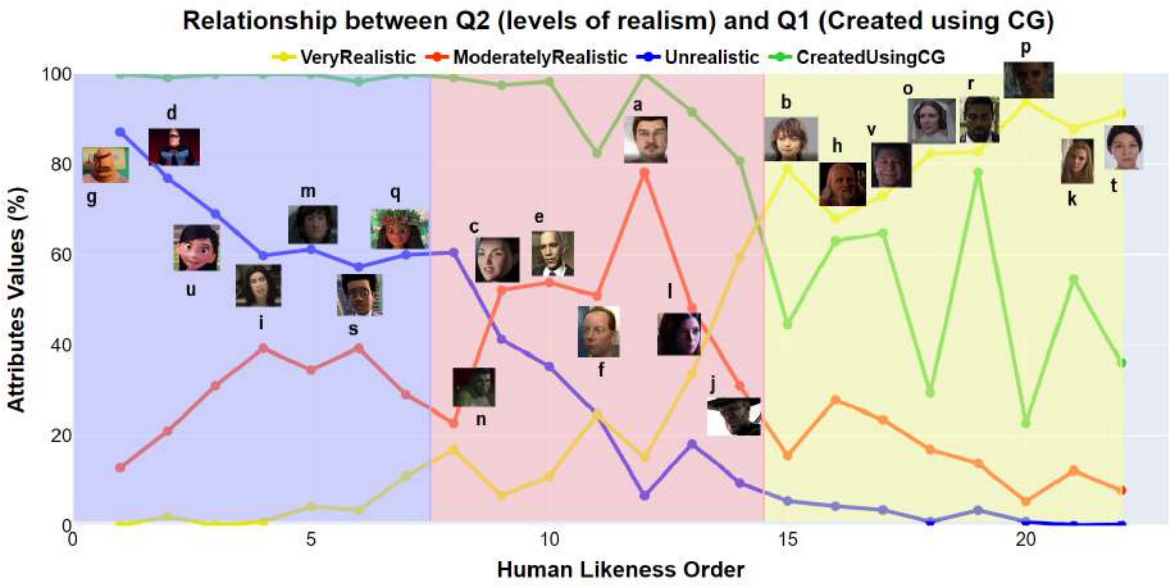


Figure 3.2: The final order of the Human Likeness axis (horizontal) of this work is presented. The vertical axis shows the percentage of answers “Unrealistic”, “Moderately Realistic”, “Very Realistic” (Q2); and "Created using CG" (Q1) of each character, respectively represented by the blue, red, yellow, and green lines. In addition, the background of each group of realism is highlighted using the same colors used in the captions of Q2’s answers.

Once the horizontal axes data was generated, the perceived comfort was used to compose the Comfort chart (remember that the comfort chart was based on the UV chart). The vertical axis (Comfort) is given by the percentage of the “No” answers to the Q4 question, which yields larger values if the character presents more comfort. The following section presents the results.

3.1.2 Results

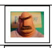
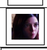
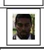
The results discussed in this section were obtained with the questionnaire shown in Section 3.1.1. The questionnaire was applied on social networks, and all participants were volunteers. It was answered by 119 participants, from which 42% were women and 58% men; 77.3% were at most 30 years old, 63.8% completed high school, and 68.1% were familiar with CG [2]. In all statistical analyses, it was used 5% of significance level (paired and independent T-tests, and *Chi-square* test). In addition, the terminology was used as defined above, using tuples where the first element is related to the data and the second to the perception. For example, (O, N) represents data acquired in 2012 (Old characters) and evaluation performed in 2020 (New evaluation). Table 3.3 presents all data used in this work, and the next sections discuss the results of this work.

Analysis of Perceived Realism of CG Characters

As shown in Section 3.1.1, question Q2 evaluates realism through answers using 3-Likert Scales. With that, each character had an average value of perceived realism, as shown in ascending order in the last column of Table 3.2, that is, following the order of Human Likeness. In addition, the average of perceived realism for each group of characters was also computed, i.e., A, N, Old and New, and according to the levels of realism (Unrealistic, Moderately Realistic, and Very Realistic), as shown in the last three lines of Table 3.3. With these data, two types of statistical analysis were performed with these values: *i)* Between levels of realism (for example, unrealistic x moderately realistic), in order to assess the differences between the groups of perceived realism; *ii)* and between the groups, (N, N) and (O, N), in order to compare the perceived realism between old and new characters. In all of these analyses, an independent T-test was used. With respect to all characters (A, N), significant p-values in all comparisons were found (< 0.001 for Unrealistic x Moderately Realistic, < 0.001 for Unrealistic x Very Realistic, and < 0.001 for Moderately Realistic x Very Realistic). Therefore, **for all characters, what can be said is that all groups of realism were different from each other.** Regarding the old characters (O, N), only in the comparison between the groups Moderately Realistic and Very realistic that no significant results were found (0.009 for Unrealistic x

[2]Flach's work does not contain demographic data, so, it was not possible to make comparisons taking into account demographic data

Table 3.2: The table presents all data (comfort, familiarity with CG, realism) in images and videos (data obtained in Flach's work presented as Old, comfort obtained in this current work presented as New) of all characters. Obs: All values are percentages, except the last column (Realism), which are averages.

Character Id	Character Figure	Image Comfort Old	Video Comfort Old	Image Comfort New	Video Comfort New	Image Familiarity New	Video Familiarity New	Image Realism New
g		61%	63%	52.1%	57.98%	63.02%	56.3%	1.10
d		88%	86%	84.87%	83.19%	92.43%	91.59%	1.23
u		-	-	67.22%	78.15%	73.1%	72.26%	1.29
i		29%	31%	24.36%	32.77%	3.36%	2.52%	1.38
m		-	-	88.23%	87.39%	80.67%	81.51%	1.39
s		-	-	89.07%	84.03%	48.73%	48.73%	1.43
q		-	-	94.95%	90.75%	78.99%	78.99%	1.49
n		-	-	71.42%	65.54%	96.63%	93.27%	1.56
c		50%	53%	26.89%	45.37%	2.52%	2.52%	1.65
e		74%	73%	65.54%	59.66%	83.19%	80.67%	1.75
f		35%	34%	35.29%	33.61%	2.52%	0.84%	1.91
a		27%	12%	41.17%	32.77%	1.68%	1.68%	2.08
l		-	-	37.81%	53.78%	31.93%	36.13%	2.1
j		65%	63%	77.31%	67.22%	21.84%	14.28%	2.43
b		70%	33%	68.9%	46.21%	3.36%	2.52%	2.5
h		88%	78%	73.1%	76.47%	14.28%	11.76%	2.54
v		-	-	79.83%	68.9%	15.96%	16.8%	2.6
o		-	-	92.43%	86.55%	71.42%	71.42%	2.67
r		-	-	81.51%	88.23%	6.72%	3.36%	2.72
p		-	-	91.59%	91.59%	17.64%	14.28%	2.73
k		-	-	91.59%	87.39%	1.68%	0.84%	2.78
t		-	-	85.71%	72.26%	4.2%	5.04%	2.79

Moderately Realistic, and 0.001 for Unrealistic x Very Realistic). Therefore, **for old characters, what can be said is that the Unrealistic group was different from the other two.** In the case of the new characters (N, N), only in the comparison between the groups Unrealistic and Moderately Realistic that no significant result were found ($<$ .001 for Unrealistic x Very Realistic, and $<$ 0.001 for Moderately Realistic x Very Realistic). Therefore, **for new characters, what can be said is that the group of Very Realistic was different from the other two.** In the comparisons between (O, N) and (N, N) (i.e., $H0_4$), a significant result was found between the Very realistic groups (.01). With that, looking at the averages in Table 3.3, **what can be said is that new characters from the Very realistic group were considered more realistic than the old characters from the Very Realistic group.**

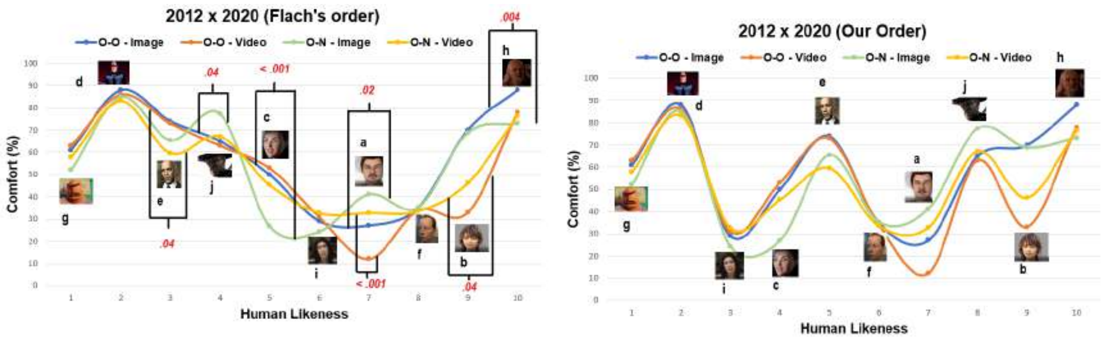
Table 3.3: Results of perceived comfort and realism (image and video for the first two, and only image for the last) of characters, in 2012 (*O*) and 2020 (*N*), and standard deviation values. For comfort, percentages were used, while for realism averages were used.

Dataset Comfort	Evaluation Time	Percentage All	Standard Deviation (SD)	Percentage Unrealistic	Standard Deviation (SD)	Percentage Moderately Realistic	Standard Deviation (SD)	Percentage Very Realistic	Standard Deviation (SD)
O_{image}	*O*	58.70%	22.72%	59.33%	29.53%	50.2%	19.71%	79%	12.72%
O_{video}	*O*	52.60%	24.12%	60%	27.62%	47%	24.3%	55.5%	31.81%
O_{image}	*N*	54.95%	21.95%	53.78%	30.28%	49.24%	21.28%	71%	2.97%
O_{video}	*N*	53.52%	18.34%	57.98%	25.21%	47.73%	15.41%	61.34%	21.39%
N_{image}	*N*	80.95%	16.06%	84.87%	12.13%	71.42%	37.81%	87.11%	5.56%
N_{video}	*N*	79.55%	11.93%	85.08%	5.37%	59.66%	8.31%	82.49%	9.43%
A_{image}	*N*	69.13%	22.74%	67.64%	25.6%	50.78%	20.07%	83.08%	8.88%
A_{video}	*N*	67.72%	19.86%	70.58%	20.88%	51.14%	14.28%	77.20%	14.99%
Dataset Realism	**Evaluation Time**	**Average All**	**Standard Deviation (SD)**	**Average Unrealistic**	**Standard Deviation (SD)**	**Average Moderately Realistic**	**Standard Deviation (SD)**	**Average Very Realistic**	**Standard Deviation (SD)**
O_{image}	*N*	1.86	0.52	1.24	0.13	1.96	0.3	2.52	0.02
N_{image}	*N*	2.13	0.64	1.4	0.08	1.83	0.38	2.71	0.07
A_{image}	*N*	2.01	0.59	1.33	0.13	1.93	0.3	2.67	0.1

Analysis of Perceived Comfort of CG Characters - Comparing (*O*, *O*) and (*O*, *N*) by Character:

Firstly, it is essential to mention that Flach's work used another ordering criterion for the Human Likeness axis, based on the evaluation performed in 2012. The two ordering schemes in Figure 3.3 were compared, where (a) shows both evaluations performed in 2012 and 2020 concerning data from 2012, using Flach's order. On the other hand, Figure 3.3(b) shows the same data but using the new data. When comparing the two orders, it is interesting to note that only characters (g) and (d) (the first two characters in the two charts) remained in the same positions. The characters (j) and (i) were the ones that moved the most, the first being in the fourth position in Flach's work, becoming more realistic in perception in 2020, while the second dropped three positions. Also, one can see that the Valley is present in Figure 3.3(a) (Flach's order), while in (b), there was more than one Valley. While Flach [FdMM+12] defined the order on (*X*-axis) based on data observed in 2012, scores from Q2 answers were used, as evaluated in 2020.

Regarding comfort analysis in Figure 3.3, comparing only images, the green line (evaluation in 2020) was only above the blue line (evaluation in 2012) in characters (a) and (j), i.e., the perceived comforts of all other characters have decreased or remained the same, over the years. Assuming that perception responses from 2012 and 2020 are independent data, the *Chi-square* test was used and perceived comfort was evaluated in image and video for each character. In Figure 3.3(a), in part above the lines, the significant results (highlighted in red) of the differences be-



(a) Flach's order

(b) Our order

Figure 3.3: All the characters used in the work of Flach et al. [FdMM+12] with Flach's order in (a), and new order in (b). Both blue and orange lines, in Figure (a) and (b), represent the percentages of comfort of each character in image and video, as perceived in 2012. The green and yellow lines represent the same in (b), however evaluated in 2020. In addition, in (a) we can see the significant results (highlighted in red) of the comparisons of the characters perceived in 2012 and 2020 (the results related to images were above the lines, the results related to videos were below the lines).

tween characters from 2012 and 2020 in images are presented, while the significant results of the videos are presented below the lines. In addition, as what is can also be seen in Table 3.2, characters (c) and (h) were perceived as more comfortable in 2012 and (a) and (j) were more comfortable in 2020. In videos, character (e) was more comfortable in 2012, and characters (a) and (b) were more comfortable in 2020. In conclusion, **only** $4/10$ **characters present differences in perceptual data in images, and** $3/10$ **in the video, and only one in both media.**

Analysis of Perceived Comfort of CG Characters - Comparing (O, O) and (O, N), globally:

Regarding hypothesis $H0_1$ $((O, O) = (A, N))$, a paired T-test was performed to compare perceptual comfort percentages assessed in 2012 and 2020 (shown in Table 3.3). However, no significant results (image and video) were found, i.e., **in general, the comfort perceived by the participants is similar when comparing the assessments made in 2012 and 2020 concerning the old characters.**

In another analysis, an independent T-test was used to compare comfort ratings for each level of realism ("Unrealistic", "Moderately Realistic", and "Very Realistic") separately. As Flach's work did not provide levels of realism, only the new order (Human Likeness) was used in this analysis. However, no significant results were found either. With that, **what can be said is that both in 2012 and in 2020, the perceived comfort about characters from 2012 was not influenced by levels of realism.**

Analysis of Perceived Comfort of CG Characters - Comparing (O, O) and (A, N):

Comparing the perceived comfort between the 10 characters in 2012 and 22 characters in 2020 ($H0_2$) (the chart of all characters is shown in Figure) 3.5, no significant p-values were obtained (for image or video). Hence, **the results indicated that considering old and new characters, people in 2020 felt as comfortable as people in 2012 about old characters.** An independent T-test was also used for the analysis of all characters (A, N) at different levels of realism. In this case, significant results were found in the comparisons between moderately realistic and very realistic characters (0.001 in images and 0.004 in videos), and between unrealistic and moderately realistic characters (0.03 in images). Therefore, **what can be said is that for all characters, in image analysis, people felt more comfortable with very realistic characters than with moderately realistic ones. In videos, people felt more comfortable with unrealistic and very realistic characters than with moderately realistic ones.** These conclusions are in accordance with Mori [MMK12], in the UV theory.

Analysis of Perceived Comfort of CG Characters - Comparing (O, O) and (N, N):

Regarding $H0_3$ ($(O, O) = (N, N)$) (the chart of the new characters is shown in Figure) 3.4, an independent T-test was also used to measure the difference between the obtained comfort percentages (shown in Table 3.3), and obtained the significant p-values 0.01 (images) and 0.002 (videos). As the comfort perceived in 2020 was superior to one reported in 2012, **we can conclude that people in 2020 felt more comfortable with current characters than people in 2012 with characters from 2012 (at that time).** Regarding levels of realism for (N, N), significant results were found in the comparisons between moderately realistic and very realistic characters (0.01 in images and 0.02 in videos), and between unrealistic and moderately realistic characters (0.009 in images). Therefore, **what can be said is that for new characters, in the analysis of images, people felt more comfortable with very realistic characters than with moderately realistic characters. In the videos, people were more comfortable with unrealistic and very real-**

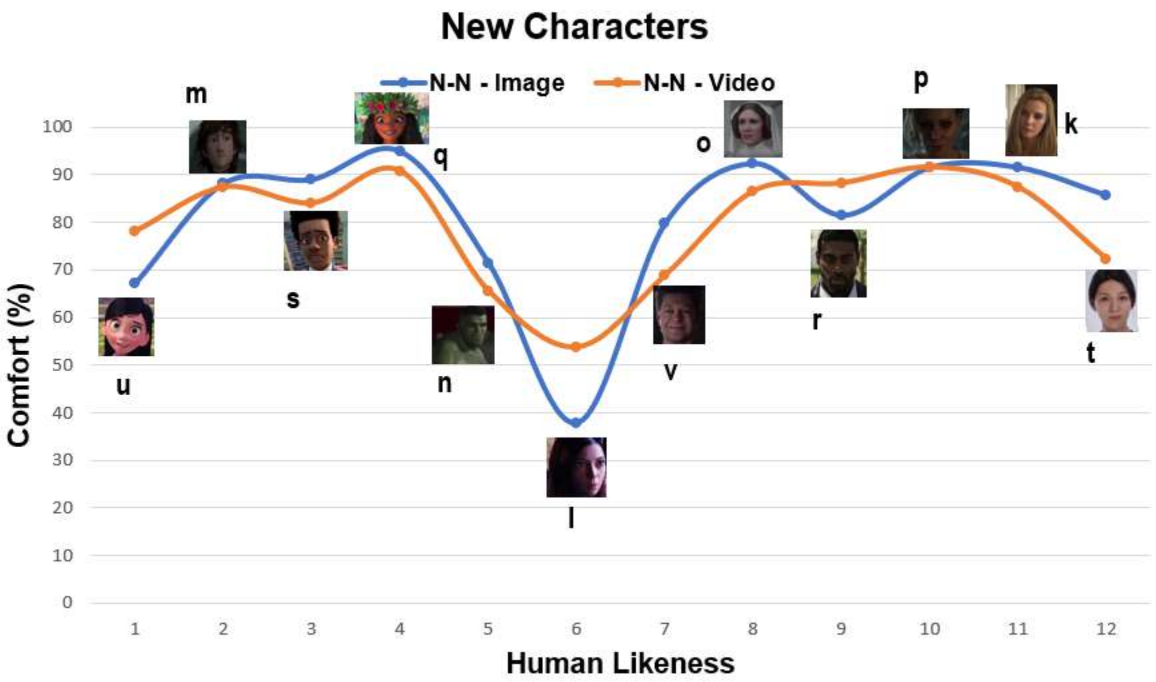


Figure 3.4: Recent characters evaluated in this work. The blue and orange lines represent the comfort percentages of each of the new characters w.r.t., image and video.

istic characters than with moderately realistic characters. As for (A, F), these conclusions were in accordance with the UV theory.

Analysis of Perceived Comfort of CG Characters - Comparing (O, N) and (N, N):

Regarding $H0_4$, an independent T-test was also used, and again obtained significant p-values (0.004 for images and < 0.001 for videos). Hence, **we can conclude that people are more comfortable with the new characters than with the old ones**. In the analysis of the three levels of realism in HO_4, a significant difference was found between very realistic characters (0.009 in images) through the independent T-test. With that, **what can be said is that people in 2020 felt more comfortable with very realistic new characters than with very realistic characters from 2012.**

Comparing Movement Effect in the Perceived Comfort Between 2012 and 2020:

To measure the theory of movement proposed by Mori [MMK12], $H0_5$ was defined: $|(O_{image}, O) - (O_{video}, O)| = |(N_{image}, N) - (N_{video}, N)| = |(A_{image}, N) - (A_{video}, N)|$,

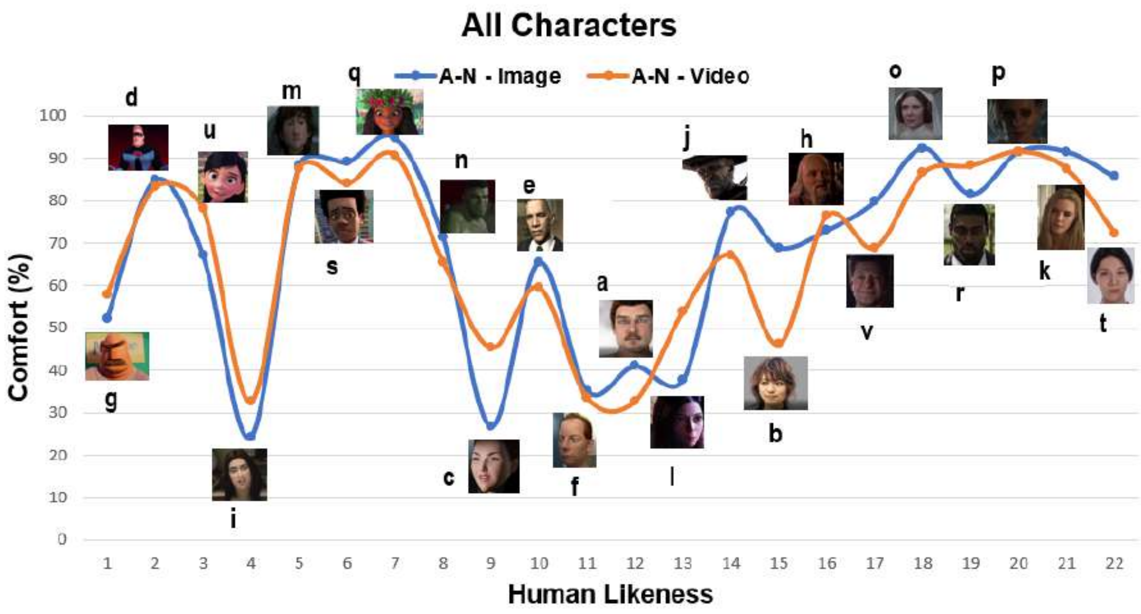


Figure 3.5: All Characters evaluated in this work. Characters (g), (d), (i), (a), (e), (c), (f), (j), (h) and (b) were the 10 characters that were also evaluated in 2012 [FdMM+12]. The blue and orange lines represent the comfort percentages of each character analyzed in 2020.

that is, the difference between image and video comfort perception is similar over the years. The difference module of comfort between image and video, for each one of the performed analysis, are as follows:

1. $|(O_{video}, O) - (O_{image}, O)| = 7.5\%$, (SD = 11.34%);

2. $|(A_{video}, N) - (A_{image}, N)| = 7.75\%$, (SD = 5.82%);

3. $|(O_{video}, N) - (O_{image}, N)| = 8.65\%$, (SD = 6.96%); and

4. $|(N_{video}, N) - (N_{image}, N)| = 7\%$, (SD = 4.87%).

To test $H0_5$, a paired T-test was used, but the results were not significant. Concerning the three levels of realism, the results were also not significant. So, it indicates that the characters' motion (videos) did not influence the perceived comfort, contrary to Mori's movement theory.

Qualitative Analysis of Perceived Comfort and Realism Involving Diversity in Terms of Gender

The perceived realism and comfort over the years (2012 and 2020) was also evaluated from a gender perspective. In the work of Flach et al., the 2012 dataset contained seven male characters and only three female characters. In our work, there were seven female characters and five male characters. Firstly, because of the small number of female and male characters in each group, only the averages and percentages were evaluated without statistical tests, that is, only qualitatively. So, the notations were as follows:

1. (O, O_{male}) = 73% of perceived comfort;
2. (O, O_{female}) = 49.66% of perceived comfort;
3. (O, N_{male}) = 59.29% of perceived comfort and 1.86 of perceived realism;
4. (O, N_{female}) = 40.05% of perceived comfort and 1.84 of perceived realism;
5. (N, N_{male}) = 82.01% of perceived comfort and 1.94 of perceived realism;
6. (N, N_{female}) = 80.18% of perceived comfort and 2.26 of perceived realism;

Looking at the perceptual data from 2012, we can see that the percentage of perceived comfort was higher for (O, O_{male}) than (O, O_{female}), with a difference of 23.34% between them. For 2020 perceived comfort data, the same happened but the difference was smaller, that is, $|(O, N_{male}) - (O, N_{female})|$ = 10.24%. Furthermore, still regarding the 2020 data, female and male characters had similar averages of perceived realism.

Comparing the years, both (O, N_{male}) and (O, N_{female}) had higher perceived comfort percentages than, respectively, (O, O_{male}) and (O, O_{female}). However, male characters had a higher percentage difference ($|(O, O_{male}) - (O, N_{male})|$ = 22.71%) than female characters ($|(O, O_{female}) - (O, N_{female})|$ = 9.61%).

Looking at characters from 2020, the difference in percentage of perceived comforts between female and male characters was low ($|(N, N_{male}) - (N, N_{female})|$ = 1.83%). The perceived realism for female characters was higher than for male characters (the difference was 0.32). Comparing 2020 data between characters from 2012 and 2020, in terms of realism, both (N, N_{male}) and (N, N_{female}) had higher realism values than, respectively, (O, N_{male}) and (O, N_{female}). However, female characters had a higher average difference ($|(N, N_{female}) - (O, N_{female})|$ = 0.42) than male characters ($|(N, N_{male}) - (O, N_{male})|$ = 0.08). In terms of perceived comfort, the difference between years in relation to female characters was also higher (40.13%) than in relation to male characters (31.71%).

Qualitative Analysis of Perceived Comfort and Realism Involving Diversity in Terms of Skin Color

The realism and comfort perceived over the years were also evaluated from the perspective of skin color, with characters divided into dark and white colored skins. The 2012 dataset contained eight characters with white skin color and only one character with black skin color (which was the Obama cartoon). The character (j) was removed due to the fact that the skin color was inconclusive. In our work (2020), there were eight white colored skin characters and only three characters with dark colored skin. The character Hulk (n) was removed because of his green skin color. Firstly, due to the small number of characters with different skin colors in each group, only averages and percentages were evaluated without statistical tests. Thus, the notations were as follows:

1. (O, O_{dark}) =74% of perceived comfort;
2. (O, O_{white}) =56% of perceived comfort;
3. (O, N_{dark}) =65.54% of perceived comfort and 1.75 of perceived realism;
4. (O, N_{white}) =50.83% of perceived comfort and 1.79 of perceived realism;
5. (N, N_{dark}) =88.51% of perceived comfort and 1.88 of perceived realism;
6. (N, N_{white}) =79.3% of perceived comfort and 2.29 of perceived realism;

Both looking at the 2012 data and the 2020 data (respectively, (O, O_{dark}) and (O, N_{dark})), the character (only Obama) with dark colored skin had higher average percentage of perceived comfort than the CG Characters with white colored skin (respectively, (O, O_{white}) and (O, N_{white})). Comparing Obama between years ($|(O, O_{dark}) - (O, N_{dark})|$), there was a small difference in perceived comfort, where people in 2012 felt 8.46% more comfortable than people in 2020. Between CG Characters with white colored skin there was also a small difference in the average percentage of perceived comfort (5.17% more perceived comfort for people in 2012).

As was the case with the 2012 characters, dark colored skin CG Characters from 2020 conveyed more comfort than white colored skin characters from 2020 ($|(N, N_{dark}) - (N, N_{white})|$, with a difference of 9.21%). In terms of perceived realism, the opposite happened, where white colored skin characters were perceived as more realistic than dark colored skin characters (with a difference value of 0.41). Comparing characters from 2012 versus 2020, both (N, N_{white}) and (N, N_{dark}) had higher perceived comfort and realism values than, respectively, (O, N_{white}) and (O, N_{dark}).

Analysis of Perceived Familiarity of CG Characters

Regarding familiarity with characters, statistical analysis was performed to compare percentages of familiarity for each level of realism separately, $(0, N)$ = (N, N) (also similar to $H0_4$), and analysis between image and video ($H0_5$). Regarding the comparisons between the levels of realism, significant results were found in (A, N) (0.004 in images and 0.005 in videos) and N, N (0.008 in images and 0.008 in videos) in the comparisons between the unrealistic and very realistic levels. So, **we can say that people were more familiar with unrealistic characters than very realistic characters, both in (A, N) and in (N, N).** In $(0, N)$ = (N, N), no significant results were found. Regarding images versus videos, significant results were also found in (A, N) (0.02 in all characters without separation into groups of realism) and in (N, N) (0.02 also in all characters). So, **we can say that people were more familiar with characters in images than in videos, both in (A, N) and in (N, N).**

3.2 Perception of CG Characters From a Gender Perspective

This section aims to try to answer the following questions:

- Do people perceive comfort and realism in female and male CG characters in the same way?
- Does this change when participants are separated into women and men?

To try to answer these questions, the perceptual data of comfort and realismcollected and presented in Section 3.1 are also used in this present section. This was done since in the previous section data were collected from women and men about different characters, also having female and male characters. With this, the following analyzes were performed: Initially, from A, the same amount of female and male characters were chosen, and the participants were separated into women and men. With that, the perceptions regarding female characters and the perceptions about male characters were compared quantitatively, and the perceptions of women participants and the perceptions of men were also compared. Below are the definitions to facilitate the interpretation of these analyses: Regarding comfort and realism, consider (PGe, CGe) being a tuple that returns perceptual data, where PGe indicates the participants' gender and CGe the characters' gender. Therefore, $PGe = W$, regarding the data answered by women participants, and M, representing men. In addition, $PGe = A$, all participants are considered. CGe is related to the character's gender, $CGe = F$ representing the female characters, $CGe = M$ representing the male characters, and $CGe = A$ all characters. With this, four hypotheses were raised, using the defined terminology:

- $H0_6$ defining that $(A, F) = (A, M)$, i.e., the perceptual data of all the participants about the female characters indicate comfortsimilar to the data related to male characters;
- $H0_7$ defining that $(W, F) = (W, M)$, i.e., the perceptual data of women participants about the female characters indicate comfortsimilar to that obtained about male characters;
- $H0_8$ defining that $(M, F) = (M, M)$, i.e., a hypothesis similar to $H0_2$, but in relation to men;
- $H0_9$ defining that $(W, A) = (M, A)$, i.e., the data obtained by women indicate comfortsimilar to that obtained by men, in relation to all characters.

This section is divided into Section 3.2.1, which presents the methodology, and Section 3.2.2, which presents the results.

3.2.1 Methodology

The methodology is divided into three parts: Section Female and Male CG Characters presents all female and male characters that are used in this part of this work, and Section Creating the Comfort Chart presents the details of the Comfort chart.

Female and Male CG Characters

Regarding the characters, the same number of female and male characters were selected from the character dataset that was created in Section 3.1. As can be seen in Figure 3.6, ten female characters were chosen, respectively, from character (a) to (j), and ten male characters, respectively, from (k) to (t). Thinking about the distribution of characters from left to right on the X-axis (Human Likeness) of the Comfort chart were chosen as follows: *i)* Unrealistic characters: female (h) and (j), and the male (l) and (s); *ii)* Moderately Realistic characters: female (a), (b), (c) and (e), and the male (k), (m), (n) and (q); and finally *iii)* the Very Realistic characters: female (d), (f), (g) and (i), and the male (o), (p), (r) and (t).

Creating the Comfort Chart

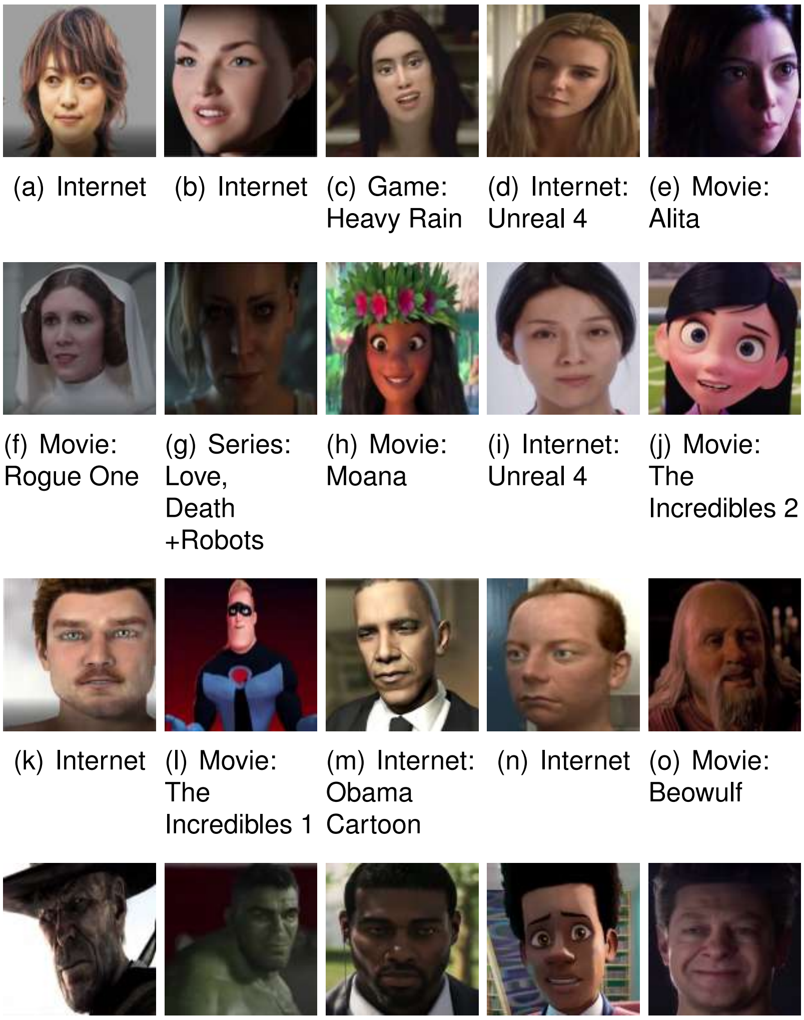
To build the Comfort chart in relation to the characters in Figure 3.6, the response data presented in Section 3.1 was used. With this, question Q2 covers the human likeness criterion, and subjects' answers influenced the order in which the characters are placed on the horizontal axis (Human Likeness), of the Comfort chart. To order the characters on the horizontal axis, the same methodology presented in Section 3.1 was used, both when the data were used in general and when they were separated into women and men.

3.2.2 Results

In this section, the (PGe, CGe) tuple defined earlier was used, to represent comfortand realism obtained from users perception. For example, (A, F) states for perceived data from all participants (A, that is, men and women) about the female characters (F).

Realism Analysis

Table 3.4 shows the levels of realism (first column), the groups of participants (second column), and the characters present at each level (third column). In addition, Table 3.4 also presents realism medians of all characters present at each level (fourth column), realism medians of the female and male characters (fifth and sixth columns, respectively).



(a) Internet (b) Internet (c) Game: Heavy Rain (d) Internet: Unreal 4 (e) Movie: Alita

(f) Movie: Rogue One (g) Series: Love, Death +Robots (h) Movie: Moana (i) Internet: Unreal 4 (j) Movie: The Incredibles 2

(k) Internet (l) Movie: The Incredibles 1 (m) Internet: Obama Cartoon (n) Internet (o) Movie: Beowulf

(p) Movie: Rango (q) Movie: Thor Ragnarok (r) Game: Overkill's TWD (s) Movie: Spider-Verse (t) Internet: Unreal 4

Figure 3.6: From (a) to (j) are the female characters, and from (k) to (t) the male characters. Remembering all the figures of the characters and short sequences have been taken from YouTube videos.

Table 3.4: The table presents the characters present in each level of realism, both in the perception of all participants, and in the perception of women and men. In addition, it also presents medians of perceived realism of all characters, female and male characters present in each level. Note: for this part of the work, values were not used for all participants and all characters (being A, A in the notation used in this work), since gender analysis was the focus.

Realism Level	**Obtained Data**	**Characters**	**All Characters (Median)**	**Female Characters (Median)**	**Male Characters (Median)**
All levels	All participants	All Characters	-	2.30	2.00
Unrealistic	All participants	(c, h, j, l, s)	-	1.38	1.33
Moderately Realistic	All participants	(b, e, k, m, n, p, q)	-	1.87	1.91
Very Realistic	All participants	(a, d, f, g, i, o, r, t)	-	2.73	2.60
All levels	Women	All Characters	2.03	2.24	1.95
Unrealistic	Women	(c, j, l, q)	1.24	1.23	1.28
Moderately Realistic	Women	(a, b, e, h, k, m, n, p, s)	2.84	1.8	1.84
Very Realistic	Women	(d, f, g, i, o, r, t)	2.7	2.76	2.6
All levels	Men	All Characters	2.13	2.34	2.07
Unrealistic	Men	(h, j, l, s)	1.34	1.37	1.29
Moderately Realistic	Men	(b, c, e, k, m, n, p, q)	1.89	1.69	2.05
Very Realistic	Men	(a, d, f, g, i, o, r, t)	2.69	2.69	2.6

A statistical analysis (using Python and the Scipy library) was performed with Wilcoxon signed-rank, Kruskal-Wallis and Mann-Whitney rank tests. These tests were used to try to avoid false positives due to the various analyzes with small samples. Therefore, it was not necessary to measure the normality of the data. Once there were a lot of tests to be performed (gender participants (2), gender characters (2), realism levels (3), image (1) = 12 in total), here just the tests which results are considered significantly using 5% of significance level are presented. The Wilcoxon test was used for paired samples with the same size, for example, women's perception of female characters x women's perception of male characters. For independent samples with different sizes, first the Kruskal-Wallis test was used to compare all three levels of realism, for example, women's perception of characters Unrealistic x Moderately realistic x Very realistic. In case of significant p-value in the Kruskal-Wallis test, the Mann-Whitney test was used in comparisons between two of the three levels of realism, for example, Unrealistic x Moderately, Unrealistic x Very, and Moderately x Very. In addition, the Mann-Whitney test was also used in comparisons between groups of equal realism, but with different sample sizes, for example, women's perception of unrealistic characters x men's perception of unrealistic characters, or women's perception of female characters unrealistic x unrealistic male characters.

Regarding all participants perception: With regard to all participants, in the perception of realism about female characters (A, F), a significant result was found (p-value = .021) when comparing the three levels of realism (Unrealistic x Moderately x Very), showing that the groups of CG characters are different in perception of realism. When comparing one group of realism with another (for example, Unrealistic x Moderately), significant results were only found when when there were a comparisons between the very realistic group with the Unrealistic group (.018), and with moderately (.039), that is, the very realistic characters group (higher percentage of realism, as shown in Table 3.4) was considered significantly different from the others. Comparing the three groups of male characters (A, M), the result was also significant (.022), also showing that the male characters' levels of realism were considered different. In the separate comparisons between groups, significant results were only found when there were comparisons between the unrealistic and moderately groups (.041), and between the moderately and very realistic (.018) groups. In these cases, **the results show that groups of CG characters are different in terms of realism perception in both female and male characters.** Comparing female and male characters (A, F x A, M), no significant results were found in the general comparison (without separations in levels of realism). Already separating the characters into levels of realism, a significant result was found when there was a comparison between very realistic female characters and very realistic male characters (.018), that is, **people considered the very realistic group of female characters more realistic than the group of male characters.**

Regarding women perception (W, A): Significant results were found when all groups were compared ($<$ .001), and also when the groups were separated (.003 for Unrealistic and Moderately, .005 for Unrealistic and Very, and .001 for Moderately and Very), that is, **for women, the groups of all characters were considered different in terms of realism.** The same happened when comparing all realism groups of female (W, F) characters (.019), and significant only in the comparison between the unrealistic and very realistic groups (.015). In the comparison of all groups of realism of the male characters (W, M), a significant result (.022) was also found. However, in the comparisons performed separately, significant results were only found when there were comparisons between between the unrealistic and moderately realistic groups (.041), and between the moderately and very realistic groups (.018). **In the comparisons between female and male characters (W, F x W, M) no significant results were found.**

Regarding men perception (W, A): In the analysis of men about all characters (M, A), significant results were found in the general comparison ($<$ .001) and in relation to separate comparisons (.004 for Unrealistic and Moderately, .004 for

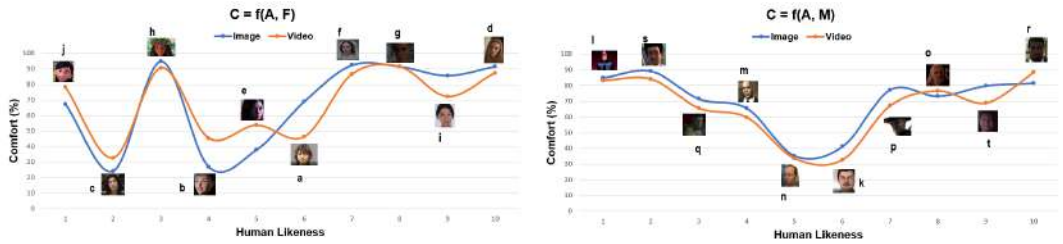
Unrealistic and Very, and $<$ .001 for Moderately and Very). On female characters (*M*, *F*) significant results were also found in the general comparison (.021), and significant in the comparisons between the unrealistic and very realistic groups (.039), and between the moderately and very realistic groups (.018). The same happened in the general comparison (.022) of the male characters (*M*, *M*), and significant in the comparisons between the unrealistic and moderately realistic groups (.041), and also between the moderately realistic and very realistic groups (.018). Therefore, **the results show that for men, the groups of realism of all characters, and the groups of female and male characters were considered different.** As mentioned earlier, analyzes between women's perceptions and men's perceptions (e.g., (*W*, *A* x *M*, *A*), (*W*, *F* x *M*, *F*), (*W*, *M* x *M*, *M*)) were performed. However, no significant results were found in any of the comparisons.

Comfort Analysis

Regarding perceived comfort, Table 3.5 shows the percentage of perceptual comfort and standard deviation obtained by the participants (all, women, and men) over the characters (all, female and male, and at all levels of realism) using the percentages of "No" responses. For all participants, the Comfort charts are shown in Figure 3.7, for the female characters in Figure 3.7(a) and for the male characters in Figure 3.7(b). Regarding the perceived data by women, Figure 3.8 shows the comfort charts for the female and male characters, 3.8(a) and 3.8(b) respectively. In relation to the data obtained by the responses of men, Figure 3.9 shows the Comfort charts. In addition, we can observe the *X*-axis orders of presented charts. The blue and orange lines represent, respectively, the comfort values obtained in the image and video analyses.

As in the previous section, a statistical analyses with Wilcoxon signed-rank, Kruskal-Wallis and Mann-Whitney rank tests were performed. For comfort analysis, there was also a large amount of tests to be performed (gender participants (2), gender characters (2), realism levels (3), video or image (2) = 24 in total), and a 5% level of significance was also used.

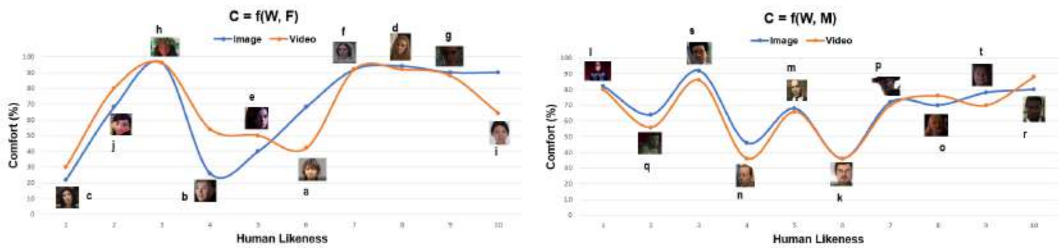
Regarding all participants perception: Firstly, with respect to the comfort perception of all participants about female characters (*A*, *F*), no significant result among all groups of realism was found (both in image analysis and in video), that is, realism did not influence the perceived comfort about female characters. Regarding the perception of comfort about male characters (*A*, *M*), significant results



(a) All Participants perceiving Female Characters - (A, F)

(b) All Participants perceiving Male Characters - (A, M)

Figure 3.7: The figure presents the charts of the perception of Comfort (X-axis) and Human Likeness (Y-axis), with respect to the female characters (a) and the male characters (b). The blue and orange lines represent, respectively, the perception of comfort in image and video.

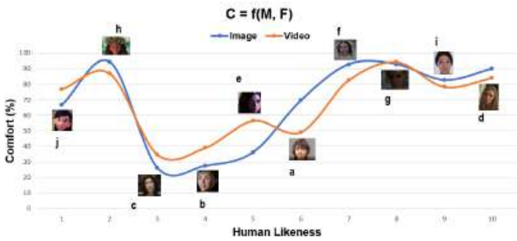


(a) Women participants perceiving Female Characters - (W, F)

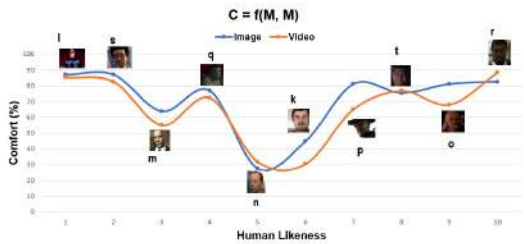
(b) Women participants perceiving Male Characters - (W, M)

Figure 3.8: The figure presents the charts of women's perception of Comfort (X-axis) and Human Likeness (Y-axis), in relation to the female (a) and male characters (b). The blue and orange lines represent the perception of comfort in the image and video, respectively.

were found in the general comparisons of the realism groups in the image (.033) and video (.032) analysis, showing that, **for people, the perceived realism influenced the perception of comfort about male characters.** In separate comparisons between groups, both in image and video, significant results were also found between the unrealistic and moderately realistic groups (respectively, .041 and .041), and between the moderately and very realistic groups (.037 and .018). As shown in Table 3.5, both the unrealistic and very realistic male character groups had higher comfort percentages than the moderately realistic group. With that, **the results show that people felt more comfortable with characters present in these two groups (Unrealistic and Very Realistic) than with characters present in the Moderately realistic group.** No significant results were found when images and videos were compared. Regarding the $H0_6$ hypothesis ((A, F) = (A, M)), both in general ana-



(a) Men participants perceiving Female Characters - (M, F)



(b) Men participants perceiving Male Characters - (M, M)

Figure 3.9: The figure presents the charts of men's perception of Comfort (X-axis) on Human Likeness (Y-axis), in relation to the female (a) and male characters (b). The blue and orange lines represent the perception of comfort in the image and video, respectively.

lyzes (without separating into groups of realism) and in analyzes dividing characters into groups of realism, in image and video, no significant results were found.

Regarding the perception of women about all characters (W, A): no significant results were found in the general comparison of groups of realism (both in image and video analysis), that is, realism did not influence the comfort perceived by women over all characters. With this result, no comparisons between groups separately were performed. No significant results were also found in the comparisons between image and video. As in the analysis of all characters, the same happened for female characters (W, F) and for male (W, M) characters, that is, without significant results. Therefore, for both female and male characters, realism did not influence perceived comfort by women. In addition, no significant results were also found in the comparisons of comfort in image and video. Regarding the comparison between female and male characters ($H0_7$ hypothesis - (W, F) = (W, M)), a significant result was found when very realistic female and very realistic male characters in images (.025) were compared. With that, **the results show that, in images, women felt more comfortable with very realistic female characters than very realistic male characters.** This result is in line with Draude's work [Dra11], that is, women can be more comfortable with female characters.

Regarding the perception of men, unlike the perception of women about all characters, significant results were found in images and videos (M, A) in the general comparisons of all groups of realism (respectively, .006 and .004). Therefore, **the results show that realism in this case influenced the perception of comfort.** In the separate comparisons, both in images and videos, significant results were found in the comparisons between groups of unrealistic and moderately realistic (.011 and .004), and between moderately and very realistic groups (.002 and .003). **So for men, unrealistic and very realistic characters were more comfortable than moderately realistic.** This result is in line with the empirical review by Katsyri et al. [KFMT15], which the authors presented a series of work that showed results of positive perceptual comforts on less realistic characters and more realistic characters, and negative results on moderately realistic characters. In addition, no significant results were found in the comparisons between images and videos. With regard to female (M, F) and male (M, M) characters, no significant results were found in the general comparisons of the realism groups, that is, having no influence on perceived comfort, and also no significant result in comparisons between images and videos. Regarding $H0_8$ ((M, F) = (M, M)), unlike the perception of women, no significant results were found in the comparisons between female and male characters. Therefore, for men, the gender of the characters did not influence the perception of comfort. Regarding $H0_9$, which was defined as (W, A) = (M, A), no significant results were found.

Table 3.5: Percentages of perceptual comfort and standard deviation obtained by the participants (all, women and men) over the characters (all, female and male) at different realism levels.

Participant Gender (*PGe*)	**Character Gender** (*CGe*)	**Image Comfort All**	**SD**	**Image Comfort Unrealistic**	**SD**	**Image Comfort Moderately**	**SD**	**Image Comfort Very**	**SD**
A	*F*	68.15%	28.39%	62.18%	35.56%	32.35%	7.72%	86.05%	9.95%
A	*M*	69.91%	18.05%	86.97%	2.97%	58.15%	18.76%	78.15%	4.44%
W	*F*	68.6%	29.18%	45%	32.52%	57.5%	31.04%	91.5%	1.91%
W	*M*	68.8%	16.84%	73%	12.72%	62.8%	22.16%	76%	5.29%
M	*F*	67.82%	27.90%	80.43%	19.47%	29.95%	5.48%	85.5%	9.82%
M	*M*	70.72%	19.78%	86.95%	0%	58.84%	22.46%	79.71%	3.83%
W	*A*	68.7%	23.19%	59%	25.84%	60.44%	24.79%	84.85%	8.93%
M	*A*	69.27%	23.59%	83.69%	11.85%	48%	22.81%	83.33%	8.27%
Participant Gender (*PGe*)	**Character Gender** (*CGe*)	**Video Comfort All**	**SD**	**Video Comfort Unrealistic**	**SD**	**Video Comfort Moderately**	**SD**	**Video Comfort Very**	**SD**
A	*F*	68.48%	21.96%	67.22%	30.49%	49.57%	5.94	76.8%	18.58%
A	*M*	65.96%	19.51%	83.61%	0.59%	51.76%	17.18%	77.87%	9.73%
W	*F*	68.8%	23.87%	55%	35.35%	60.5%	24.18%	84%	13.46%
W	*M*	66.4%	18.59%	68%	16.97%	58.8%	22.11%	78%	9.16%
M	*F*	68.26%	21.39%	81.88%	7.17%	43.47%	11.5%	77.68%	16.91%
M	*M*	65.65%	20.73%	84.05%	2.04%	51.01%	19.15%	77.77%	10.17%
W	*A*	67.6%	20.86%	61.5%	23.85%	59.55%	21.55%	81.42%	11.35%
M	*A*	66.95%	20.54%	82.97%	4.48%	48.18%	16.2%	77.71%	13.89%

Uncanny Valley Analysis

Relating with the presented Comfort charts, we can notice that the one in Figure 3.7(b)(A, M) has the indication of a valley (observe the characters n and k), while in Figure 3.7(a)(A, F), one can say there are two valleys. As can be seen, this may have happened because the characters (c) and (h), in Figure 3.7(a), may be incorrectly positioned in vertical axis. Although the orders are similar in X-axis, qualitatively the generated curves can be very different, as can be perceived, for instance, if we compare Figure 3.8(a) and Figure 3.9(a). Observing the charts in Figures 3.8((W, F)(a) and (W, M)(b)), we can see that only the first seems close to the UV theory, regarding the expected valley, since the second has several variations. Unlike such graphs, Figures 3.9(a) and (b) present perceived comfort by man participants, with respect to perceived comfort of female and male characters, through curves that resemble the structure of the UV chart [MMK12]. So, **assuming that the valley should exist, in these cases, the results show that men seem to categorize according to expected with the UV theory. Anyway, the results also show that men and women could perceive the realism of CG characters differently.**

3.3 Perception of CG Characters From a Skin Color Perspective

The goal of this section is to investigate people's perceptions about dark colored skin CG characters already established by various media to try to answer the question:

- Are comfort and realism similarly perceived in a CG characters with different skin colors?

As we saw in the previous sections (Figures 3.2 and 3.5, and also in the work of Flach et al. [FdMM+12, DFH+12]), the dark colored skin characters seem to be perceived with higher comfort values and lower realism values, in comparisons with the characters with white colored skin. Therefore, we raised the following hypotheses to be studied in this section:

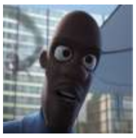
- $H0_{10}$ defining that the UV effect is similar for white and dark colored skin characters, with similar levels of realism. First, to test this hypothesis, we used the data of white colored skin characters from the Section 3.1. To compare these results, we recreated the experiment used in the Section 3.1 only with dark colored skin characters.

- $H0_{11}$ defines that UV's effect on different colored skin characters is similar for people with different racial identifications. By testing this hypothesis, we can discuss in-group and out-group advantages.

3.3.1 Methodology

The same dataset from Section 3.1 was used, which contains perceptual data about CG characters from different media. Therefore, only the perceptual data of CG characters with white skin were used from that dataset. In addition, the participants' perceptions were evaluated with respect to the dark colored skin characters. Therefore, this section is organized as follows: Section 3.3.1 presents the studied characters, and Section 3.3.1, presents the questions and the stimuli presented to the participants.

CG characters With Dark and White Colored Skins


(a) Movie: The Incredibles 2


(b) Movie: Soul


(c) Series: Arcane


(d) Game: GTA San Andreas


(e) Game: The Walking Dead from Telltale


(f) Movie: Encanto


(g) Movie: Spider-Verse


(h) Movie: Moana


(i) Game: True Crime New York City


(j) Internet: Obama's Cartoon


(k) Game: GTA V


(l) Game: Mortal Kombat 11


(m) Game: Fifa 19


(n) Game: Call of Duty Black Ops 2


(o) Game: Horizon Zero Down


(p) Internet: MetaHuman Creator

Figure 3.10: All dark colored skin characters presented in this work.

To test $H0_{10}$ ("defining that the UV effect is similar for white and dark colored skin characters, with similar levels of realism"), the perceptual data from white colored skin characters were used from the dataset present in Section 3.1. Remember that, concerning skin color and race/ethnicity, it was focus only on visual attributes; that is, without using narratives and contexts involving the characters in the media. Therefore, data from six (in Figure 3.1, being CG characters s, q, n, e, j, and r) of the 22 characters in the dataset were not used. Indeed, the removed characters (analyzed data) that, in our view, did not have white skin color and characters that we were not sure about their skin color, e.g., Hulk (n) and his green skin color. Considering the UV theory, these characters removed from the analysis do not negatively influence the methodology for creating the comfort chart (based on the UV chart), because the three levels of realism proposed in the work of [KFMT15] still remain: Unrealistic, Moderately Realistic, and Very Realistic (the characters with white colored skin can be seen in Figure 3.11(b)).

Regarding the dark colored skin characters, we chosen characters to be classified in the three levels of realism, as in the group of white colored ones, as discussed by Katsyri and his colleagues. Figure 3.10 shows the characters with dark colored skin used in this section. Following the methodology proposed in the previous sections, we looked for unrealistic characters (they are usually cartoon characters, for example, characters "a", "b", "f", "g" and "h"), moderately realistic characters (they are not a cartoon, but they do not have high levels of realism, for example, "c", "d", "e", "i", "j" and "l"), and very realistic characters ("k", m", "n", "o" and "p"). However, we needed to follow the methodology presented in the last sections to reduce differences in comparisons between perceptual data related to dark and white colored skin characters. As these characters are "public ready", again, we looked for the characters in Youtube videos, channels with many followers, and videos with many views that had permission for content transmission[3].

Questions and Stimuli

Before asking questions related to the characters, we asked participants the following demographic questions: *i)* The first question was "In terms of skin color and race/ethnicity, do you identify as a person:" with the answer options "Black", "Indigenous", " Yellow", "Brown", "White", "Other" (with a free text field), and "I do not want to answer"[4]. Note: As the previous section did not have the question of color and race/ethnicity, we did not use such data in the analyzes of $H0_{11}$; *ii)* "Do

[3]Still, we invoke the "fair use policy: Copyrighted images reproduced under “fair use policy” for protected images and videos.

[4]Those are the official categories of geographic institute used in the census of Brazil

you identify with the gender:" with the options "Female", "Male", "Nonbinary", "Other" (with a free text field), and "I do not want to answer"; *iii)* "Age:" with the options "18 to 20 years old", "21 to 29", "30 to 39", "40 to 49", "50 to 59" and "Over 60"; *iv)* "Inform the level of education" with the options "Incomplete high school", "Complete high school", "Complete higher education", and "Postgraduate".

In the Section 3.1.1, to compare with the work of Flach et al. [FdMM+12, DFH+12], we asked participants to answer five questions about the characters, for each image or video. However, in this current section, as we wanted to focus only on perceived realism and comfort (a possible UV chart related to skin color differences), we removed the questions that were not the focus. With this, we used only two questions, as the following: *i)* Question Q1 (”How realistic is the character") which the possible answers were ”Unrealistic", ”Moderately realistic", and ”Very realistic", is a 3-Likert Scale, and responsible for shaping the order of characters in terms of Realism. That is, each character had a response average between 1-Unrealistic and 3-Very Realistic, and we rank them based on these averages; and *ii)* Q2 (”Do you feel some discomfort/strangeness looking at this character?"), which was a ”Yes" or ”No" question. As in the last sections, we used the percentage of "No" answers for each character, that is, how much comfort was perceived.

We used the same methodology used in the Section 3.1.1 to divided the characters into two stimulus stages. In the first stage, we presented one image of each character, while in the second stage, we presented one video. All media were presented in a random order to avoid bias. The images, videos, and questions were applied through an online form created with Google Forms and transmitted to volunteer respondents through social networks. We presented a term of commitment approved by the ethics committee[5], where we presented all the possible risks of the respondent's participation. The images and videos of the characters have the same size in the form, approximately 520x520, measured on a 15-inch Full HD 1080 screen. The videos had an average duration of 4.937 seconds and it was focused on the characters' faces in both images and videos to avoid body bias.

[5] Research Ethics Committee of Pontifical Catholic University of Rio Grande do Sul, Brazil - Project Number: 46571721.6.0000.5336

3.3.2 Results

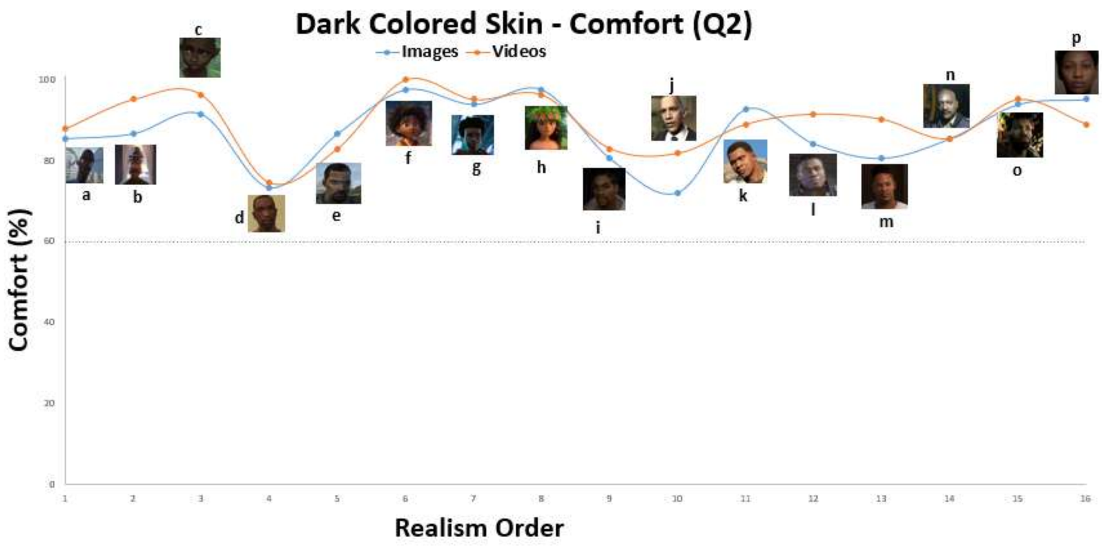
The questionnaire was answered by 82 volunteers, from which: 81.7% answered "White", 13.4% answered "Brown", and 4.9% "Black" for the question of identification with skin color and race/ethnicity. In this case, as we had many more white participants, we separated the data into white and brown/black (in Brazil, black and brown people can be grouped into one category [Gom19]) participants as they declared themselves. In addition, participants were 54.9% women and 41.5% men; 41.1% were between 18 and 20 years old; 48.8% had completed high school at most. In all statistical analyses, we used 5% of significance level (*Factorial ANOVA* test).

This section evaluates the $H0_{10}$ ("defining that the UV effect is similar for white and dark colored skin characters, with similar levels of realism") and the $H0_2$ ("defining that the effect of UV on different colored skin characters is similar for people with different racial identifications.") hypotheses. So, we compared white colored skin characters and dark colored skin characters (Figure 3.10) in a UV theory perspective, that is, perceptions of realism and comfort.

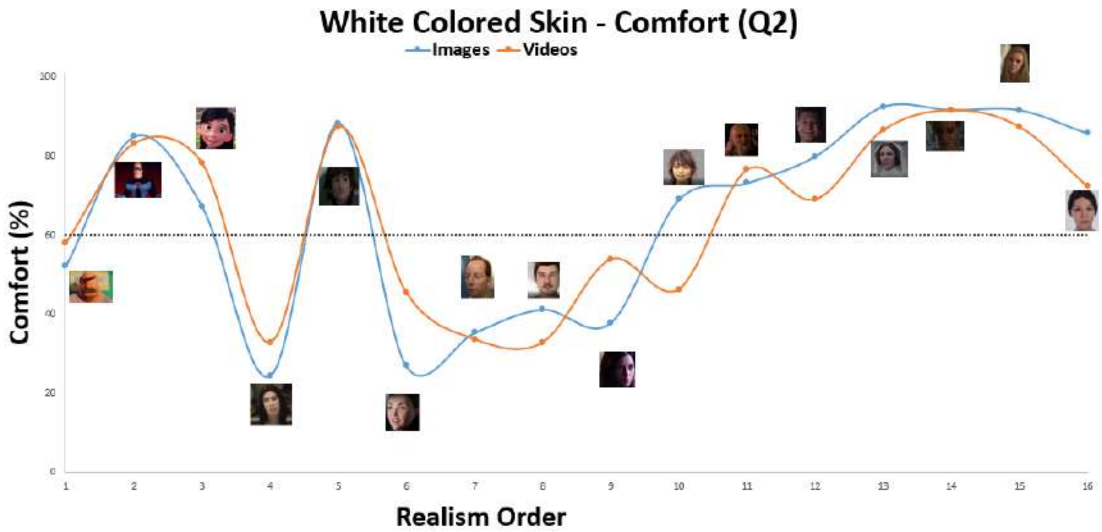
Analysis of Perceived Realism

In Figure 3.11(a), we can see the ordering of the dark colored skin characters in our work on Realism (horizontal axis). To order the characters, we used each character's averages of realism (answers to question Q1) Regarding $H0_{10}$, we found a significant result in the main effect of Skin Colored feature (F(1, 3214)=22.96 $p<$ .001), in which the average realism of white colored skin characters was 2.049, and the average of dark colored skin characters was 1.9. With this, **we can say that dark colored skin characters were considered less realistic than the white colored skin characters, so this refutes partially $H0_{10}$ hypothesis because participants do not evaluate the realism of black and white colored characters similarly.** The second part of the hypothesis, focused on the perception of comfort, will be presented in Section 3.3.2.

Regarding $H0_{11}$, we did not find a significant main effect of the Participant's Race. Remembering that, as in the previous sections we did not have racial identification data, we used only the data related to this current section in such analysis. That is, participants who identified themselves as white versus participants



(a) Dark Colored Skin Characters



(b) White Colored Skin Characters

Figure 3.11: In a), There are all dark colored skin characters evaluated in this work. The characters were ordered according to the perceived Realism (Q1 question). The blue and orange lines represent the comfort percentages (Q2 question) regarding images and videos, respectively. The dashed line represents 60% of comfort values. In b) all white colored skin characters from [AMDM21] are illustrated in this work.

who identified themselves as brown/black in relation to dark colored skin characters (respectively, averages of 1.89 and 1.91). In this case, with this, **we cannot refute the $H0_2$, in terms of perceived realism, for two reasons: First, we do not know the racial identification of the participants related to the white colored skin**

CG characters, and second, we did not find a significant result between the different racial groups present in our work.

Analysis of Perceived Comfort and UV Theory

Concerning the perceived comfort (Q2 question), the Figure 3.11 also shows the perception of comfort. Contrary to what happens with white colored characters (illustrated in Figure 3.11(b)), we can see that the characteristic shape of UV theory does not appear (Figure 3.11(a)), i.e., the visual aspect of the Valley. Considering the comfort values, in (a), all dark colored skin characters were above 60%, whereas in (b), six white colored skin characters had comfort values below 60% for videos, and five for images. Concerning $H0_{10}$, both for image and video (respectively, F(1, 3214)=212.64 and $p<$ .001, and F(1, 3214)=277.58 and $p<$ .001), we found significant results in the main effects of Skin Colored feature. The white skin colored characters had average comfort percentages of 65.07% in the image and 64.65% in the video, while the dark colored skin characters in our work had 87.27% and 89.55%. Therefore, **we can say that dark colored skin characters conveyed more comfort than white colored skin characters, which fully refutes hypothesis $H0_{10}$ (the evaluation of realism was presented in Section 3.3.2). Looking the Figure 3.11, we can also say that the UV effect was different for dark colored and white colored skin characters, refuting the $H0_{10}$ hypothesis.**

Regarding $H0_{11}$, as well as in the realism analysis, we did not find significant results. As in the analysis of realism, the values of perceived comfort about dark colored skin characters were higher for black/brown participants (respectively, 89.73% and 90.62%) than for white participants (86.66% and 89.17%). With this, **we can say that the perception of comfort about dark colored skin characters was similar for participants with different racial identifications, that is, confirming the $H0_{11}$ hypothesis. However, we need to point out that this could change if the sample of the black/brown participants were larger.**

3.4 Chapter Considerations

This section discusses the main findings of this chapter, and it is divided into: Section Discussion about Perception of CG characters in Relation to the Advances in Technology presents the discussions on the results of Section 3.1; Section Discussion about Perception of CG characters from a Gender Perspective presents the discussions on the results of Section 3.2; and finally .

3.4.1 Discussion about Perception of CG characters in Relation to the Advances in Technology

First the findings about the Section 3.1. In relation to **perceived realism**:

- For ALL characters, all groups of realism are significantly different; and
- Very realistic group from NEW characters is more realistic than very realistic group from OLD characters

We can conclude that people could perceive the different levels of realism, and also can assess the difference between the very realistic groups in 2012 and 2020. The recommendation here is that we can invest, if there are resources, to provide very realistic characters because people will have this qualitative assessment.

Regarding the **perceived comfort**, comparing ALL, NEW and OLD characters:

- For OLD characters, all subjects, in 2012 and 2020, perceived similar comfort and it was not impacted by the level of realism, i.e., $(O, O) = (O, N)$, so $H0_1$ is confirmed;
- People evaluated in 2020 felt similar comfort w.r.t ALL characters, as people in 2012 about OLD characters, i.e., $(O, O) = (A, N)$, so $H0_2$ is confirmed;
- In 2020, people felt more comfortable with current characters than people in 2012 with characters from 2012 (at that time), i.e., $(O, O) < (N, N)$, so $H0_3$ is refuted;
- In 2020, people felt more comfortable with current characters than with characters from 2012, i.e., $(O, N) < (N, N)$, so $H0_4$ is refuted. Besides, in 2020, people felt more comfortable with very realistic characters from 2020 than 2012; and
- No significant results were found concerning the comparison between comfort in images and videos, globally and with different levels of realism, in 2012 and 2020, so $H0_5$ is confirmed;

Based on tested evaluations, we can confirm that perceived comfort in 2020 is higher than in 2012 when evaluating OLD and NEW characters. So, designers can invest in more advanced resources to model the characters. However, it is also interesting to see that the comfort perceived w.r.t. 2012 characters did not change as a function of time. We can speculate that even if the audience can perceive the improvement in realism and comfort, OLD likewise characters can still cause comfort. Hence, designers have a wider range of possibilities, e.g., from advanced to not so advanced techniques to model the characters, which can be good news to those designers that do not have many resources. Still regarding the **perceived comfort**, comparing image and video stimuli:

- For ALL characters, in image analysis, people felt more comfortable with very realistic and unrealistic characters than with moderately realistic ones. In videos, people felt more comfortable with unrealistic and very realistic characters than with moderately realistic ones; and

- For NEW characters, in image analysis, people felt more comfortable with very realistic and unrealistic characters than with moderately realistic characters. However, only between very realistic and moderate characters a significant difference exists. In the videos, people were more comfortable with unrealistic and very realistic characters than with moderately realistic characters.

These results align with expectations of the UV theory in terms of perceived comfort, being higher in very realistic and unrealistic characters than in moderately realistic. However, we could not see the expected difference in terms of comfort obtained in videos and images, as suggested by Mori [MMK12].

The present work also assessed differences between years in perspective on gender and skin color diversities. Regarding gender, female characters have had greater growth in perceived realism and comfort over the years (old versus new technologies) than male characters. In this case, we can hypothesize that newer technologies may have been more effective in increasing female realism than male realism. And this may also have an effect on the issue of perceived comfort. In cases involving skin color between years, both in 2012 and 2020, the perceived comfort in relation to dark colored skin characters was greater than in relation to white colored skin characters. In terms of realism, for both the 2012 and 2020 characters, people in 2020 perceived more realism in white colored skin characters than in dark colored skin characters. Based on these results, the next section discuss more about gender and skin color.

3.4.2 Discussion about Perception of CG characters from a Gender Perspective

Following the results of Section 3.2.2, the levels of realism designers/animators put in CG characters with binary gender can affect perceived comfort by the audience. Although some analyzes on women and men, and female and male characters have not shown significant differences, some data obtained are interesting and are highlighted in this section.

In general, people felt similarly comfortable with male and female characters, both in images and videos. But going deeply by analysing women and men perceptions, we can observe some relevant aspects. Overall, unrealistic and very realistic characters tend to be more comfortable to people than moderately realistic characters. In this case, professionals, who are concerned with modeling comfortable characters for viewers, need to take into account that moderately realistic (for example, characters who have flaws in their anthropomorphic levels, such as the green skin of the hulk, Alita's big eyes, smiles with disproportionate sizes, among others) characters can cause discomfort. With regard to all the characters observed by women, realism did not influence the perceived comfort. In particular, with a significant difference, women felt more comfortable when watching very realistic female characters than male characters. It can indicate that in realistic games, women can feel more comfortable playing with female characters than with male ones. This result goes in the same direction as in-group advantages ([GBBM07, EA02, KST+15]) and Social Identity Theory ([TT04, Bro20]), since women and female CG characters can theoretically be in the same social group. We saw in the previous sections that technological evolution seems to have affected the perceived comfort of female characters more than that of male characters. And we can hypothesize that this may have been beneficial to the result in that it shows that women may feel more comfortable with very realistic female characters. Furthermore, it is in line with Draude's [Dra11] work, which mentioned that women might feel more comfortable with female CG characters. Unlike women, realism influenced the comfort perceived by men, being that they felt more comfortable with unrealistic and very realistic characters than with moderately realistic characters, both in image and video. Although no significant difference was found, men perceived very similar comfort of very realistic male and female characters, in images and videos. Therefore, only the results of women were in agreement with the results of the work of Tinwell et al [TNC13], which showed that male characters were stranger than female characters. However, contrary to this current work, the authors used only few different characters, and used morphing between levels of realism. Therefore, thinking about products with high levels of realism (films, games, series, etc.), the industry needs to guarantee representation if it wants to attract female audiences, that is, the ideal is to use female characters. The male audience, on the other hand, does not matter if the characters are male or female for perceived comfort. In this case, the important thing is that the characters cannot be moderately realistic.

3.4.3 Discussion about Perception of CG characters from a Skin Color Perspective

This section discusses some results presented in Section 3.3. Firstly, regarding the analyzes involving $H0_{10}$, this work compared the perception of all participants from Section 3.1 about white colored skin characters and the perception of all participants from Section 3.3.1 about dark colored skin characters.

The results showed us some interesting points: the first is that the UV effect [MMK12] was different for the dark colored skin characters, which conveyed more comfort and were considered less realistic than the white colored skin characters. These results are in line with those found in comparisons between 2012 and 2020, which also showed more perceived comfort and less realism in relation to dark colored skin characters, as discussed in Section 3.4.1. These results also can answer our question "Do characters with dark colored skin cause low or high sensations of discomfort in human perception?". As a consequence, in terms of UV theory, the Figure 3.11(a) did not present the expected Valley, i.e., not presenting characters with lower values of comfort. We can hypothesize at least two situations that could interfere with human perception causing the high values of comfort obtained in this work. Firstly, this may be related to rendering and illumination techniques. Kim [KRD+22] mentioned that most CG techniques are focused on VHs with white colored skin and that dark colored skin VHs often have exaggerated skin color lighting. Still, according to Kim, techniques such as "translucency and the corresponding physical mechanism of subsurface scattering has become synonymous with human skin in rendering. However, translucency is only the dominant visual feature of young, white Europeans and fair-skinned East Asians". In this case, the author still cited several scientific studies on this subject, which involve only examples of VHs with white skin color, which becomes a scientific bias in creating future

VHs. In the same line, the observed decrease in perceived realism of dark colored characters may be related to the theory of Dehumanization [KL22], since "algorithmically" the main methods were modeled and tested with white colored skin [KRD+22]. So, we can hypothesize that, in their perceptions, the participants did not "humanize" the dark colored skin characters, which could indicate an issue with rendering methods. Therefore, this question needs further study to draw more substantial conclusions. While one explanation may be the rendering and illumination techniques, another possibility can be the extension of human in-group and out-group effects. However, as this work did not have many participants who considered themselves black/brown, the work did not achieve significant results in this aspect. Lastly, it is important to mention why this type of research is relevant for the CG community. According to Katsyri et al. [KFMT15], the experience we have when we observe an artificial being is related to human identification. So, our ultimate goal is to contribute with methodologies that provide more equity and representation in the modeling of VHs, providing better experiences to users.

4. GENDER BIAS IN TERMS OF GENDER ATTRIBUTION

This chapter aims to carry out analyzes regarding the perception of VHs in terms of binary and nonbinary genders. In this sense, this chapter is also divided into four parts: *i)* Section 4.1 presents a perceptual study which goal is to to assess people's perception of genderless VHs. In this case, the section presents a perceptual study related to gender bias when people watch a VB; *ii)* Section 4.2 presents a perceptual study which goal is to assess gender bias when people interact with a VB; *iii)* Section 4.3, as in the Section 4.1, presents a perceptual study which goal is to assess people's perception of genderless VHs. However, in this case, the section presents a perceptual study related to gender bias when people watch an adult VH; *iv)* And finally, Section 4.4 presents the discussions of the results presented in the other three sections.

4.1 Gender Bias in Relation to Observing VBs

As mentioned in the previous section, so far, only characters with female and male characteristics were mentioned, but what about when the VHs do not have a defined gender (i.e., androgynous, nonbinary, genderless VHs). Draude [Dra11] made an analogy, stating that Alan Turing's Imitation Game [Tur09] has the potential to deconstruct gender through Computer Science, that is, without gender stereotypes. This game was proposed by Turing in his work "Computing Machinery and Intelligence" [Tur09] to try to answer the question "Can machines think?". In this case, the Turing test challenges the ability of a computer to engage in possible human-like conversations. In the game, Turing proposes that three people play, being a man (A), a woman (B) and a third person (C). This third person was the investigator, who can be of either sex, stays in a separate room from the other two, and has the objective of trying to identify the gender of the first two. In addition, the interrogator knows the other two participants by the labels X and Y, and at the end of the game it is necessary to say who is A and who is B. The woman had a specific role in the game, which was to be the interrogator's assistant. The role of the man in the game was to try to make the interrogator have wrong answers about the identifications. With that, Turing suggested that the woman be sincere in carrying out her role. However, as the man needed to perform his function, then he could claim to be a woman. Therefore, both the woman and the man needed to argue with the interrogator that they are women. Turing then suggested replacing the original question, ”Can machines think?”, by the question, ”What will happen when a machine takes the part of A in this game?” With that, Turing proposed that a machine should be a man, indicating that it was easier for the male machine to imitate a woman than to propose a female machine. According to Draude [Dra11], this suggestion caused the game to become a gender-biased game. Thus, from the perspective of the third actor (the interrogator), even with a gender bias, the author stated that there was potential for gender deconstruction, since the machine is essentially genderless. In this line of thinking, gender is a subject of examination explored in the studies of Scott [Sco07] and Wallach [Wal10], and is intertwined with social and cultural constructs regarding roles assigned to women and men. So, in other words, a machine is born essentially without an assigned gender, and the person who assigns it a gender is its creator. With that, what about virtual humans? Is there this potential for gender deconstruction in a genderless virtual human?

There are several games in which the player can choose the design of a VH, such as gender, race, face and body characteristics, clothes, among others [1] [2]. Through data obtained in these games it is possible to measure patterns of player choices. In one example, the role-playing game developer Larian Studios used analysis tools to learn the most popular choice of character created in their Baldur's Gate game [3]. In this game, there are several characteristics that can be modeled from an adult character with no defined gender, however, the results showed that the most created type was a male character. The preference to attribute a non-gendered character into a male character can be a problem of gender bias that is rooted in human culture. However, the characters are adults in these games, but what if the VH was not an adult humanoid? As mentioned in previous paragraphs, there are more male main characters in games, meaning this is still fueled by the gaming industry. In the movies industry, specifically when we talk about cartoons, female characters are usually more stereotyped [GAHH19] both in terms of visuals [4] and movement (facial and body emotions, among others). This makes it necessary to exaggerate the design of female characters for the public to recognize them. But are these exaggerations really necessary for the attribution of a female VH?

[1] https://twinfinite.net/2018/08/games-nailed-character-creation-systems/
[2] https://starloopstudios.com/character-design-in-video-games-types-classes-and-characteristics/
[3] https://www.avclub.com/baldurs-gate-3-developers-reveal-baldurs-gate-3-players-1845331378
[4] https://thedissolve.com/news/655-disney-animator-says-women-are-hard-to-draw-becaus/

We can also relate these findings to studies that showed that people tend to have an advantage in attributes categorization of perception among groups to which they belong, as in-group advantages, as mentioned in studies of [EA02, KST+15]. As discussed in the work of [AMRM19], this type of advantage also extends to gender, which states that women tend to identify better emotions expressed by other women. In examples involving gender attribution, studies by [ZHRM15, HAJK04] have shown that it may be easier for participants to identify the gender of opposite-sex actors presented in an incongruent model (e.g., a VH without gender stereotypes). In the case of VHs, the work by [GBBM07] showed that both women and men presented in-group advantages when evaluating persuasive speeches by female and male VHs. For example, when assessing the attitude of VHs with the same animations, women evaluated that a female VH had more attitude than a male VH, while men evaluated the opposite.

To eliminate substantial evidence of biases [DSKP22, KRD+22], when incorporating gender into the design of VHs, it is crucial for professionals such as designers, programmers, and researchers to eradicate their own gender biases to deconstruct gender stereotypes [GFGW+23]. Moreover, they should consider the diversity of the audience's gender identities and experiences, visually and behaviorally [MJH+07, MJH+09], to create inclusive and representative VHs. In the study conducted by Armando et al. [AOR22], the authors posed the following question (a question similar to what was said by Turing [Tur09] and by Draude [Dra11]): "Can Androgynous Virtual Agents be a Potential Solution to Combat Gender Stereotypes?" Additionally, the authors extended this question to include genderless VHs. Therefore, this current section aims to try to answer the following questions:

- How do people perceive genderless VHs?
- Does this change with respect to participant gender?
- Do we have a bias to perceptive gender?

To try to answer these questions, this section replicates and expands the work of Condry and Condry [CC76], in which the authors measured the perception of men and women about a video containing a real baby reacting to different stimuli. The authors separated the participants into two groups, one receiving the video of the baby with a male name and the other receiving the video of the baby with a female name, but the baby was always the same, that is, only changing the name. Therefore, people in different groups received the same video of the baby, but only changing the baby's name (female and male). In addition, the authors also asked the participants whether or not they were familiar with young children. The results showed that women who were familiar with babies perceived more emotions in the baby with a female name, while men perceived more emotions in the baby with a male name. In other words, there were in-group advantages for both the women's and men's groups. The research showed a gender bias in a non-adult human (a baby) being who does not have female or male stereotypes. So this means that people's perception is influenced by the gender characteristic, even if that characteristic is just the identification of the gender or the name. Remembering, in the experiment, the baby was always the same for all groups of people, only the name and gender were changed for each of the groups. With that, the idea is to replicate Condry and Condry's experiment using a Virtual Baby (VB), and expand the experiment by having a group of people who will receive an unnamed baby (in addition to the other two groups that will receive the baby with a female and male name). Therefore, to help answer these questions, the following hypotheses were created:

- $H0_{12}$ defining that women and men perceive emotions similarly in female, male and unnamed VBs;

- $H0_{13}$ defining that the unnamed VB is recognized as genderless;

This section is divided into Section 4.1.1, which presents the methodology, and Section 4.1.2, which presents the results. Furthermore, two important observations are made to prevent any confusion throughout the next sections:

i) When referring to the ”gender attribution feature", we are specifically addressing the gender attribution made by the participants about the VHs they encountered in the study. This feature focuses on the perceptions and interpretations of the participants regarding the gender of the VHs.

ii) When mentioning "Textually Assigned Gender," we are specifically referring to the genders that have been assigned to the VHs based on their names or textual descriptions.

By providing these clarifications, we aim to ensure a clear and consistent understanding of the terminology and concepts used in the work.

4.1.1 Methodology

This methodology of this part of the work is divided into: Section Creation of Stimuli presents the creation of stimuli used in this part of the work; and the Section Questionnaire presents the proposed questionnaire.

Creation of Stimuli

In order to answer the hypothesis $H0_{12}$ and $H0_{13}$, the experiment of Condry and Condry [CC76] using a VB was recreated. First, explaining Condry and Condry's experiment, the authors presented to participants a video of a real 9-month-old baby sitting in a baby chair facing a mirror (with a camera mounted behind the mirror) reacting to four stimuli (the teddy bear, a jack-in-the-box, a doll, and a buzzer). The baby had neutral clothes to avoid gender stereotypes. Then, the authors separated the participants into two groups, with a group receiving the baby with a female name, and the other group receiving the baby with a male name. Thus, the authors asked the participants to use predefined scales to assess the baby's emotional in relation to the four stimuli presented. The main goals of the authors were to know whether the baby's gender influenced the perception of emotional levels and the gender attribution.

In our work, we used a 3D model of a VB purchased on website [5] to replicate the work of Condry and Condry [CC76] in the virtual environment. The model had animations of crawling, walking and playing with a ball (which was an animation where the baby interacts with a ball). In addition to the model, it was necessary to have objects that the VB could interact with. In the work by Condry and Condry, the authors reported that the baby reacted positively (smiled, laughed, reached out) to the teddy bear and the doll, and reacted negatively (turned away, stared, cried) to the other two objects. Three stimulus objects were used for our VB to interact. Firstly, the animation of the VB playing with a ball, which in our hypothesis was perceived as a positive reaction from the baby. Secondly, the possible negative reaction was hypothetically created using the same object as in the original work, i.e., a virtual model of a jack-in-the-box [6], which contains an animation of "Jack" jumping out of the box and back into the box. The crawl animation was used for the VB to reach the jack-in-the-box, and a simple facial animation (mouths opened as in a surprise facial

[5] https://www.cgtrader.com/3d-models/character/child/game-ready-baby
[6] https://www.blendswap.com/blend/27680

expression) was created (using the facial blendshapes of the baby). Regarding a object that caused a hypothetically neutral reaction, a 3D model of a colored unicorn[7] was used, and the VB's reaction was to crawl in the opposite direction to the unicorn (without interest). Furthermore, this object was also chosen because it was also necessary to have a stereotyped object associated with the non-masculine, for example, like the doll in the original paper. Bearing in mind that, even though the non-male stereotyped object (colored unicorn) and the male stereotyped object (ball) were used to cause, respectively, neutral and positive reactions to the VH, the goal was centered on analyzing whether the VH's defined gender would influence the people's perception, not whether the stereotype of the object would cause the influence. The objects used in this work can be seen in Figure 4.1.

(a) Ball

(b) Unicorn

(c) Jack-in-the-box

Figure 4.1: Environment and three objects that the baby interacts with: (a) the baby plays with the ball; (b) the baby crawls in a direction opposite to the unicorn; (c) the baby has a negative emotion seeing Jack jump out of the box (Jack-in-the-box).

Three videos as stimuli were created, each video for each object (ball, jack-in-the-box, unicorn), to present to the research participants. For all videos, a virtual scenario was created in which it was possible to apply the reactions of the VB to the objects (explained in the previous paragraph). The set contained a room which has a table, a sofa, an armchair, a curtain, an open window showing the sky with clouds, and a wooden floor (all of these objects in the set were obtained from the

[7] https://free3d.com/3d-model/unicorn-doll-772526.html

internet). The scenario can be seen in Figure 4.1. The camera always remains in the same position, that is, pointed at the window. At the beginning of the videos, the VB always starts facing the camera and objects, and with his back to the window. Also, the baby never leaves the camera view, and the only object that moves is the ball, as the baby makes it move in the animation. The videos were between 0.06 and 0.19 seconds, and were sent to YouTube to be added to the questionnaire (explained in the next section).

Questionnaire

The questionnaire was also based on the work of Condry and Condry [CC76]. First, the participants were presented with the consent form approved by the Ethics Committee of Pontifical Catholic University of Rio Grande do Sul (CEP), referring to the research project entitled "Estudos e Avaliações da Percepção Humana em Personagens e Multidões Virtuais", number 46571721.6.0000.5336. Therefore, participants were informed of potential risks. In addition, the questionnaire was created on the Qualtrics platform [8], distributed on social networks, and without financial compensation, that is, participants were volunteers. Participants were also asked about their demographics: age, educational level, gender, and familiarity with CG (games, movies, simulations, among other examples).

After this, three types of questionnaire were created, one containing the baby with a female name, one containing the baby with a male name, and one containing the baby without a name. We randomly selected one of the questionnaires for each participant. Both the female and male names were presented in an introductory text block[9], which presented the baby with the name and gender (no name and gender in the unnamed group), and stating then the gender and age (9-month-old) of the VB. While in the work by Condry and Condry, the authors created only two groups of questionnaires, one containing the baby with a female name and the other with a male name. Our work had one more group, the group with the unnamed baby, in order to test the $H0_{13}$ hypothesis, that is, to assess whether people perceive that an unnamed baby does not have a defined gender.

After this step, the participants received an informed text about the videos and questions that would follow. As in the work by Condry and Condry, this present work also used emotional rating scales. For each video, the participants were in-

[8] https://www.qualtrics.com/pt-br/

[9] We chose two popular names: Helena and Miguel.

structed to rate, on 11-Likert Scales (as in the original study), the pleasure, anger, and fear felt by the baby. If the participant noticed that an emotion did not appear in the video, then he was instructed to score the lowest value on the scale (with 0 = "Absence of Emotion"), and the highest value if the emotion appeared as much as possible (with 10 = "High Intensity"). This block of questions is intended to help test hypothesis $H0_{12}$, and it provides a 2 (Women and Men responses) x 3 (Ball, Jack-in-the-box, and Unicorn) x 3 (Pleasure, Anger, Fear) x 3 (Female name, Male name, and Unnamed VB) design structure.

Finally, three questions about name, age and gender were asked to the participants (also based on the work of Condry and Condry). However, only the question "What was the baby's gender?" is evaluated in this work, with "Female", "Male", "Other", and "I do not know" as possible answers, as it aimed to test hypothesis $H0_{13}$. This step was 2 (Women and Men responses) x 3 videos x 3 (questions about name, age, and gender) x 3 (Female name, Male name, and Unnamed).

4.1.2 Results

The questionnaire was answered by 148 volunteers, being 79 women (25 received the female name baby, the male name had 24, and the unnamed had 30), 66 men (23 in female name, 23 in male name, and 20 in unnamed group) and three people who chose another option. [10] Regarding age, from 145, 105 people were younger than 36 years old. About education, 105 people completed undergraduate studies. 130 responded that they were familiar with CG. In the statistical analyses[11][12], we used the nonparametric tests: *Kruskal-Wallis*, *Mann-Whitney*, and *Chi-Square*, with 5% of significance.

Explaining the results related to hypothesis $H0_{12}$, Table 4.1 presents the averages of the emotional rating scales (11-Likert Scales) for all videos (Ball, Unicorn, and Jack-in-the-box). The columns represent the separate emotional scales (Pleasure, Anger, and Fear), and all emotions together[13]. The first four data columns refer to the answers of the participants who received the VB with the female name, followed by the VB with the male name, and the unnamed baby. Therefore, the first lines refer to women's answers, and the last lines refer to men's answers.

[10] We removed these three participants because this group was very small.
[11] https://www.statsmodels.org/dev/index.html
[12] https://docs.scipy.org/doc/scipy/index.html
[13] Note: The global analysis of emotions was also based in the cited paper.

Table 4.1: Table of results referring to the averages of emotional rating scales. The first four data columns correspond to the responses of the participants who received the questionnaire containing the baby with a female name, followed by the columns of the groups that received male, and unnamed names. In addition, "Avg Videos" states the average emotional rate value of the three videos together (Ball+Unicorn+Jack-in-the-box), and "Avg Emotions" presents the average value of all emotions together (Pleasure+Anger+Fear). The first three lines refer to women's data, and the last three refer to men's data.

Women												
	Female Name				Male Name				Unnamed			
Stimuli	Pleasure	Anger	Fear	Avg Emotions	Pleasure	Anger	Fear	Avg Emotions	Pleasure	Anger	Fear	Avg Emotions
Avg Videos	5.18	0.98	1.65	2.60	4.40	0.66	1.38	2.15	3.51	0.84	1.44	1.93
Men												
	Female Name				Male Name				Unnamed			
Stimuli	Pleasure	Anger	Fear	Avg Emotions	Pleasure	Anger	Fear	Avg Emotions	Pleasure	Anger	Fear	Avg Emotions
Avg Videos	4.34	0.54	0.98	1.96	4.94	1.04	1.85	2.61	5.11	0.88	2.03	2.67

Concerning women population, in general, the emotional perception was higher for the baby with a female name than for the other groups. We found significantly different values in pleasure (.006) and global analysis (Avg Emotions) (.03). Furthermore, using a *Dunn* test as a *post hoc* with a *Bonferroni* correction, the women's perception of emotion (.03) and pleasure (.004) had the most significant difference in the comparison between the group that received the female name and the group that received the unnamed baby. Therefore, **we can say that the women perceived that the baby with a female name was more emotional, in general, and felt more pleasure than the baby with a male name and the unnamed baby.** **About the results of men**, the opposite happened, i.e., in most cases, the average values were lower for participants who received the VB with a female name. The *p*-values were significant in fear and in relation to all emotions "Avg Emotions" (.01). Regarding the *post hoc*, in the perception of all emotions (.01) and in the perception of fear (.03), the most different groups were the ones who received the groups of female and male names. Therefore, **we can say that the men perceived both the baby with a male name and unnamed as more emotional and feeling more fear than the female named baby. In general, it was more evident when we compared men's perceptions of the baby with male and female names.** Bearing in mind that the baby is the same for both women and men groups, only the name was changed. Comparing women and men, regarding emotions in general, we found significant values in the three groups (.006, .01, .01, respectively for female, male names, and unnamed). With this, we can say that **in relation to $H0_{12}$, women perceived more emotion in the babies with a female name than men, and men perceived more emotion in the babies with a male name and unnamed than women**.

In relation to the gender attribution, when the baby was not assigned a gender, 18 women (60% from the total) and 11 men (55%) responded that the baby was male. We compared the correct answers, i.e., "female" for who received the female name (88% for women and 86.95% for men), "male" for the male name (91.66% for women and 100% for men), and "Other"/"I do not know" for without name (30% for both, women and men). We found no significant results, that is, **women and men were similar in the gender attribution, they attributed according to gender when the babies were assigned a gender (female and male names), and did not attribute according to gender in a similar way when the baby was not assigned a gender (unnamed baby). In relation to $H0_{13}$, even if the baby does not have visual/behavioral gender stereotypes, people will still point out that it is male.**

4.2 Gender Bias Regarding Interaction with VBs

In the last sections, we presented videos of a gender-neutral VB reacting to some toys to three different participant groups, with the aim of reproducing a study on gender bias in the psychology literature (that is, the work of Condry and Condry [CC76]). We assigned a distinct gender to the VB by providing a name for two groups (one group received a female name, another group had a male name), and a third group watched the VB without any assigned name. However, we only relied on participants' observations while watching a movie and did not test their perception of how they feel if they could interact with the VB. Therefore, there is evidence in the literature that adults carry their gender biases when interacting with infants [SKZ75, SL80, WSD76, SL78]. For instance, in the study by Seavey et al. [SKZ75], the authors conducted a study similar to the Condry and Condry [CC76] study but included interactions between infants and adults. One of the results indicated that participants tended to choose more "feminine perceived toys" when the baby had a female name. So, this current section aims to try to answer the following question:

- Would the gender-based preference also emerge when interacting with a genderless VB?

With this, the main goals of this section are: *i)* to assess whether participants' perception changes as a function of interaction in 3D environments, i.e., comparing watching videos of VBs versus interacting with them. These analyzes are important to measure the interaction between humans and genderless machines (that is, VHs). As in Section 4.1, this current section is in line with what was stated by Draude that genderless machines have the potential to deconstruct gender, or combat gender stereotypes as stated by Armando et al. [AOR22]; and *ii)* to assess whether gender-biased behaviors of psychology literature in the interaction between infants and adults also occur in interactive virtual environments.

Then, we raised the following hypotheses:

- *$H0_{14}$: People perceive emotions similarly in VBs with different genders:*

  $H0_{14-1}$ Only in interactive environments.

  $H0_{14-2}$ Comparing interactive with video setups.

- *$H0_{15}$: People attribute gender similarly to VBs with different genders:*

  $H0_{15-1}$ Only in interactive environments.

  $H0_{15-2}$ Comparing interactive with video setups.

- *$H0_{16}$: People interact (toy choices, emotional reactions, and proximity) similarly with VBs.*

To test these hypotheses, we based on the work did in Section 4.1, but considering a 3D interactive environment. Drawing on previous literature exploring adult-infant interaction, we asked participants to play with a VB using toys within an interactive environment. Therefore, we chose some types of interactions used in the literature, such as choosing a toy for the baby, the participant's emotional reaction towards the baby, and the participant's movement around the baby in the virtual environment. Regarding the participant's movement, we also tried to measure this attribute due to the fact that in several studies involving virtual environments, both the gender of the participant and the gender of the VH can influence proximity involving Virtual Reality (VR) [ZNO+22], perception of facial characteristics [GFGW+23], perception of emotions [ZHRM13], perception of movement [ZNO+20b], and others. As far as we know, this is the first time this kind of active interaction has been proposed to evaluate human gender bias and compare it with real life.

4.2.1 Methodology

Our experiment consisted of interactions between participants and a VB in a 3D virtual environment accessed through web browsers. The participants accessed the application link through a form in which they answered questions after the interaction. The following sections detail our methodology.

Stimuli Creation: Baby and Toy Animations

We used the same 3D VB model (purchased[14]) used in our last sections, and also its animations, the toys (Ball, Unicorn[15], Jack-in-the-box[16]), etc. The three virtual toys were hypothetically linked as female (pink Unicorn), male (Ball), and neutral (Jack-in-the-box) toys. For example, the ball was used as a male toy in most of the studies cited above, and the Jack-in-the-box is a toy that was used to evoke surprise in the VB, regardless of the baby's gender. Concerning the unicorn, to reinforce a possible female bias, we changed the original body color of the unicorn from white to pink. Furthermore, if we search for toys for boys and colors for girls on the internet, in most cases we can find that ball is assigned as a boy's toy and pink is assigned as a girl's color. However, we know that this may be a limitation of our work, where in the future we could carry out studies with different toys for different genders. The VB, the three toys, and the virtual environment can be seen in Figure 4.2. Respectively, the ball in (b), the unicorn in (c), and the Jack-in-the-box in (d). Therefore, as we have already said in the previous sections, in relation to the VB, there was no visual indication (accessories, clothes, etc.) that would relate it to any gender.

In the interactive environment, the VB always started from an idle animation (as shown in Figure 4.2(a), remained seated looking at both sides of the virtual room), and returned to the idle animation. In addition, we had the three animations where the VB plays with the toys. Figures 4.2(b), (c), and (d) respectively show frames of the animations referring to the ball (the VB could play with the ball - animation time = 30 seconds), a unicorn (the VB could crawl in an opposite direction - animation time = 12 seconds), and Jack-in-the-box (the VB could crawl towards Jack and open the mouth - animation time = 14 seconds).

[14] https://www.cgtrader.com/3d-models/character/child/game-ready-baby
[15] https://free3d.com/3d-model/unicorn-doll-772526.html
[16] https://www.blendswap.com/blend/27680

Participant's Actions

(a) VB in Idle Animation

(b) VB interacting with the Ball

(c) VB interacting with the Unicorn

(d) VB interacting with the Jack-in-the-box

Figure 4.2: VB actions (Idle (a), Interactions with Toys - Ball (b), Unicorn (c) and Jack-in-the-box (d)) and the 3D interactive virtual environment.

(a) Positive Emoji

(b) Other position of the participant

Figure 4.3: The participant's possible actions that appear on the screen (buttons for choosing toys and emojis; clock with interaction time; and indications through figures of keyboard and mouse buttons on how to move around the environment). In (a), an example of choosing a positive emoji (representing a positive emotion felt by the participant during the specific moment of the interaction). In (b), the example of the participant in another position of the virtual environment, looking at the VB through another angle.

As in the studies presented in literature, we decided to analyze the interactions (actions) of the participants with the VB for three minutes. The choice of interaction time was based on the studies by Seavey et al. [SKZ75] and Sidorowicz and Lunney [SL80], and on the fact that a longer time could be exhausting for the participants, who would still need to answer a questionnaire after the interaction. Participants always started the experiment facing the baby, as shown in Figure 4.2(a). In the upper right corner of the screen, participants could see a timer that indicated the interaction time (three minutes of interaction). So, there were three possible actions from the participants:

- Baby Toys (Toy Choice): With a click of the left mouse button on top of the toy figures (on the right of the interface), participants could choose one of the three toys, as many times as they wanted, during the interaction, as shown in Figure 4.2 in (b), (c), (d). However, one toy could be chosen at a time. After the participant chose a toy, the animation related to that toy started, and the other two toys could not be chosen until the animation ended (as shown in Figure 4.3(a)). So, we can assess whether the gender of the participant influences the choice of the toy (for example, women choosing the unicorn more than the ball);
- Participant Emotion (Reaction with Emojis): With a click of the left mouse button on top of the emoji figures, participants could also choose positive, negative, and neutral emojis to represent what they were feeling during the experiment. In the studies in the literature shown in Chapter 2, the authors recorded or observed participants' facial emotions during interactions with real babies. In our case, we used emojis as resources because we were limited to an experiment in which we had no way to record the participants' facial expressions. When the participant chose an emoji, it appeared at the bottom of the screen and its position was animated vertically on the screen (as shown in Figure 4.3(a)) until it disappeared on the top of the screen. The participant could not choose another emoji until it disappeared from the screen. Thus, we can evaluate the participants' emotional reactions during the experiment and if the gender of the participant and VB interfere with these reactions.
- Participant Movement in the 3D virtual world: As we can see in the upper left corner of the screens in Figures 4.2 and 4.3, we added visual indications of how participants could move around the virtual environment. The participants could rotate the camera in the X and Y axes by pressing the left mouse button. To avoid the participants' dizziness, we limited the rotation in the X axis to 30 degrees up and down. Participants could use the keyboard's up

and down arrow keys to move forward and backward along the Z axis. Figures 4.3(a) and (b) show the participant's camera in different positions and angles. We chose these possible actions so participants could freely move and look at the 3D virtual environment, since in the studies we used as a basis ([SKZ75, SL80, Gre89]), adults could also move freely in the environment to interact with babies. With the participant's movement during the interaction, we can also assess whether user proximity is influenced by both the participant's gender and the gender of the VB. Indeed, as shown in the work of Greeno [Gre89], the proximity between the participant and a baby can vary depending on the gender of both. So, we evaluated the interaction time (i.e., relative to the three minutes) that participants spent in four Hall's [Hal66] proxemics/space distances of the VB. The distances are: *i)* Public - more than 3.65 meters; *ii)* Social - 1.21 to 3.65 meters; *iii)* Personal - 0.76 to 1.21 meters; and *iv)* 0 to 0.76 cm. Furthermore, the Hall Distance is a measurement widely used in experiments with VR, so it makes sense to use it in our 3D interactions.

Experiment Setup

We used three technologies to create the experiment: Unity engine to develop the 3D virtual environment, Unity WebGL to run the environment on Internet browsers, and Firebase [17] Realtime Database to save the actions of each participant. The use of the project through Unity WebGL was due to trying to get more responses from participants by keeping the application available on the web. We used all the default object textures compatible with WebGL and closed the curtain in the window to avoid showing the blue sky outside the virtual room. We generated a WebGL file containing the virtual environment and hosted it on the Web through Itch.io [18], which is a platform that allows running Web files generated by Unity. This platform also allows other people to access the generated executable file. At Itch.io, we scaled the experiment viewport so that participants were limited to interacting on an 840x560 screen. As soon as the participant finished the three minutes of interaction with the VB in the virtual environment, the data of the participant's actions were automatically sent to the Firebase, which generated a Json file containing the data. Firebase is a set of back-end cloud computing services and application development platforms provided by Google. Unfortunately, as a limitation of our work, only people

[17] https://firebase.google.com/
[18] https://itch.io/

using Windows and MacOS could participate in the experiment because only these two platforms are compatible with Unity's WebGL.

Questionnaire

The online questionnaire was created on the Qualtrics platform and distributed to volunteers through social media (Instagram, LinkedIn, Discord, WhatsApp) and institutional emails. First, we presented the consent form approved by by the Ethics Committee of Pontifical Catholic University of Rio Grande do Sul (CEP), referring to the research project entitled "Estudos e Avaliações da Percepção Humana em Personagens e Multidões Virtuais", number 46571721.6.0000.5336. Participants' Demographic information, such as age, gender, educational level, and a question regarding CG familiarity was also collected.

Furthermore, we introduced a concise yet informative text outlining our research's purpose. This text served as a means for participants to understand the study's objective, which involved evaluating their interactions with a VB (9-month-old) over a three-minute period. We also randomly (through the Qualtrics system) divided the participants into three groups, and each of the groups received, in addition to the initial text, the name of the VB. That is, one group received information about a VB with female name (Helena), one group received information about a VB with male name (Miguel), and the last group did not receive a name. Since a participant could withdraw from responding/participating in the study at any time, we did not control exactly the number of participants in each group of users. Before sending the participants the link to the 3D virtual environment, they also read information about the possible actions they could take during the experiment through a simple tutorial. As in the initial text, we informed the participants again that they would help a VB play with some toys and were free to do this through their possible actions (through images, showing examples of keyboard and mouse button options). So, firstly, we used images of the environment to show examples, followed by texts mentioning that they could choose toys to play with the VB, react emotionally, and move freely in the 3D virtual environment. Regarding the possible actions, we showed figures highlighting how to choose toys and reactions through emojis, such as the interface elements on the right Figure 4.2(a - ”Baby Toys" and ”Your Emotion"). In addition, as shown in the left corner of Figure 4.2(a - movement with mouse and keyboard), we also informed how the participants move and rotate the camera in the virtual environment. Next, to confirm the information was important, we asked the participants if they had read the instructions (Yes or No question). If the answer was affirmative, the participant received the link; otherwise, they remained on the question page until they answered ”Yes”. Once participants accessed the environment link, they could start the interaction by pressing the play button.

After the three-minute interaction, a screen with text appeared to the participants, informing them that they needed to access the form to answer some questions regarding VB. In the questionnaire, they were informed that their task was to classify the perceived emotion intensity of the VB concerning the three toys in three emotional dimensions. Therefore, as in the experiment involving only videos of the VB (presented in Section 4.1), they evaluated through an 11-Likert Scale (0 = No emotion; 10 = High Intensity) how much pleasure, anger, and fear the baby felt about each of the three toys. We also state that if the VB expressed only one emotion, e.g., only anger, during action with a toy, then the intensity of the other emotions could be evaluated with zero. In addition, we also advised that participants were not asked about emotions about toys they did not choose during the interaction. After participants rated the VB's emotional intensities, we asked, "What was the VB's gender?". Possible responses were "Female", "Male", "Other" (with open text field - No participant responded to this text field), and "I don't know". This question aimed to assess gender attribution ($H0_{15}$).

4.2.2 Results

Our questionnaire was answered by 63 volunteers. Twenty-eight answered female/women in gender question (8 received the female name baby, the male name 5, and the unnamed had 15), and 35 answered male/men (13 received a female name, 14 a male name, and 8 received an unnamed baby). Thirty-four people were between 25 to 35 years old. 54 people completed undergraduate studies. 58 responded that they were familiar with CG. The number of participants was irregular between groups due to factors such as withdrawal during the questionnaire. Therefore, we cut these participants from the analyses.

In this section, we compared the results between interactive VB and VB presented in videos, both about the evaluations of the Emotional Scales (pleasure, anger, fear, and emotions in general) and Gender Attribution. The statistical analyses performed were a factorial *ANOVA* 2 (Participant Gender - Women and Men) x 3 (Baby Name - Female, Male, and Unnamed) to evaluate our results and a nonparametric approach through *Mann-Whitney U* test to compare interactive VB and VB presented in videos, considering significance levels of 5%. We used G*Power software [19] to assess the effect size of *ANOVA* using our number of participants (63). The effect size (n^2) was 0.4 (F-value(2, 6)=3.15), indicating a large effect; consequently, the significant values to our results should be greater than this F-value. In addition, we tested possible variations using the *Bartlett* test (homoscedasticity) and post-hoc tests using the *Tukey HSD* or the *Tamhane* tests.

Emotional Scales Analysis

Regarding $H0_{14}$ (*People perceive emotions similarly in VBs with different genders*), in relation to interactive environments ($H0_{14-1}$), we did not find any significant results for either Participant's Gender, Baby's Gender, or interaction effects between these variables. As shown in the chart on the right of Figure 4.4, women (blue line) perceived more emotions on average made by the VB with male name than by female named and Unnamed VBs. They also perceived more emotions made by male VB than men (red line). While men also thought male VB had more emotions than female VB. **However, not rejecting $H0_{14-1}$, we can say that people perceived emotions similarly in VBs with different genders in interactive environments.**

Concerning $H0_{14-2}$ (*Comparing interactive with video setups.*), first, we look at values globally before separating into women and men. We found a significant result when comparing the general averages related to interactive VB and VB presented in videos (*Stats*=409859.0 and $p<$ .001), showing that participants in the interactive experiment perceived more emotions in the VB (AVG=2.78) than participants who only saw VB in videos (AVG=2.29). We also found a significant result in the comparison between participants who received the VB with male name (*Stats*=41394.0 and p=.003). This showed that the participants who received the interactive VB with male name (AVG=3.16) perceived more emotions than the participants who only saw the male VB's videos (AVG=2.37).

Regarding participant gender, we found three significant results only in the data related to women. In the general comparison between women who interacted and who only watched VBs in general (*Stats*=100968.5 and p=.001, respectively, AVGs of 2.89 and 2.21), and also VBs with male name (*Stats*=6173.0 and p=.001,

[19] https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

respectively, AVGs of 3.71 and 2.15) and unnamed VBs (*Stats*=20677.0 and *p*=.01, respectively, AVGs of 2.66 and 1.93), women perceived more emotions in interactive VB than in video VB. Looking at the two charts in Figure 4.4 (on the left there are the results with video and on the right with interaction), in all cases, both women and men rated more emotions in interactive VBs (respectively for female, male and unnamed). In general, **rejecting $H0_{14-2}$, we can say that people who interacted with VB perceived more emotions than people who only watched VB in videos.** This becomes more evident when we separate the data by participant gender, especially when we look at the results of the women, which showed that in most cases, they had higher emotional values when interacting with VB than just watching it. When the VB had a female name, it was the only case in which the perception of women about the emotions of VBs was similar for both cases, interactive environment and videos. However, the most notable difference between the two studies is that in the video presentation, women perceived more emotions in the VB with female name, whereas men perceived more emotions in the virtual being with male name. In contrast, during the interactive phase, women perceived more emotions in VB with male name than in videos.

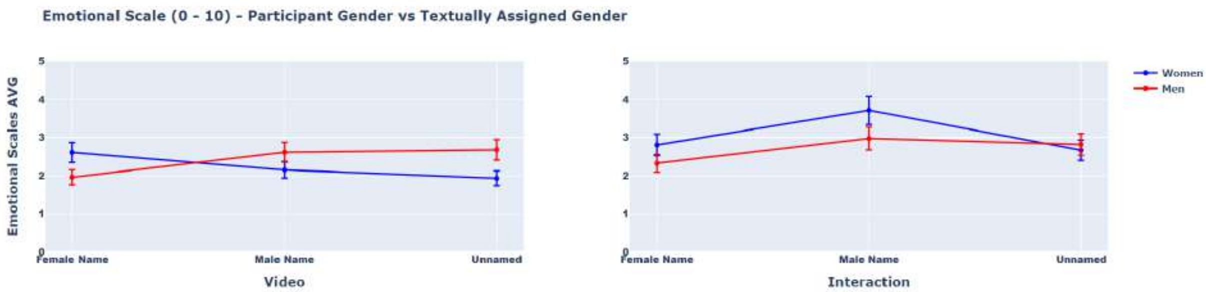


Figure 4.4: Emotional Scale Averages by VB Name between Video and Interaction Stimuli. Blue and red lines are related to women's and men's data.

Gender Attribution Analysis

To assess $H0_{15}$ (*Similarly, people attribute gender to VBs with different genders*), we devised a scoring system ranging from 0 to 2 to capture the participants' responses regarding VB gender attribution concerning the question *What was the VB's gender?*. By employing this scoring system, our objective was to quantitatively assess the participants' perceptions of VB gender attribution. The score for Gender Attribution was determined as follows: *i)* We assigned a score of 2: If the participant's response matched (correctly) the VB textually assigned gender, i.e., VB with female name, VB with male name, and Unnamed VB with "I do not know"; *ii)* We assigned a score of 1: If the participant's response is "I don't know" in the VB with female and male cases, and "Other" in Unnamed case; *iii)* We assigned a score of 0: If the participant answered wrongly the VB gender, i.e., not selecting female for female VB, male for male VB, and male/female for unnamed.

Statistically, in relation to $H0_{15-1}$ about interactive environment, we found a significant result related to a Baby Name's main effect ($F(2, 61)$=5.46 and p=.006; no equal variance - Stats=20.95 and $p< .001$). Then, concerning the post-hoc, we only found significant results in the comparisons between the VB with male name and the other two conditions (female name - p=.03; unnamed VB - $p< .001$). Comparing the results between interactive environment and video stimuli ($H0_{15-2}$), we found two significant results about the female VB, considering all participants in general (Stats=378.0 and p=.009; AVG=1.38 for interactive scenarios, and AVG=1.76 for videos), and from the perspective of men (Stats=100.5 and p=.01; AVG=1.23 for interactive scenarios, and AVG=1.76 for videos). As we can see in the charts of Figure 4.5, both in virtual environment and video, people attributed a gender for the unnamed VB that was not according to the expected ("I don't know"). In this case, most participants chose the "Male" answer for the unnamed VB. Therefore, **rejecting $H0_{15-1}$, we can say that people more correctly attributed a gender accordingly when VB already had a male gender assigned than when it did not have a gender assigned or when it had an assigned female gender. Furthermore, considering only the VB that have names, the gender attribution score was higher when the VB had a male name than when it had a female name.** This ultimate effect (participants offering more precise responses regarding the gender of the male baby) can be attributed to the observation that men made more errors in gender attribution when encountering the VB with female name. In this case, **we can partially reject $H0_{15-2}$, because the gender attribution correctness referring to the female VB in the interactive environment was lower than in the video environment.**

4.2.3 Participant Actions Analysis

In this section, we aim to compare our results with literature regarding interactions between participants and babies in the real world. In this case, we test our hypotheses $H0_{16}$ (*People interact (toy choices, emotional reactions, and proxim-*

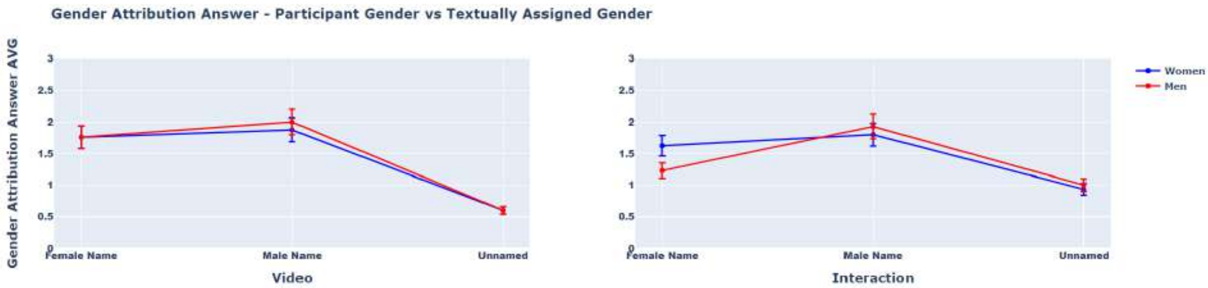


Figure 4.5: Gender Attribution Score Averages by VB Name between Video and Interaction Stimuli. Blue and red lines are related to women's and men's data.

ity) similarly with VBs.) with the equivalent statistical analysis performed in the last sections.

Toys Analysis

Concerning the choice of toys, we found only one significant result regarding an interaction effect ($F(2, 57)$=3.44 and $p<$ .03) between Participant's Gender and Baby's Name when participants chose the Unicorn toy. However, when performing the post-hoc after confirming the equality of variance (Stats=7.19; p=0.20) between groups, we found no significant differences. Table 4.2 shows the average number of toy choices during interactions with the participants. The first thing we can observe in Table 4.2 is a possible link between toys for girls and boys. In general, the unicorn was the most chosen toy on average. In addition, women had a higher average number of toy choices than men. We can see that the unicorn (a pink toy assumed to be more feminine) was chosen on average more by the women who received the VB with female name than by the women who received the male VB (a lesser number of choices) and Unnamed VBs. Additionally, men chose the ball on average (a toy that was hypothesized to be more masculine) more for male VB than for female VB and Unnamed VB. Also, when participants interacted with a baby with female name, women chose more toys than men. When the participants received the VB with male name, the men chose more toys than the women. Women chose more toys than men for an Unnamed VB. Even so, **the results presented in Table 4.2 show a possible indication that women and men chose the unicorn according to the VB's gender. However, in relation to toys, we did not find statistically significant differences ($H0_{16}$).**

Table 4.2: Average number of toy choices, both individually (Unicorn, Ball, Jack-in-the-box) and toys in general. The first four data columns correspond to the responses of the participants who received the questionnaire containing the baby with a female name followed by the columns of the groups that received male names, unnamed, and in general (all). The first three lines refer to women's data, and the last three refer to men's data.

Women																
	Female Name				Male Name				Unnamed				General			
Stimuli	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys
Choice AVG	3.25	2.12	2.00	2.45	2.22	2.00	2.00	2.06	2.40	2.13	2.13	2.22	2.60	2.10	2.07	2.26
Men																
	Female Name				Male Name				Unnamed				General			
Stimuli	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys	Unicorn	Ball	Jack	Toys
Choice AVG	2.38	1.84	1.84	2.02	2.64	2.14	2.00	2.26	2.37	2.12	1.87	2.12	2.48	2.02	1.91	2.14

Emotion Reaction Analysis

We did not find significant results regarding the choice of emojis to represent the participant's emotions. In global averages, as shown in Table 4.3, the positive emoji was the most chosen. Looking at the table, we can see that women selected more emojis than men. Women also chose the positive emoji more than the other two for all VBs. Only for unnamed VB, men select more neutral emojis than positive ones, while women gave more emojis to Unnamed VB than to the other two name groups. Yet, men gave more emojis to the male VB. **These indications make us not rejecting $H0_{16}$ regarding the choice of emojis.**

Table 4.3: Average number of selected emojis, both individually (Positive, Neutral, Negative) and Emojis in general. The first four data columns correspond to the responses of the participants who received the questionnaire containing the baby with a female name, followed by the columns of the groups that received male names, unnamed, and in general (all). The first three lines refer to women's data, and the last three refer to men's data.

Women																
	Female Name				Male Name				Unnamed				General			
Stimuli	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis
Choice AVG	3.25	1.65	1.12	2.00	4.80	1.20	2.60	2.86	4.93	2.40	1.66	3.00	4.42	1.96	1.67	2.69
Men																
	Female Name				Male Name				Unnamed				General			
Stimuli	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis	Positive	Neutral	Negative	Emojis
Choice AVG	2.92	1.84	1.23	2.17	3.42	2.14	1.71	2.45	2.00	2.12	0.87	1.58	2.91	2.20	1.34	2.15

Distance Analysis

One possible action for participants was to navigate the 3D interactive environment freely. We analyzed the amount of time that participants spent in each of the Hall Spaces [Hal66] (Intimate, Personal, Social, Public) concerning their proximity to the VB. As the interaction lasted 3 minutes (fixed time for all), we analyzed participants' average time (seconds) in the four Hall spaces. The participants always started the interaction from the farthest space (Public) from VB. Statistically, we did not find significant results concerning intimate and social spaces. As shown in Table 4.4, our results showed that participants spent less time in intimate and personal than in social and public spaces. Concerning these two closest spaces, only the men, when they received a VB with male name, stayed more than 10 seconds in the personal space. Concerning public and social spaces, statistically comparing, we found two significant main effects of Participant Gender ($F(1, 61)=7.39$ and $p< .008$ for social space; $F(1, 61)=6.35$ and $p< .01$ for public space), one for each of these two spaces. Looking at Table 4.4, we can say that women spent more time in the social space than men. In the case of the public space, women spent, on average, fewer seconds than men. So, **overall, we reject $H0_{16}$ concerning distance, women spent more time close to the VB due to the fact that they spent more time in social space than men, while men spent more time in public space than women.**

Table 4.4: Average interaction time spent by the participants in each Hall proxemics [Hal66] (Intimate, Personal, Social, Public). The first four data columns correspond to the responses of the participants who received the questionnaire containing the baby with a female name and followed by the columns of the groups that received male names, unnamed, and in general (all). The first three lines refer to women's data, and the last three refer to men's data.

Women																
	Female Name				**Male Name**				**Unnamed**				**General**			
Stimuli	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**
AVG	0.00	0.83	153.87	25.32	1.24	3.46	158.54	16.80	0.25	2.37	150.28	27.14	0.35	2.13	152.78	24.77
Men																
	Female Name				**Male Name**				**Unnamed**				**General**			
Stimuli	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**	**Intimate**	**Personal**	**Social**	**Public**
AVG	0.94	8.68	132.50	37.89	2.82	11.06	103.59	62.55	8.06	5.08	117.16	49.10	3.32	8.81	117.43	50.32

4.3 Gender Bias in Relation to Observing Adults VHs

In the previous sections we evaluated gender attribution and emotion evaluation in relation to observation and interaction with VBs. The present section aims to try to answer the question:

- Would the gender-based preference also emerge when observing a genderless virtual adult instead of a VB?

So, to this, we investigated the impact of various features on adult VHs gender attribution and emotion recognition, including: *A)* Participant gender, *B)* VH textually assigned gender, *C)* Gender of Real Actor and Actress, *D)* Pre-processed motion styles, and *E)* through visual features, such as facial shape and mouth. Tables 4.5 and 4.6 summarizes all the studied features and research questions we aim to answer in this work. We aim to investigate the gender attribution of adult VHs, which refers to how users perceive a adult VH along a continuum from masculine to feminine and also measure the impact of gender on emotion recognition. In this context, the endpoints of the continuum would represent the binary gender categories, while the midpoint represents the nonbinary (genderless as well as the analyzes on VBs) category. It is important to note that we are not investigating VH's gender identity, which pertains to one's personal sense of gender. However, we will examine the correlation between the textually assigned gender of VHs and the gender identity of the users. By doing so, we hope to shed light on how users interpret and respond to VHs with different gender presentations, including binary (female and male) and nonbinary 3D models. It is important to address the definitions utilized in our research, per the guidelines provided by APA 6th and 7th Editions[20]. Nonbinary gender, as understood within these frameworks, may or may not align with an individual's assigned sex at birth, presumed gender based on sex assignment, or primary or secondary sex characteristics. In the context of our study focusing on VHs, we have conducted interpolations between binary models representing binary characteristics. This approach enabled us to create a 3D adult model that we refer to as nonbinary, meaning a VH that users could not identify applied tests as a binary gender. For further details, please refer to Section 4.3.1.

Contrary to what was proposed in our last sections, which aimed at studying the gender attribution of VBs, in this section, we focused on the textually assigned gender of a nonbinary (genderless) adult VH, in addition to male and female VHs, during the task of emotion recognition and gender attribution. To establish a ground truth for the VHs with textually assigned genders, a focus group evaluated the three VHs and their genders were attributed based on this analysis. Similar to the studies conducted by Zibrek et al. [ZHRM13, ZHRM15], we assessed the recognition of emotions using data from real actor or actress who drove the animation of the VHs, with a specific focus on facial expressions.

Our study contributes significantly to the field by examining two critical questions about emotion recognition and gender attribution in VHs. Firstly, we investigate how individuals perceive happiness in VHs presented as binary and nonbinary genders (through different 3D facial models and different textual assigned names)

[20]https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/gender

and animated using a variety of stimuli. Secondly, we explore how the same features - face models and textual names - influence gender attribution in VHs. While there is some literature on similar topics, our study is unique in its use of realistic nonbinary, male and female 3D VH models to assess participants' gender attribution of VHs and also in comparing the perception of emotion in different VHs with textually assigned genders. At this point, we believe that our work offers methodology, results, and discussions that are beneficial to show that genderless VHs with textually assigned gender can be a solution to combat gender stereotypes, which was an issue raised in the work of Armando et al. [AOR22]. Additionally, we contribute to the field by incorporating facial styles generated as a function of groups of people's faces. We demonstrate that these animation styles are readily perceivable by participants, even when compared to performances by human actors. To our knowledge, this is the first study to incorporate these elements.

Relating to the originality of this work, different than the last sections, we present an analysis regarding our *RQ*s presented in Tables 4.5 and 4.6. Concerning *RQ*1 and *RQ*3: the results related to the VBs (Sections 4.1 and 4.2) showed that textually assigned genders (using gendered names) could influence the evaluation of emotions about a nonbinary VB, but it was not yet tested with nonbinary adults VHs, and also using binary adult VHs. Regarding *RQ*2 and *RQ*4, several studies in the literature (for example, studies of Zibrek et al. [ZHRM13, ZHRM15] and Ennis et al. [EHEM13]) have shown that the emotion expressed by actors of different genders can influence the recognition of emotion in female and male VHs (without using nonbinary VH). About *RQ*5, some studies evaluated the accuracy of emotion recognition in female versus male models ([ZHRM13, ZHRM15, EHEM13, BB17]); however, we did not find comparisons of binary versus nonbinary models, as in the present work. Concerning *RQ*6, we also measured whether the textually assigned gender of a nonbinary VH impacts gender attribution, but only for a VB, and finally, concerning the *RQ*7, there are papers ([MJH+07, MJH+09, NY20, NVDSN09, GFGW+23]) in the literature that measure whether female, male and androgynous VHs assess the gender attribution. However, there are no reports involving textually assigned gender.

This research is relevant in CG because computer-generated environments should be inclusive spaces where all users feel comfortable ([MMK12, KFMT15]) and represented ([ZHRM13, ZHRM15, BB17]). Additionally, as mentioned by Draude [Dra11], Computer Science applications have the potential to deconstruct biases. Our study is also a significant step forward in understanding how users perceive VHs and the

Table 4.5: The table presents all the research questions presented in our work. In addition, the table links the questions to the experiments and sections in which they are analyzed.

Research Questions
Experiment 1 - Virtual Human with Nonbinary Gender - Section 4.3.1
RQ1: Is there a difference in how people perceive emotions when watching a nonbinary VH with gendered names assigned through text? *This question will examine whether women and men have different emotional responses to a nonbinary VH depending on the gendered names assigned to them.*
RQ2: Is there a difference in how people perceive emotions when presented with a nonbinary VH whose facial expressions are portrayed by both male and female actors? *This question aims to investigate whether the gender of the actors portraying a nonbinary VH's expressions affects emotional perception differently for men and women.*
Experiment 2 - VHs with Female and Male Visual Characteristics - Section 4.3.3
RQ3: Is there a difference in how people perceive emotions when watching a binary VH with gendered names assigned through text? *This question will examine whether women and men have different emotional responses to a binary VH depending on the gendered names assigned to them.*
RQ4: Is there a difference in how people perceive emotions when presented with a binary VH with gendered names assigned through text whose facial expressions are portrayed by various motion stimuli (actors and facial styles)? *This question aims to investigate whether the gender of the actors portraying a binary VH's expressions affects emotional perception differently for men and women.*
Comparing perception of people with respect to binary and nonbinary VHs - Section 4.3.5
RQ5: Is there a difference in how people perceive emotions when presented with binary versus nonbinary VHs with gendered names assigned through text? *This question aims to investigate whether there are gender differences in users emotional perception when presented with VHs that are either binary (having only male or female models) or nonbinary (having a model that is neither strictly male nor female).*
Gender Attribution to VHs - Section 4.3.5
RQ6: Is there a difference in how people attribute gender to VHs with textually defined gendered names? *This question aims to investigate whether there are gender differences in how individuals attribute gender to VHs based on the gendered names assigned to them in text.*
RQ7: Is there a difference in how people attribute gender to VHs with varying gender models (female, male, and nonbinary)? *This question aims to investigate whether there are gender differences in how individuals perceive the gender of VHs when they are presented with different gender models.*

potential impact of in-group biases on these perceptions. By examining the role of bias in adult VH gender attribution, we aim to encourage the development of more inclusive and representative computer-generated environments.

Table 4.6: The table presents the correlations between the features investigated in our study and the research questions posed. Each research question is accompanied by its corresponding analysis goal.

Question	**Participant Gender (Feature A)**	**Assigned Gender (Feature B)**	**Actor Gender (Feature C)**	**Style Gender (Feature D)**	**Model Gender (Feature E)**
Experiment 1 - Virtual Human with Nonbinary Gender - Section 4.3.1					
RQ1	X	X			
RQ2	X	X	X		
Experiment 2 - VHs with Female and Male Visual Characteristics - Section 4.3.3					
RQ3	X	X			
RQ4	X	X	X	X	
Comparing perception of people with respect to binary and nonbinary VHs - Section 4.3.5					
RQ5	X	X	X		X
Gender Attribution to VHs - Section 4.3.5					
RQ6	X	X			
RQ7	X	X			X

4.3.1 Methodology Applied in Experiment 1: Adult VH with Nonbinary Gender

Our first experiment aims to answer research questions *RQ*1 and *RQ*2, as shown in Tables 4.5 and 4.6, by examining how participants perceive the emotions of a nonbinary adult VH that displays facial animations.

Nonbinary Gender 3D Model

To begin with, we required a VH that did not possess male and female features simultaneously, thereby representing a nonbinary gender. This was a crucial step in this work as we aimed to minimize binary gender biases in this study. To create the nonbinary VH, we employed Metahuman Creator [21], which offers a wide range of photorealistic 3D models of VHs and permits manipulation of facial and

[21] https://metahuman.unrealengine.com/

(a) 100M/0F $E = 22.5$ (b) 85M/15F $E = 13$ (c) 65M/35F $E = 3.5$ (d) **50M/50F** $\boldsymbol{E = 1.5}$ (e) 35M/65F $E = 15$ (f) 15M/85F $E = 14$ (g) 0M/100F $E = 15$

Figure 4.6: All variations between Female (Ada) and Male (Bryan) VHs. In addition, the model considered the most nonbinary gender (50M/50F), as evaluated by the focal group, the one highlighted in bold (d). In addition, E presents the result of the evaluation equation for each model.

body features. Furthermore, the software allows for the merging of facial features across different models. With this in mind, we selected two standard Metahuman models - Ada (female) and Bryan (male), and generated intermediate models between them using the slider provided by the Metahuman Creator interface[22]. We generated seven variations by merging the two models, with 50%-50% being an intermediate variation. Additionally, we created three variations with more prominent characteristics of one model (that is, 100%-0%, 85%-15%, 60%-40%) and repeated the same variations with more features of the other model. To ensure consistency and reduce possible biases, we removed the hair from all VHs. While a VH without hair might be biased, we felt that hair and hairstyle could be stronger gender cues and represent more significant biases (as stated in the studies of Hess et al. [HAJK04] and Ghosh et al. [GFGW+23]). All variations are illustrated in Figure 4.6.

After creating the VH models, we conducted a focus group study to determine the most nonbinary model from the seven variations. The focus group consisted of 27 participants, including 12 women, 14 men, and one nonbinary individual. We asked the focus group to rate each variation's femininity and masculinity inspired in the scales used in the work of Shively and Hall [SH77] with two questions, using 5-Likert scales for each one (in the interval [1;5], where one is the least of a specific gender). This step minimizes subjectivity in selecting the nonbinary VH from all the created variations. We then used Equation 4.1 to determine the most nonbinary model:

$$E_{k,N} = \frac{1}{2N}(|\sum_{i=1}^{N} m_{k,i} - 3|, |\sum_{i=1}^{N} f_{k,i} - 3|), \tag{4.1}$$

[22] We tested the facial feature transformations in both directions (from Ada to Bryan and vice versa) and obtained similar individuals. Hence, we used the transformation only from Ada to Bryan.

where k represents each of the 7 models, N states for the number of answers, m states for the answers regarding the masculinity 5-Likert scale, f for the femininity, and $E_{k,N}$ defines how the model k was evaluated. To determine the degree of femininity (f) and masculinity (m) associated with each VH, we calculated the difference between their perceptual value and the central value of the 5-Likert scale (3) and then averaged these differences. Among the evaluated VH models, the 50%-50% model was considered the most nonbinary in terms of gender, as it had the lowest value of $E_{k,N}$ (highlighted in bold in Figure 4.6(d)). It is important to note that this equation yields smaller values for E when participants evaluate the face closer to the central value (e.g., 3 in our case) and when their opinions are contradictory. For instance, if half of the participants chose five and the other half chose 1. In this case, we define the more nonbinary VH as the one that participants find difficult to categorize as either feminine or masculine, that is, escaping from a binary spectrum.

Creation of Stimuli

To answer $RQ1 - RQ2$ in Table 4.5, we created facial emotion stimuli based on filmed actors' facial expressions, and we chose to use only happiness and sadness emotions as proposed in the work of Kohler et al. [KTS+04], that suggests that these are the two most easily recognized emotions. To create these stimuli, we used real videos of an actress and an actor performing expressions of happiness and sadness. The videos range from 9 to 12 seconds and show facial expressions starting from a neutral one. To generate the nonbinary VH animation, we used the Metahuman Creator to export the 3D model (VH (d) in Figure 4.6) and then imported it into the Unreal Engine[23] (4.27 version) using the Quixel Bridge software[24]. To capture the facial expressions of the actress and actor in the videos and transfer them to the nonbinary VH, we used the Live Link Face plug-in[25], which records facial expressions in real-time and transmits them to Unreal engine. A limitation of our work is that we recorded expressions directly from the actors' videos, which can generate some loss in capturing some facial movements. To freeze the head rotation of the VH, we adjusted the Live Link Face settings. Each video was focused on the VH's face, presented frontally, and had a gray background (with a 3D cube added and positioned behind the VH). We generated four videos in total[26], with each video

[23] https://unrealengine.com/

[24] https://quixel.com/bridge

[25] https://dev.epicgames.com/community/learning/tutorials/lEYe/unreal-engine-facial-capture-with-live-link

[26] Examples of all videos are presented in the supplemental material.

featuring the nonbinary VH model and the two actor genders (one female and one male) expressing happiness and sadness (as shown in Figure 4.7).

(a) nonbinary Model - Actress Happy (b) nonbinary Model - Actress Sad (c) nonbinary Model - Actor Happy (d) nonbinary Model - Actor Sad

Figure 4.7: Emotions of our nonbinary model. (a) and (b) represents happiness and sadness performed by an actress, and (c) and (d) by an actor.

Questionnaire

We utilized the Qualtrics platform[27] to design our questionnaire and disseminate it via social media without imposing any restrictions on the participants' device types. Before responding to the questions, participants (volunteers) were presented with the consent form approved by an Ethics Committee of our University[28]. Demographic information of the participants, such as age, gender, educational level, and CG familiarity, were also collected. Regarding gender, we asked two questions: one about "assigned sex at birth" and the other related to "self-identification currently" (both based on APA 7th Edition[29]).

Likewise in the two studies involving VBs, we randomly divided the participants into three groups: in one group, the VH was given a female name (Helena), another a male name (Miguel), and the last group was given a VH without a name. However, the VH was the same for all groups. The VH was presented with a name (either female, male, or no name) and an introductory text. Participants were informed that they would watch four videos of the VH expressing emotions (or not). Participants had to indicate which emotion was most prominent in the video from a list of Ekman's six main emotions ([Ekm92]): happiness, sadness, anger, surprise, fear, neutral, or "I do not know." Finally, with the aim of analyzing gender attribution, participants were asked about the gender of the VH, where the possible answers

[27]https://www.qualtrics.com

[28]Research Ethics Committee of Pontifical Catholic University of Rio Grande do Sul, Brazil - Project Number: 46571721.6.0000.5336

[29]https://apastyle.apa.org/

were female, male, other, and I do not know. Overall, the experiment included three names for the VH (female, male, and unnamed) x 1 VH model (nonbinary gender model) x 2 actor genders as motion stimuli (1 female and 1 male) x 2 emotions expressed (happiness and sadness).

4.3.2 Results

In our results, we used a 2x3x2 Factorial ANOVA analysis (we considered a significance level of 5%) to investigate the effects of participant gender versus VH textually assigned gender versus gender of actor/actress. Specifically, we conducted the Factorial ANOVA analysis to explore the main effects of each variable and any potential interactions between them. Following the ANOVA, we conducted two post-hoc tests to examine further the significant effects observed in the analysis. When the groups of the data had equal variances, we used the Tukey HSD test. We conducted a Bartlett test for each significant effect observed to ensure that the assumption of equal variances was met. In cases where the groups of the data did not have equal variances, we used the Tamhane test.

In total, we had 115 participants, 88 completed the undergraduate studies, 82 were between 18 and 35 years old, and 108 had CG familiarity. About assigned sex at birth, 56 answered female and 59 male, and three answered "Neither of the two options". Regarding current self-identification, women (69.29% were between 18 and 35 years old) were the same that were assigned female at birth, 58 men (approximately 72% were between 18 and 35 years old), and the same three participants answered nonbinary, and another person answered "other". However, because of the low quantity of nonbinary and others, we removed these four people from the analysis. Women and men were divided, respectively, 13 and 25 receiving Helena, 25 and 14 receiving Miguel, and 18 and 19 receiving a VH without a name. One observation was that, in relation to the two models, we tried to control the number of participants per group of names (Helena, Miguel, and Unnamed). However, some participants did not respond until the end of the questionnaire, and therefore we only considered participants who answered the entire questionnaire.

Table 4.7: Percentages of all response options (Happy, Sad, Surprise, Fear, Anger, Other, I do not know) referring to videos that contained Happiness and videos that contained Sadness.

Emotions	Happy	Sad	Anger	Surprise	Fear	Neutral	I do not know
Happy	79.56%	0.43%	0.86%	7.82%	3.91%	2.60%	4.78%
Sad	0.43%	36.95%	13.04%	9.56%	16.08%	3.47%	20.43%

Emotion Recognition Analysis

First of all, to perform the Factorial ANOVA, we created a score of the responses on the recognition of the emotions presented in the videos. Thus, we created an Emotion Score considering = 2 when the participant correctly responded to the emotion presented in the video (Happiness or Sadness responses), = 1 to "I do not know" answers, and = 0 when the participant incorrectly responded to the emotion presented. To analyze the users' correct answers regarding emotion recognition, we first analyze a simple percentage of each detected emotion, as shown in Table 4.7. The obtained results indicate that Happiness was correctly recognized in 79.56% of faces, while Sadness obtained lower correctness (36.95%). Given the low percentage of correct answers for sadness and the significant proportion of participants answering "I do not know" (20.43%), we decided to focus our exploration of human perception solely on the emotion of happiness in the rest of the studies. We chose to limit our investigation to happiness because accurate emotion recognition is crucial for measuring participants' correct perception. By selecting an emotion that is more easily recognized, we aim to minimize potential biases and ensure reliable and meaningful results.

So, using only the participants' happiness recognition (Happiness Score), we report our analysis on the impact of the features *Textually Assigned Gender* and *Actor gender* with the participant gender to answer the questions: *RQ*1) Is there a difference in how people perceive emotions when watching a nonbinary VH with gendered names assigned through text?, and *RQ*2) Is there a difference in how people perceive emotions when presented with a nonbinary VH whose facial expressions are portrayed by both male or female actors?, as shown in Table 4.5. Results are discussed next.

Regarding question *RQ*1 (Participant Gender vs. Textually Assigned Gender, as shown in Tables 4.5 and 4.6, Features A vs. B), we did not find significant results (Table 4.8). However, looking at the charts in Figure 4.8, we can see that men (red lines) had higher values for recognizing happiness when they received

Miguel. Conversely, women had a higher value when the VH had no name but without significant difference. Therefore, **answering *RQ*1, we can conclude that there was no significant difference in how people perceived happiness when they watched a nonbinary VH with Textually Assigned Genders.**

Table 4.8: Table refers to the results of the Factorial ANOVA (2x3x2) performed in Experiment 1 concerning the nonbinary model (using Happiness Score). The table shows the comparison between Participant Gender versus Textually Assigned Gender versus Actor Gender features. In addition, results referring to main and interaction effects. The table also presents the *F*-values and *p*-values (* = significant values, considering a 5% significance level) and the research question representing the comparison.

Nonbinary Model			
Participant Gender - Textually Assigned Gender - Actor Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 228) = 0.01	.90	*RQ*1
Textually Assigned Gender	**F**(2, 227) = 2.55	.07	*RQ*1
Actor Gender	**F**(1, 228) = 4.57	.03*	*RQ*2
Textually Assigned Gender-Participant Gender	**F**(2, 227) = 0.66	.51	*RQ*1
Textually Assigned Gender-Actor Gender	**F**(1, 227) = 2.09	.12	*RQ*2
Actor Gender-Participant Gender	**F**(1, 228) = 5.12	.02*	*RQ*2
Textually Assigned Gender-Actor Gender-Participant Gender	**F**(2, 227) = 0.74	.47	*RQ*2

Regarding *RQ*2 (Features A vs. B vs. C), as we can see in Table 4.8, we found a main effect in relation to the animation stimuli, i.e., the Actor's Gender (Figure 4.8). We also found a significant result in the interaction between the Participant Gender and Actor Gender variables (equal variances - Stats = 1.37 and p=.71). However, we did not find significant results when performing the post-hoc test. In general, people recognized happiness more correctly when they saw the emotion expressed by the actress than by the actor. **Responding to *RQ*2, we only found a significant difference concerning the motion stimuli, i.e., people, in general, recognized the actress's happiness more than the actor's happiness on the nonbinary VH.**

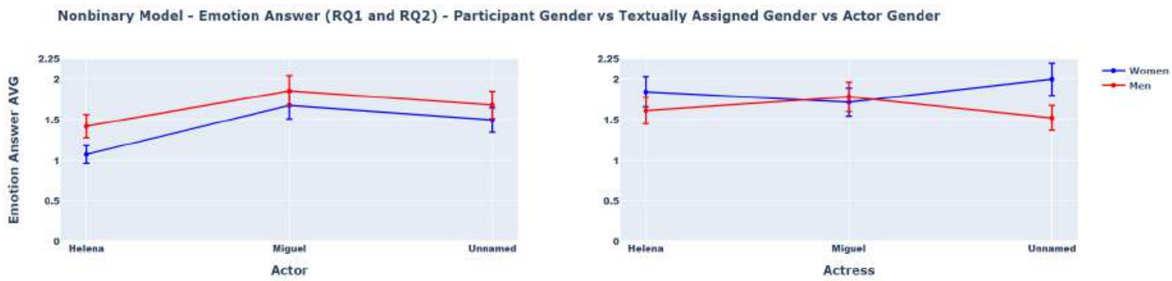


Figure 4.8: The charts show the average of Happiness (Emotion Recognition - *Y*-axis - values from 0 to 2, as explained in Section 4.3.2) answer in relation to the videos from the nonbinary model. The charts show the comparisons between Gender Participant (women and men, respectively, blue and red lines) versus Textually Assigned Gender (*X*-axis) versus Actor Gender (chart on the left presents data relating to the actor's videos, and the chart on the right refers to the actress's videos).

4.3.3 Experiment 2: Adult VHs with Female and Male Visual Characteristics

The second experiment addressed questions $RQ3$ [30] and $RQ4$ [31] presented in Tables 4.5 and 4.6. Our study investigated how humans textually assigned gender to virtual humans when presented with binary (male and female), all animated to display emotions. Additionally, we utilized the emotion styles introduced by [MQM22] to compare with the emotions portrayed by the actor and actress.

Creation of Female and Male 3D Models

We utilized the Metahuman Creator tools to create the male and female VHs. We used the nonbinary VH (Section 4.3.1) as the base and merged it with the binary models (Ada and Bryan, as in Experiment 1) without hair to ensure comparability. Additionally, our designer was free to try to include more female and male details in the shape, e.g., to increase the mouth, to model two binaries. To evaluate the femininity and masculinity aspects of the female and male models, we asked a focus group of 25 participants (15 women and 10 men) to rate them using a 5-point Likert scale. The results showed that the female model was rated as more feminine (Avg=3.44, STD=0.94) and less masculine (Avg=2.4, STD=0.89), while the

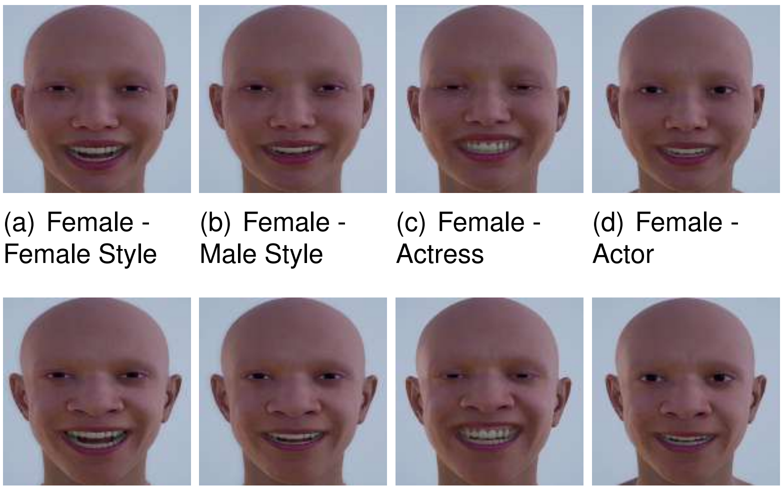
[30] RQ3: Is there a difference in how people perceive emotions when watching a binary VH with gendered names assigned through text?

[31] RQ4: Is there a difference in how people perceive emotions when presented with a binary VH with gendered names assigned through text whose facial expressions are portrayed by various motion stimuli (actors and facial styles)?

male model was rated as more masculine (Avg=4.24, STD=0.7) and less feminine (Avg=1.64, STD=0.74), as expected. The models are depicted in Figure 4.9. It is important to note that the numerical values for the female model are more central in the Likert scale than the male. Indeed, we decided to keep studying such female model because the participants referred to the missing hair and accessories as some aspects that made them evaluate in a more central way. Furthermore, we wanted to focus only on textually assigned gender as a possible gender indicator, so hair could include a more considerable bias, as shown in the studies by Hess et al. [HAJK04] and Ghosh et al.[GFGW+23].

(a) Female (b) Male

Figure 4.9: The binary models used in our work: female (a) and male (b).



(a) Female - Female Style (b) Female - Male Style (c) Female - Actress (d) Female - Actor

(e) Male - Female Style (f) Male - Male Style (g) Male - Actress (h) Male -Actor

Figure 4.10: From (a) to (d) happiness of the female model expressed by, respectively, female style, male style, actress, and actor, while (e) to (h) shows happiness applied to the male model.

In our study, we also utilized a technique proposed by Melgaré et al. [MQM22] to visually depict feminine and masculine styles of expressing happiness. This approach allows us to analyze female and male VHs from the perspective of individuals (actor and actress, as we did in Experiment 1) and from an average of individual characteristics (styles). We selected happiness emotion because it had the best results in Experiment 1, as described in Section 4.3.1. The technique of Melgaré et al. involves processing a dataset of facial images and using a clustering method to group similar facial expressions. To represent each group's style, the technique employs the Facial Action Unit Coding System [EF78] to numerically quantify the facial expressions present in the dataset using Action Units (AUs), which are individual components of facial muscle movement that are activated with various intensities. To generate a representative face for each cluster, we considered the gender annotated in the dataset as the selection factor to create two groups. Then, we averaged the AUs intensities for each cluster and exaggerated the values by 50%.

Table 4.9: The table shows the mapping of Action Units (AUs) with the features present in Unreal's facial control rig, both in relation to the male and female style values.

Action Unit Mapping for Unreal Facial Control Rig				
Facial Control Rig Features	**Face Side**	**Action Units**	**Male Style Values**	**Female Style Values**
jaw	**Center**	**AU-26**	y = 0.18	y = 0.252
jaw-ChinRaiseD	**Right and Left**	**AU-17**	0.16	0.0228
mouth-upperLipRaise	**Right and Left**	**AU-10**	0.363	0.468
mouth-cornerPull	**Right and Left**	**AU-12**	0.674	0.785
mouth-dimple	**Right and Left**	**AU-14**	0.356	0.351
mouth-cornerDepress	**Right and Left**	**AU-15**	0.032	0.006
mouth-stretch	**Right and Left**	**AU-20**	0.064	0.0236
mouth-lowerLipDepress	**Right and Left**	**AU-25**	0.352	0.597
mouth-tightenD	**Right and Left**	**AU-23**	0.18	0.0388
mouth-tightenU	**Right and Left**	**AU-23**	0.18	0.0388
eye-cheekRaise	**Right and Left**	**AU-6**	0.564	0.55
brow-raiseOut	**Right and Left**	**AU-2**	0.082	0.0182
brow-raiseIn	**Right and Left**	**AU-1**	0.035	0.0956
eye-faceScrunch	**Right and Left**	**AU-4**	0.074	0.079
eye-eyelidU	**Right and Left**	**AU-5**	0.143	0.201
eye-eyelidD	**Right and Left**	**AU-7**	0.47	0.469
nose-nasolabialDeepen	**Right and Left**	**AU-9**	0.166	0.208

Using the AUs information of female and male happiness styles we obtained, we applied these styles to our Metahuman models. Since each Metahuman exported to Unreal includes a facial rig control with various features that have associated values, we mapped the AUs data acquired from both the female and male styles to our female and male models (as presented in Table 4.9). To generate animations, we began with a neutral face and applied the extracted AU values for the female happiness style, which we applied to both the female and male models (see Figure 4.10 (a) and (e)). Additionally, we used the male happiness style to both binary models, as shown in Figures 4.10 (b) and (f) for the female and male models, respectively.

To ensure consistency with the first experiment, we followed the same procedure for creating stimuli related to the happiness of actors and the questionnaire (refer to Section 4.3.1 for details). To display the actors' happiness, we used the animations from the first experiment, which are shown in Figure 4.10 for actress and actor happiness in the female VH ((c) and (d)) and the male VH ((g) and (h)).

Regarding the survey, we divided it into two separate questionnaires, one for the female VH and the other for the male VH. For each questionnaire, we used the same three names to textually assign gender, namely, Helena, Miguel, and VH without a name. Therefore, this experiment consisted of two VHs (one female and one male) versus three gendered names (one female, one male, and one VH without name) x two gendered actors (one female and one male) versus two emotional gender styles (one female and one male) versus one emotion (happiness). In the next section, we will discuss the results obtained from this experiment.

4.3.4 Results of Female and Male Models

We used 2x3x2 Factorial ANOVA analysis (also using Tukey HSD or Tamhane as post-hoc tests according to the data variance and significance level of 5%) to investigate the effects (related to the emotion recognition result of the questionnaire of Section 4.3.3) of Participant Gender versus Textually Assigned Gender versus Actor/or Style Gender (2x3x2 analysis).

Regarding the evaluation of the female model, we had 167 volunteers; 142 completed the undergraduate, 113 were at least 36 years old, and 129 had CG familiarity. In relation to assigned sex at birth and self-identification, 93 answered female/women (66% were at least 36 years old) and 74 male/men (65.91% were at least 36 years old). With this, 31 out of 93 women answered the form about the female named VH, 31 about the male, and 31 about the VH without a name. In addition, 74 men also answered the questionnaire (female name 29, male 24, VH without name 21).

Regarding the evaluation of the male model, we had 181 volunteers: 151 completed at least undergraduate, 133 were at least 36 years old, and 147 had CG familiarity. About assigned sex at birth and self-identification, 100 female/women (that answered our research about a female named VH 37, male 32, VH without name 31, and 64.2% were at least 36 years old) and 81 male/men (female name 26, male 27, VH without name 28, and 81% were at least 36 years old).

Analyzing the female and male (binary) models in the perceptual analysis of women and men (*RQ*3 - Section 4.3.4 - Participant Gender versus Textually Assigned Gender, and *RQ*4 - Section 4.3.4 - Participant Gender versus Textually Assigned Gender versus Actor Gender/or Style Gender), we first split the data into two groups, i.e., when the motion stimuli were captured from the actors, and when facial styles generated the motion stimuli. Furthermore, in Section 4.3.4, we compare Styles versus Actor Gender.

Emotion Recognition Analysis (Participant Gender versus Textually Assigned Gender)

Regarding *RQ*3 (Features A vs. B from Tables 4.5 and 4.6), considering only the data containing stimuli captured from the actors and the two models (female and male), as we can see in Tables 4.10 and 4.11, we did not find any significant results. Looking at the charts in Figure 4.11, we can visually see that the results are similar between women and men in relation to Textually Assigned Gender, even if we separated for male and female models. **Answering *RQ*3 (in data containing only actor/actress happiness), people perceived happiness similarly when they observed a binary VH with Textually Assigned Gender.**

Still related to *RQ*3, when we consider only the data containing stimuli captured from the women/men facial style (Figure 4.12), as we can see in Table 4.10, we found a main effect about Textually Assigned Gender when we measured the female model data (no equal variance - Stats = 7.90 and p=.01). So, in the post-hoc, we found significant p-values in the comparisons between Helena versus Miguel (p = .002) and Helena versus Unnamed (p = .01). Therefore, as we can see in Figure 4.12 related to the female model, we can say that when participants received

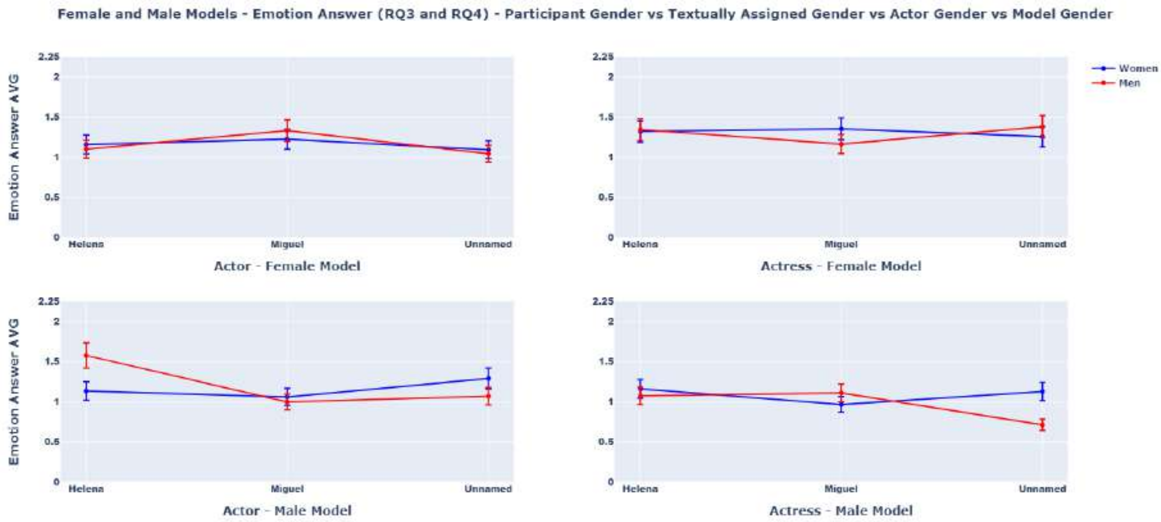


Figure 4.11: The charts show the averages of Emotion answer (Y-axis) related to the videos containing only Actor/Actress's Motion Happiness Stimuli from the Female (two charts at the top) and Male (two charts at the bottom) models in the comparisons between Gender Participant (women and men, respectively, blue and red lines) versus Textually Assigned Gender (X-axis) versus Actor Gender (the two charts on the left and the two charts on the right).

Helena, they recognized Happiness Style more accurately than people who received Miguel or unnamed VH. Therefore, **answering** *RQ*3 **(in motion stimuli containing only style), the participants perceived more happiness in the female model named Helena than in the female model named Miguel, and then in the female model unnamed VH. We did not obtain significant results when we tested the male model (Table 4.11).**

Emotion Recognition Analysis (Participant Gender versus Motion Stimuli: Actor Gender or Style Gender)

Regarding *RQ*4, firstly considering only videos containing actors as motion stimuli (Features A vs. C), we only found a main effect of Actor Gender when we measured the male model data, as can be seen in Table 4.11. With that, **answering** *RQ*4 **(on data containing only actors motion)**, as we can see in Figure 4.11, **people recognized more correctly the happiness of the Actor than the happiness of the Actress when they received a Male Model. We did not find a significant result for the female model (Table 4.10). In addition, regarding the data that**

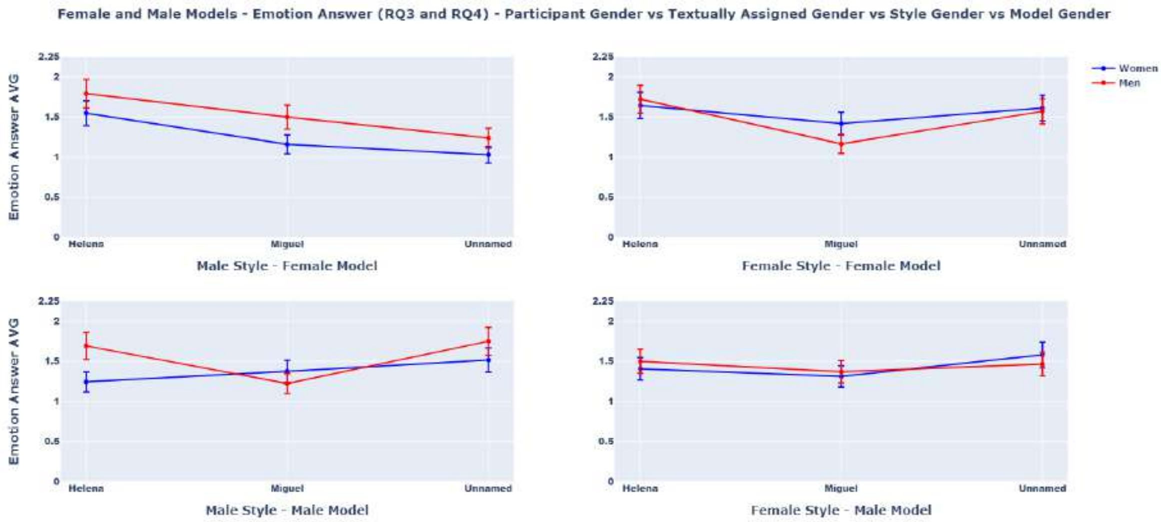


Figure 4.12: The charts show the averages of Emotion answer (Y-axis) about the videos containing only data of Style Happiness from the Female (two charts at the top) and Male (two charts at the bottom) models in the comparisons between Gender Participant (women and men, respectively, blue and red lines) versus Textually Assigned Gender (X-axis) versus Style Gender (the two charts on the left and the two charts on the right).

only contained gender facial styles (Features A vs. D), we also did not find significant results.

Emotion Recognition Analysis (Facial Style gender versus Actor gender)

Furthermore, we conducted comparisons between facial styles and actors' movements in our analysis. Employing 2x2 Factorial ANOVA, we examined the interaction between Motion Stimuli (Facial Style/Actor Movement) and Participant's Gender. However, we only observed a significant result in relation to the main effect of the Motion Stimuli (**F**(1, 1390)=7.51 and p=.006). Upon examining Figures 4.11 and 4.12, **it becomes evident that participants were more adept at accurately recognizing facial styles of happiness compared to identifying happiness conveyed through actors' motion**. **Comparing the Figures 4.11 and 4.12, it becomes apparent that men demonstrated greater proficiency in correctly identifying facial styles of happiness compared to recognizing happiness expressed through motion of actors. Conversely, this distinction was not as pronounced among women, where the difference in recognition between facial styles and actors' motion was less evident.**

Table 4.10: Table refers to the results of the Factorial ANOVA (2x3x2) performed in Experiment 2 in relation to the Female model (using Happiness Score). The table is divided into two parts about comparisons between the Participant Gender versus Textually Assigned Gender Actor Gender/Style Gender features. The first part regarding the female model is related only to data containing videos with actors' emotions, and the second part contains videos with emotional styles. The table shows the results of main and interaction effects. The table also presents the *F*-values and *p*-values (* = significant values, considering a 5% significance level) and the research question that represents the comparison.

Female Model - Only Containing Actor/Actress Videos			
Participant Gender - Textually Assigned Gender - Actor Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 332) = 0.27	.60	*RQ*3
Textually Assigned Gender	**F**(2, 331) = 0.11	.89	*RQ*3
Actor Gender	**F**(1, 332) = 0.93	.33	*RQ*4
Textually Assigned Gender-Participant Gender	**F**(2, 328) = 0.05	.94	*RQ*3
Actor Gender-Participant Gender	**F**(1, 330) = 0.11	.73	*RQ*4
Actor Gender-Textually Assigned Gender	**F**(2, 328) = 0.48	.61	*RQ*4
Actor Gender-Textually Assigned Gender-Participant Gender	**F**(2, 328) = 0.53	.58	*RQ*4
Female Model - Only Containing Style Videos			
Participant Gender - Textually Assigned Gender - Style Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 332) = 1.04	.30	*RQ*3
Textually Assigned Gender	**F**(2, 331) = 6.29	= .002*	*RQ*3
Style Gender	**F**(1, 332) = 2.38	.12	*RQ*4
Textually Assigned Gender-Participant Gender	**F**(2, 328) = 0.14	.86	*RQ*3
Style Gender-Participant Gender	**F**(1, 330) = 3.20	.07	*RQ*4
Textually Assigned Gender-Style Gender	**F**(2, 328) = 2.69	.06	*RQ*4
Textually Assigned Gender-Style Gender-Participant Gender	**F**(2, 328) = 0.49	.69	*RQ*4

4.3.5 Comparing Both Experiments - Nonbinary and Binary models

We used 2x3x2x3 and 2x3x3 Factorial ANOVAs (Tukey HSD or Tamhane as post-hoc tests according to the data variance and a significance level of 5%) to compare the results of both presented experiments and to answer *RQ*5 [32], *RQ*6 [33], and *RQ*7 [34]. The following sections detail such analyses.

[32] RQ5: Is there a difference in how people perceive emotions when presented with binary versus nonbinary VHs with gendered names assigned through text?

[33] RQ6: Is there a difference in how people attribute gender to VHs with textually defined gendered names?

[34] RQ7: Is there a difference in how people attribute gender to VHs with varying gender models (female, male and nonbinary)?

Table 4.11: Table refers to the results of the Factorial ANOVA (2x3x2) performed in Experiment 2 in relation to the Male model (using Happiness Score). The table is divided into two parts about comparisons between the Participant Gender versus Textually Assigned Gender Actor Gender/Style Gender features. The first part relates only to data containing videos with actors' emotions, and the second contains videos with emotional styles. The table shows the results of main and interaction effects. The table also presents the *F*-values and *p*-values (* = significant values, considering a 5% significance level) and the research question that represents the comparison.

Male Model - Only Containing Actor/Actress Videos			
Participant Gender - Textually Assigned Gender - Actor Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 332) = 0.008	.92	*RQ*3
Textually Assigned Gender	**F**(2, 331) = 1.11	.32	*RQ*3
Actor Gender	**F**(1, 332) = 4.18	.04*	*RQ*4
Textually Assigned Gender-Participant Gender	**F**(2, 328) = 1.68	.18	*RQ*3
Actor Gender-Participant Gender	**F**(1, 330) = 0.55	.45	*RQ*4
Actor Gender-Textually Assigned Gender	**F**(2, 332) = 0.82	.43	*RQ*4
Actor Gender-Textually Assigned Gender-Participant Gender	**F**(2, 332) = 0.97	.37	*RQ*4
Male Model - Only Containing Style Videos			
Participant Gender - Textually Assigned Gender - Style Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 332) = 1.04	.30	*RQ*3
Textually Assigned Gender	**F**(2, 331) = 2.56	.07	*RQ*3
Style Gender	**F**(1, 332) = 0.08	.76	*RQ*4
Textually Assigned Gender-Participant Gender	**F**(2, 328) = 1.04	.35	*RQ*3
Style Gender-Participant Gender	**F**(1, 330) = 0.79	.37	*RQ*4
Textually Assigned Gender-Style Gender	**F**(2, 328) = 0.23	.79	*RQ*4
Textually Assigned Gender-Style Gender-Participant Gender	**F**(2, 328) = 1.02	.36	*RQ*4

Comparing Perception Concerning Binary and Nonbinary Adult VHs

First, we disregarded data referring to the styles to compare the two experiments because Experiment 1 (Section 4.3.1), with nonbinary VH, was not animated using facial motion styles. Concerning *RQ*5 (features A vs. B vs. C vs. E - Tables 4.5 and 4.6), as shown in Table 4.12, we found a main effect of Model Gender (no confirmed variance, Stats=16.91 p$<$ .001) and an interaction effect between Model Gender and Actor Gender features (no variance, Stats=27.23 p$<$ .001). Looking at Figure 4.13, we can see that emotion recognition scores were higher when people viewed the nonbinary model than the other two. In relation to the Model Gender effect, after the post-hoc, we found significant values in the comparisons between the nonbinary model and the other two models ($p < .001$ both for female and male). Indeed, concerning the interaction effect of Model Gender and Actor Gender, as shown in Table 4.13 resulting from post-hoc, we found several significant *p*-values in the pairwise comparisons between all model gender and actor gender options.

Looking at Figure 4.13 and the results in Table 4.12, we can see that, in most cases, people recognized the emotion more correctly when the nonbinary VH had the emotion performed by an actress. When an actor performed the nonbinary VH emotion, people recognized the emotion in a similar way to the other models. Furthermore, people recognized the emotion of a female VH performed by an actress more correctly than the emotion of a male VH performed by an actress. With this, **answering *RQ*5, as we can see in the results presented in Experiments 1 and 2, people recognized more correctly happiness when the VH had a nonbinary model than a VH with female and male models. Furthermore, we found indications that emotion performed by an actress may increase the correctness of perceived emotion when the VH had a female model compared to a male model.**

Table 4.12: Table refers to the results of the Factorial ANOVA (2x3x2x3) performed comparing the three models presented in our work (Nonbinary, Female, and Male models) versus the Participant Gender (using Happiness Score as outcome) versus Textually Assigned Gender (Helena, Miguel, Unnamed) versus Actor Gender (Actor and Actress). The table also presents the *F*-values and *p*-values (* = significant values, considering a 5% significance level) and the research question that represents the comparison.

Participant Gender - Model Gender - Actor Gender - Textually Assigned Gender			
Factor-Emotion Answer	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 924) = 0.02	.87	*RQ5*
Model Gender	**F**(2, 923) = 25.16	$<$ .001*	*RQ5*
Textually Assigned Gender	**F**(2, 923) = 0.21	.80	*RQ5*
Actor Gender	**F**(1, 924) = 1.12	.28	*RQ5*
Actor Gender-Textually Assigned Gender	**F**(2, 920) = 0.57	.56	*RQ5*
Actor Gender-Participant Gender	**F**(1, 924) = 3.05	.08	*RQ5*
Textually Assigned Gender-Participant Gender	**F**(2, 920) = 1.20	.30	*RQ5*
Actor Gender-Model Gender	**F**(2, 920) = 3.88	.02*	*RQ5*
Model Gender-Textually Assigned Gender	**F**(4, 918) = 1.80	.12	*RQ5*
Model Gender-Participant Gender	**F**(2, 920) = 0.04	.95	*RQ5*
Actor Gender-Textually Assigned Gender-Participant Gender	**F**(2, 920) = 0.43	.64	*RQ5*
Actor Gender-Participant Gender-Model Gender	**F**(2, 920) = 0.97	.37	*RQ5*
Textually Assigned Gender-Participant Gender-Model Gender	**F**(4, 918) = 0.79	.52	*RQ5*
Actor Gender-Textually Assigned Gender-Participant Gender-Model Gender	**F**(4, 918) = 0.94	.43	*RQ5*

Gender Attribution Analysis

First, as shown in Section 4.3.1, at the end of the questionnaire, we asked participants, "What was the VH's gender?" to investigate whether people attribute the VH gender according to the textually assigned gender. This section compares the results of the nonbinary VH versus binary VHs (related to *RQ*6 and *RQ*7, as shown in Tables 4.5 and 4.6). Before evaluating our research questions, we compared the results related to the nonbinary adult VH with the result of the VB (only

Table 4.13: Confusion Matrix of *p*-values generated by the post-hoc Tamhane test of the interaction between the variables Model Gender and Actor Gender in relation to Emotion Answer.

Models / Actor(ress)	Nonbinary Actress	Female Actress	Male Actress	Nonbinary Actor	Female Actor	Male Actor
Nonbinary Actress	1.00	$<$ .001*	$<$ .001*	.31	$<$ .001*	$<$ .001*
Female Actress	$<$ .001*	1.00	.03*	.11	.65	.75
Male Actress	$<$ .001*	.03*	1.00	$<$ .001*	.74	.54
Nonbinary Actor	.31	.11	$<$ .001*	1.00	.002*	.002*
Female Actor	$<$ .001*	.65	.74	.002*	1.00	.99
Male Actor	$<$ .001*	.75	.54	.002*	.99	1.00

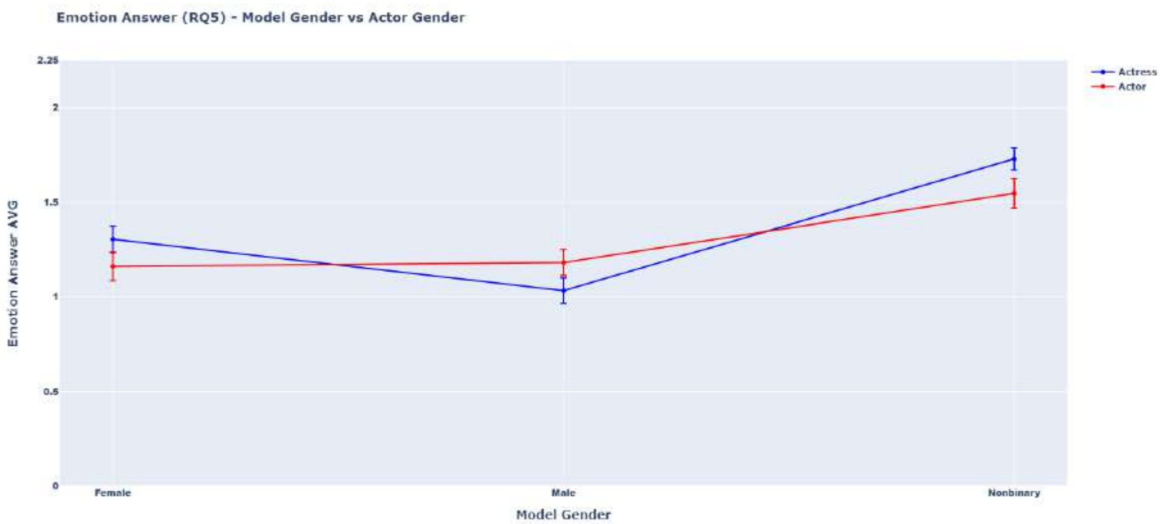


Figure 4.13: The line charts show the average of the Emotion answer (Y-axis) in the comparisons between Actor Gender (actor and actress, respectively, blue and red lines) versus Model Gender (respectively, female model, male model, and nonbinary model on the X-axis).

video stimuli of Section 4.1). Once only a nonbinary VH (VB) was used, we only used the results referring to the nonbinary adult VH, in the present section. We used the *Chi-square* test to compare Adult VH and VB in the three different name cases (2x3 analysis - considering the possible answers of the gender attribution question), and the results did not present a significant difference ($X^2(2, 86)=0.11$ and p = .73

for VHs with the name Helena, $X^2(2, 86)=0.85$ and $p=.65$ for the name Miguel, $X^2(2, 87)=0.54$ and p = .76 for VHs without names). Therefore, **for both a nonbinary virtual adult and a genderless VB, the gender attribution was similar. In this case, we can say that our methodology for creating and evaluating a genderless virtual adult was efficient, as we achieved similar results in relation to the genderless VB.** In addition, for both VHs (baby and adult), men and women tend to evaluate a nonbinary VH as male.

To assess our research questions (*RQ*6 and *RQ*7), we devised a scoring system from 0 to 2 to capture the participants' responses regarding VH gender attribution. The score for Gender Attribution was determined as follows:

- We assigned a score of 2: If the participant's response matched (correctly) the VH textually assigned gender. For instance, if the participant received a VH named Helena, who was assigned the gender of female, and responded that the VH's gender was female. Regarding Unnamed VB, a "I do not know" answer was considered correct;

- We assigned a score of 1: If the participant's response is "I don't know" in the VH with female and male cases, and "Other" in Unnamed case;

- Finally, if the participant did not agree with any of the provided options, a score of 0 was assigned. For example, the participant did not select female for female VH, male for male VH, and male/female for unnamed.

By utilizing this scoring system, we aimed to quantitatively evaluate the participants' perceptions of VH gender attribution about our research questions (*RQ*6 and *RQ*7).

In the comparison between Participant Gender and Textually Assigned Gender (*RQ*6 - Features A vs. B), we can see in Table 4.14 that we found only one significant result concerning the main effect of Textually Assigned Gender (equal variance confirmed using Bartlett's test - Stats=1.53 and p=.46). The post-hoc revealed two significant results in the pairwise comparisons between the VHs with names and the unnamed VH ($p < .001$ for both the comparison with Helena and Miguel). **Responding to *RQ*6, people attributed a gender to the VHs according to the textually assigned gender. Therefore, we can say that participants who received Miguel or Helena provided more correct responses in accordance with the textually assigned gender than participants who received the**

unnamed VH. Furthermore, as we can see in Figure 4.14, in terms of participant gender, women and men (respectively, blue and red lines) assigned gender to VHs similarly in all cases.

Regarding the comparison between Participant Gender, Textually Assigned Gender and Model Gender (*RQ*7 - Features A vs. B vs. E), as shown in Tables 4.10 and 4.11, we found significant results concerning the main effect of the Gender Model (with data variance, stats=0.54, and p = .76) and in relation the interaction between Model Gender and Textually Assigned Gender (no variance, stats=90.95, and $p < .001$). With regard to the main effect, in pairwise comparisons, we found two significant results in the comparison of the nonbinary model versus the other two binary models (p = .01 for the female model and p = .007 for the male model). Concerning the interaction effect, as you can see in Table 4.15, we found several significant p-values in pairwise comparisons between VHs with different models and names. Both looking at these values and looking at Figure 4.14, we can notice that people who received the nonbinary VH with both female and male names attributed gender more correctly than people who received the female VH with a male name, and then the male VH with a female name. Furthermore, people attributed gender less correctly when any VH did not have a name. That is, **answering question *RQ*7, people who saw the nonbinary model attributed more correctly the VH gender according to the textually defined name. It can indicate that the female or male models may have caused issues in the VH gender attribution, while the nonbinary model, with fewer female or male visual characteristics, may have made people remember the gender through the textual given name.**

Table 4.14: Table refers to the results of the Factorial ANOVA (2x3x3) performed about the question "What was the VH's gender?" (using Gender Attribution Answer Score). The table presents the results of comparing the Participant Gender, Textually Assigned Gender and Model Gender features (main and interaction effects). The table also presents the F-values and p-values (* = significant values, considering a 5% significance level) and the research question that represents the comparison.

Participant Gender - Textually Assigned Gender - Model Gender			
Factor-Gender Attribution Score	**F-stats**	**p-value**	**RQ**
Participant Gender	**F**(1, 461) = 0.03	.91	*RQ*6
Textually Assigned Gender	**F**(2, 460) = 108.28	$<$ .001*	*RQ*6
Model Gender	**F**(2, 460) = 8.71	$<$ .001*	*RQ*7
Textually Assigned Gender-Participant Gender	**F**(2, 460) = 0.33	.71	*RQ*6
Model Gender-Participant Gender	**F**(1, 457) = 0.16	.85	*RQ*7
Model Gender-Textually Assigned Gender	**F**(4, 458) = 23.97	$<$ .001*	*RQ*7
Participant Gender-Model Gender-Textually Assigned Gender	**F**(4, 458) = 0.12	.28	*RQ*7

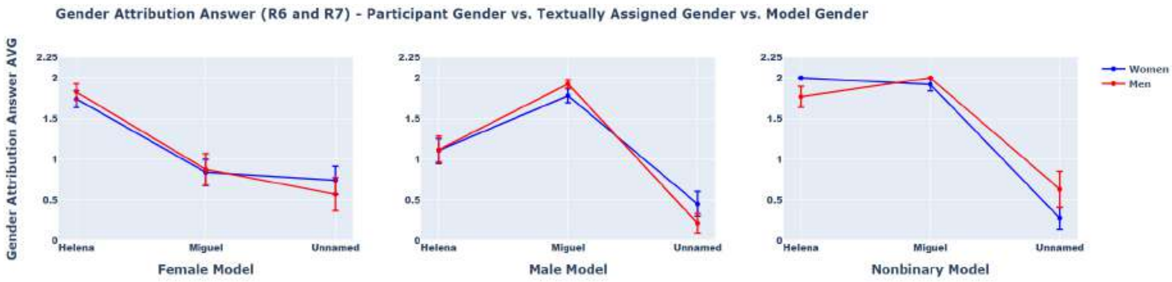


Figure 4.14: The line charts show the average of the Gender Attribution answer in the comparisons between Participant Gender (women and men represented by the blue and red lines respectively) versus Model Gender (respectively, female model in the first chart, male model in the middle chart, and nonbinary model in the last chart) versus Textually Assigned Gender (x-axis on each chart). The y-axis represents the agreeing responses on the gender previously assigned textually. So, for example, if the name of the VH was Helena, and the participants responded that the gender of the VH was female, the answer was correct. More information on counting correct answers in Section 4.3.1.

4.4 Chapter Considerations

This section discusses the main findings of this chapter, and it is divided into: Section Discussion About Gender Bias in the Perception of VBs in Video Stimuli presents the discussions on the results of Section 4.1, Section Discussion about Gender Bias in the Perception of VBs in Interactive Stimuli presents the discussions on the results of Section 4.2, and finally Section Discussion about Gender Bias in the Perception of Adult VHs in Video Stimuli presents the discussions on the results of Section 4.3.

4.4.1 Discussion About Gender Bias in the Perception of VBs in Video Stimuli

Following the results of Section 4.1, as discussed in Chapter 1, in Draude's work [Dra11], the author made an analogy with Alan Turing's Imitation Game [Tur09], pointing out that computing has the potential for gender deconstruction, since a machine is essentially genderless. To assess this, the $H0_{11}$ hypothesis was raised, and it was tested through a study where a VB without visual gender stereotypes was evaluated, but with genders defined through names presented to different groups of participants. In this case, most participants who received the genderless baby de-

Table 4.15: Confusion Matrix of *p*-values generated by the post-hoc Tamhane test of the interaction between the variables Model Gender and Textually Assigned Gender concerning Gender Attributions.

Models / Names	Female Helena	Female Miguel	Female Unnamed	Male Helena	Male Miguel	Male Unnamed	Nonbinary Helena	Nonbinary Miguel	Nonbinary Unnamed
Female Helena	1.00	$<.001$*	$<.001$*	$<.001$*	.99	$<.001$*	.99	.44	$<.001$*
Female Miguel	$<.001$*	1.00	.96	.68	$<.001$*	.01*	$<.001$*	$<.001$*	.24
Female Unnamed	$<.001$*	.96	1.00	.09	$<.001$*	.33	$<.001$*	$<.001$*	.92
Male Helena	$<.001$*	.68	.09	1.00	$<.001$*	$<.001$*	$<.001$*	$<.001$*	.002*
Male Miguel	.99	$<.001$*	$<.001$*	$<.001$*	1.00	$<.001$*	1.00	.81	$<.001$*
Male Unnamed	$<.001$*	.001*	.33	$<.001$*	$<.001$*	1.00	$<.001$*	$<.001$*	.99
Nonbinary Helena	.99	$<.001$*	$<.001$*	$<.001$*	1.00	$<.001$*	1.00	.96	$<.001$*
Nonbinary Miguel	.44	$<.001$*	$<.001$*	$<.001$*	.81	$<.001$*	.96	1.00	$<.001$*
Nonbinary Unnamed	$<.001$*	.24	.92	.002*	$<.001$*	.99	$<.001$*	$<.001$*	1.00

fined that the VB was male. This is in line with the work of Armando et al. [AOR22], who stated that people tend to attribute male gender to genderless VHs. In addition, this result may not be in line with what was scored by Draude (and this is repeated both for interaction with VB and for the virtual adult in relation to gender attribution in genderless VHs, as discussed in the next sections). Furthermore, still in the Imitation Game, Alan Turing himself suggested that it would be easier for the machine to have a masculine gender and then imitate the woman, than to be of the feminine gender. Which shows that these thoughts have been rooted in human culture for a long time. Along the same lines, in games, most people choose to create male VHs, and the movie industry has developed stereotyped female VHs, as shown in the work of Garza et al. [GAHH19]. However, with the results shown through $H0_{10}$, we can see that people can perceive emotions in a gender-defined baby even if that baby does not have gender stereotypes. Of course this can be related to gender identification or in-group advantage [GBBM07, EA02, KST+15], for example, women identify more with female VHs (as we can see throughout this present work), but this means that the industry may not worry about creating stereotyped VHs.

4.4.2 Discussion about Gender Bias in the Perception of VBs in Interactive Stimuli

In summary, just looking at the averages people perceived emotions similarly in the interactive VB with different names and genders ($H0_{14-1}$). In general, looking at the averages of perceived emotions in the chart in Figure 4.4, people perceived more emotions in male VB than in female VB. Men also perceived more emotions in male VB than in female VB. This result for men is in line with was discussed last section on video stimuli, and in line with the study of Condry and Condry [CC76], in which men also perceived more emotions in babies with a male name. Even though this result requires more data and future analysis, it may indicate an in-group advantage effect [GBBM07], that is, a preference for men to evaluate male VB more emotionally than female VB. In this case, as we discussed last section, the real-life bias was maintained in virtual environments, both in videos and interactions. Women, on the other hand, did not have results in accordance with in-group theories, as they perceived fewer emotions in the female interactive VB than in male and the unnamed interactive VBs. However, the small number of women volunteers in female and male VB's groups may have influenced this result, and also the result comparing women and men. Regarding $H0_{14-2}$, overall, our results showed that people perceived more emotions in interactive VB than in VB presented in videos. Considering VB's gender, this result is repeated only when VB name had a male name. A possible explanation is due to the different stimuli, that is, videos and 3D interactive environments. We hypothesize that the difference in stimuli can affect the users' perception. It may be related to the fact that the Games industry, which is very interactive, uses more realistic characters with facial animations that one could classify as falling in the UV [MMK12] theory if it was to watch in a movie, but works well for games. Furthermore, people who interacted with a VB in the 3D environment rated it as more emotional than people who only watched videos. It may indicate that, in interactive environments, with greater participant immersion, there may be a more straightforward perception of virtual characters' emotions, as we hypothesize a reduction of the UV effect. When we separated participants by gender, we only found differences when the participants were women. Overall, women also perceived more emotions in interactive VBs. When separating the VB by gender, women perceived more emotions in VB with male name and VB without a name. In this case, women once again had an out-group advantage, that is, they had preferences in evaluating a VB with male name emotionally.

Regarding gender attribution by participants ($H0_{15}$), the results regarding interaction with VBs were similar to the results of video stimuli, except when VB had a female name. In this case, people attributed to the VB a female gender more ”correctly" in videos than in interactive environments. In both studies, when a VB is presented with a defined gender using a male name, people tended to attribute the male gender ”correctly", that is, according to the name and gender of the VB. Contrary to what we discussed about video stimuli, just looking at the averages, people attributed gender similarly when VB had a female name comparing to when VB had no name, and differently when it had a female name comparing to when VB had a male name. Also, as in the video stimuli, people attributed more masculine gender to the unnamed VB than any other gender, even though babies are considered genderless [CC76]. These indications show that even without visual indications referring to any binary gender (such as a dress, blue color, or other accessory), a gender bias occurred. Therefore, as discussed in the previous section on video stimuli, this result is in the same line as Armando et al. [AOR22], showing that the male gender may be the most attributed to genderless VBs. In addition, men also answered more incorrectly the gender attribution of the interactive female VB than in videos. This result is in line with what we mentioned in the previous paragraph, i.e., increasing an out-group disadvantage. In this case, men had more difficulty attributing the female gender to the VB with female name in the interactive environment than in the video environment. One hypothesis is that, in interactive environments, men may have felt more comfortable with the male VB (in-group advantage because of the male name) and more uncomfortable with the female VB (out-group disadvantage because of the female name). Another interesting result related to $H0_{15}$ indicated that, as in the videos, women who interacted with a female VB attributed gender more correctly to the VB with female name than others. All these results are in the same line of gender anonymity issues ([MMLD20, KMJ+24, BL16, LNB14]), where a previously defined gender (female and male names) has different results compared to the case in which the VB had no previously defined gender (no name).

Regarding $H0_{16}$, we found patterns in comparing our results with those presented in the literature on interactions between real babies and humans. However, through statistical analysis we did not find many significant results. The different animation times and the ordering of the toys on the screen (these two cases can be manipulated in future studies) may have affected these results, and this is a limitation of our work. Looking at the results regarding the choice of toys, for example, in our work, women chose toys considered neutral (Jack-in-the-box) and feminine (unicorn) more than men, as well as in the work of Culp et al. [CCH83]. Women chose the female toy more for the female VB than for the male VB, as also in the studies by Culp et al., Will et al. [WSD76], Seavey et al. [SKZ75], and Sidorowicz and Lunney [SL80]. Men gave more female toy than male and neutral toys to the unnamed VB, and also gave more male toy than the neutral one to the unnamed VB, as in the work of Sidorowicz and Lunney. Additionally, we associated the emoji selection with the emotion the participants felt during the interaction. In our work, men gave the positive emoji more for the VB with a male name than the other VBs; that is a result similar to the work by Culp et al., in which men smiled more at the male baby. Therefore, we can consider that behaviors and biases of interactive experiments between babies and adults can also happen in virtual interactive environments. Still, in $H0_{16}$, regarding distances, the participants spent little time in intimate and personal spaces. Women spent more time in social spaces than men, while men spent more time in public spaces than women. So, we can say that the women preferred to stay in a space closer to the VB than the men. This result is not entirely in line with Greeno's work [Gre89], which stated that both the baby's gender and the adult's gender influence the proximity between them. The proximity measure is generally more used in studies containing VR. Indeed, our results showed indications that proximity changed based on the participant's gender, which can be investigated in a future VR experiment. For example, in some VR studies, this result is contrary to the results presented in studies by Zibrek et al. [ZNO+20a, ZNO+20b, ZNO+22], in which women preferred to stay at greater distances than men in relation to VHs with different genders. Of course, the cited studies presented adult VHs, and we presented VBs for the participants.

4.4.3 Discussion about Gender Bias in the Perception of Adult VHs in Video Stimuli

In this section, we summarize the main findings of results related to the Section 4.3. Regarding *RQ*1 (Tables 4.5 and 4.6), in general, men recognized happiness better when the nonbinary VH was named Miguel (male name) than when was Helena (female name), while women did not present such effect when the nonbinary VH was named Helena (see Figure 4.8). This result, for women participants, was a contrary effect to what happened regarding the evaluation of emotion in video stimuli related to VBs, but similar results to the stimulus related to interaction with VBs. In other words, there was no group effect for women, but there was for men when they recognized happiness in the nonbinary VH. Also, unlike VB in video stimuli, we did not find significant results comparing women and men. This may be related to the fact that a VH without hair may have been perceived more as male, as it is rarer in the real world for a female to be without hair. Therefore, the lack of hair ([HAJK04, GFGW+23]) in the VH may have been a bias in this case. Furthermore, this may have happened due to the difference in stimuli between the two studies. That is, we had real emotions to be classified behind a nonbinary adult VH made by an actress and actor, while in the other case, we studied a VB without explicitly expressing emotions.

The findings we obtained concerning *RQ*2 align with previous studies conducted by Zibrek et al. [ZHRM13, ZHRM15], Ennis et al. [EHEM13] and McDuff et al. [MKKL17] on emotion recognition, and even about women expressing more happiness than men. Specifically, in our case, we explored the effectiveness of emotions expressed by actors in animating a nonbinary VH. Our results indicate that all participants could recognize emotions portrayed by the actress more readily than emotions conveyed by the actor. This trend is represented in Figure 4.8, which data is presented in Table 4.8. Furthermore, we found an indication that women had an advantage in perceiving happiness better in an actress than in an actor. It shows that our result is in line with the literature on in-group advantage ([GBBM07, EA02, KST+15]), and the fact that women have an advantage in recognizing emotions ([AMRM19]). It also indicates that female representativeness (even in this case, we are only talking about female animation) could be important for simulated human behavior in VHs of various media since we observed a higher value of corrected emotion recognition with actress motion. However, this statistical comparison was not significant, which requires further investigation in the future.

Regarding *RQ*3, we only obtained one significant result when the motion stimuli were based on styles, as shown in Tables 4.10 and 4.11, i.e., the participants perceived happiness better in the female model named Helena versus the female model named Miguel, and in the unnamed female model. As can be seen in Figures 4.11 and 4.12, the remaining results were very similar if we compared women x men and female with male models. Even if they are very similar in both charts, it is possible to observe a tendency to have more corrected answers when the VH is named Helena.

Regarding *RQ*4 (see Table 4.11), we only found results when analyzing the male model, where people had higher score values of recognition of the actor's happiness than the actress's. In these cases, we can say that the visual model of the VH influenced the recognition of happiness. When we look only at motion styles, in most cases for the participants, the correct perception of female or male happiness style was higher than the perception of actress or actor happiness. We hypothesize that stimuli related to facial styles can be more explicit of a specific gender in showing emotions than individual actors due to the following facts: *i)* The facial styles were built based on the average of a group of facial expressions and then exaggerated, while the actor/actress motions were a real person smiling; and *ii)* as the facial style is built using pictures we generated only the last facial posture and interpolated from the neutral one, while the actor/actress motions were transferred to the VHs. This is interesting to the CG media industry because it shows the possibility of using pre-defined facial styles instead of having actors generate facial motions.

We found significant differences related to *RQ*5, i.e., the nonbinary VH was more likely to have happiness perceived than the other two binary models (see Table 4.12). We hypothesize that the participants had an out-group advantage, i.e., our participants, who self-identified as binary, had their perception impacted when comparing binary and nonbinary VHs models and focused more on the expressed emotion of the nonbinary model. Just as we discussed in Section 4.4.1, according to Draude, a machine is born essentially genderless, as it does not have visual and behavioral characteristics that place it on a binary gender spectrum. Returning to our results, our hypothesis that there was an out-group effect may have shown that a genderless machine (VH in our case) deconstructed the participants' gender bias in recognizing the emotion happiness. In other words, when the VH was a binary model (female or male), gender bias may have interfered with the recognition of happiness. When VH was nonbinary (genderless), people may have focused more on recognizing happiness. In another hypothesis, along the lines of UV ([MMK12, HM10]), the perceived realism/human likeness of the nonbinary VH was greater than the perceived realism of the other two VHs, because these other two had female and male modifications based on the nonbinary VH observed by binary identified participants. This is a question for further analysis in the future (including other emotions and levels of realism/human likeness).

By analyzing the outcomes of Gender Attribution for the three VHs in the context of *RQ*6 and *RQ*7, our results suggest that participants experienced some difficulty in accurately attributing gender to the 3D model. We can see in Figure 4.14

that people provided more accurate answers when the VH was named Helena and Miguel than the unnamed VH. In addition, participants attributed genders less accurately to female and male VHs than to nonbinary, also when considering that the VHs were named. Our findings align with the literature ([MJH+07, MJH+09, NVDSN09, NY20, GFGW+23]), particularly in the context of VHs exhibiting clear male or female features. However, it is essential to note that our study introduces a novel perspective by examining the impact of a nonbinary VH model. Based on our observations, we propose a new hypothesis: when the VH model possesses distinct female or male features, participants tend to prioritize the model's appearance over the assigned name in gender attribution. In contrast, when presented with the nonbinary VH model, participants appear to focus more on the given name, resulting in more accurate responses. Furthermore, our results regarding the nonbinary adult VH on gender attribution were similar to the results related to the VBs experiments (both, in video and interactive stimuli), which the VBs were also considered genderless. These results suggest that VH's visual characteristics influence participants' attention and subsequent gender attribution. One of our results that illustrate this was noted when the female and male VH models presented one of the lowest gender attribution scores, and to most cases in which VH did not have a name. Therefore, visual characteristics with more feminine features combined with a masculine name seem to confuse gender attribution. So, in accordance with Armando et al. [AOR22], we hypothesize that the attribution of gender through the possibility of presenting a character with a nonbinary virtual model may be a possible solution to combat gender stereotypes. In addition, these findings provide valuable insights into the interplay between visual attributes, assigned names, and gender attribution in the context of VHs, extending the existing body of research on this topic both for psychological and CG literature.

5. FINAL REMARKS

This work proposes a set of experiments to evaluate how people perceive VHs in different contexts that involve gender and skin color biases. The main goal of this work was to conduct studies which assists in the design of VHs that consider gender and skin color diversities, while also taking human perception into account. The work presented studies on the influence of technological evolution and its importance on gender and skin color diversities, on the influence of demographic characteristics, and on in- and out-group influences in terms of gender and skin color of VHs. Below we present some conclusions that we have already obtained from our studies:

- Regarding the question "Are people feeling more comfortable with newer CG characters than older ones?", we can say that the percentage of perceived comfort has increased during those years and that the percentage of comfort about old characters remained similar to 2012 (year of the paper that was reproduced and compared) [FdMM+12]. In relation to gender, both male and female characters also conveyed more comfort when they were created using new technologies. However, the difference in perceived comfort between years was greater for female characters. In the case of skin color, both characters with white and dark colored skins conveyed more comfort with newer technologies than older ones. As we only had one character with dark colored skin created with old technologies, we were unable to conclude whether newer technologies provided more comfort than older ones. However, we found an indication that people felt more comfortable and perceived less realism in characters with dark colored skin than characters with white colored skin;

- In relation to "Do people perceive comfort and realism in female and male CG characters in the same way?" and "Does this change when participants are separated into women and men?", we can say that there is a strong relation between perceived comfort and gender of participants and characters. Men perceived very similar comfort of realistic male and female characters, while women felt more comfortable with realistic female characters. For men, the most important in relation to perceived comfort was the comparison between unrealistic and very realistic characters versus moderately realistic, in which men felt less comfortable;

- In the question "Are comfort and realism similarly perceived in CG characters with different skin colors?", the results were similar to those found in the study on differences between old and new technologies. In other words, people felt more comfortable and perceived less realism in characters with dark colored skin than characters with white colored skin.

- About the questions "How do people perceive genderless VHs?", "Does this change with respect to participant gender?", and "Do we have bias regarding gender attribution?", women perceived more emotions in a VB defined as female than men, and men perceived more emotions in male and unnamed VBs than women. In this case, the VB was always the same, that is, just changing the name. Furthermore, the group of participants who received the unnamed VB attributed a male gender to it, again, even though the VB did not have gender stereotypes;

- In relation to "Would the gender-based preference also emerge when interacting with a genderless VB?", the results showed that there were biased behaviors when participants chose toys (however, this indication of results requires more data and future analysis). Women and men tended to choose toys based on VB gender, with women choosing more toys for female VB, and men for male VB. In other results, as in the video stimuli, we can see that participants in the 3D interactive environment tended to correctly attribute gender to VB with male gender previously assigned through text. However, contrary to video stimuli, they tended to attribute gender less correctly to a VB with female name.In terms of perception of emotions in the interactive VB, the results showed similar trends to video stimuli when male participants perceived more emotions in the interactive male VB. On the other hand, the same did not happen when female participants interacted with VB with a female name;

- Finally, with regard to question "Would the gender-based preference also emerge when observe a genderless virtual adult instead of a virtual baby?", the results showed that participants tend to attribute genders more accurately when the textual and visual features are coherent and aligned, i.e., the shape face more feminine with a female name. This result is in line with the results of VBs, both from video and interactive stimuli.

Answering these questions, we can list three main findings of this work in relation to VHs, as shown below:

- The gender binary (female and male) expressed in VHs can influence people's perception in several aspects. Mainly, in terms of perception/recognition of emotions and perceptions of comfort and realism, in-group effects are highlighted when participants respond about VHs that belong to their groups. In this case, we can conclude that modeling VHs (in games, movies, etc) based on a binary gender can bring representation to the group to which that gender belongs, but can cause disadvantages when the audience needs to recognize/perceive emotions of the group of the opposite gender. Examples: *i)* in Section 3.2, we showed that women felt more comfortable with very realistic female VHs than with male ones; *ii)* in Section 4.1, when the VB had a female name in the video stimuli, women perceived more emotions in this VB than in the VB with a male name. Men, on the other hand, perceived more emotions in VB with a male name; Section 4.3, in relation to emotion recognition, we observed an indication of in-group effect when women had the advantage of perceiving happiness in the actress compared to the actor, and it shows that our results are in line with the literature on in-group advantage ([GBBM07, EA02, KST+15]).

- Regarding gender attribution (Chapter 4), we can conclude that the industry and academic community may not need to create VHs with gender stereotypes to inform the audience of the VH's assigned gender. Our results showed that people attributed a gender according to the previously assigned gender through text (both VB and adult VH). In those cases, the VHs (adult and baby) were considered genderless in our methodology. In other words, a female character in a game does not need to show stereotypical gender traits to show the audience that it is a female character. In this case, only a prior narrative informing the character's gender would be necessary for the audience to understand that the character was female. As mentioned by Draude [Dra11], a machine is born essentially genderless; that is, people attribute gender to it. Along the same lines, as mentioned in Chapter 4, Armando et al. [AOR22] raised the question "Can Androgynous Virtual Agents be a Potential Solution to Combat Gender Stereotypes?". So, in the case of this present work, our results indicate that genderless machines/androgynous virtual agents/genderless VHs could be solutions to combat gender stereotypes.

- Regarding skin color (Section 3.3), we were unable to measure in-group and out-group effects due to the lack of diversity of participants with different skin colors. However, in the context of our participants, when evaluating virtual characters through a UV theory perspective, our results showed that characters with dark colored skin conveyed more comfort and were perceived as less realistic than characters with white skin colors. As discussed in Section 3.4.3, algorithms designed to create dark skin colors are often adapted from those initially intended for generating white skin colors [KRD+22], potentially leading to unreliable results for characters with darker skin tones. As discussed in UV literature [MMK12, MGHK09, KFMT15, MC16], the audience tends to feel comfortable with unrealistic CG characters, e.g., cartoons. So, maybe linking our results of perceived realism and strangeness with the results of this present work, the low level of realism perceived in relation to dark colored skin characters made these characters convey more comfort. This hypothesis still needs to be tested in future work for at least one important reason: the lack of diversity of our participants with different skin colors. We can conclude that stimuli involving VHs with dark colored skin need to increase the level of realism, may be creating new algorithms to render the dark colored skin, or varying characteristics in the environment that cause the perceived realism to be increased, such as the study by Khoo et al. [KLV23], which manipulated lighting elements for unrealistic VHs with different skin colors.

Of course, this work had limitations, such as those mentioned in the previous paragraphs in relation to the lack of diversity in the skin color of the participants. Furthermore, we can list some other limitations such as the not so controlled stimuli in Chapter 3, which contained VHs from different media. In Chapter 4, in Section 4.3, some limitations were related, such as not using VHs with dark colored skin, the non-diversity of nonbinary genders of the participants, among other things that could perhaps change some results.

However, we specify that the results found in this work are important for contexts in relation to discussions about gender and skin color in our society. Furthermore, this work leaves open hypotheses for both gender and skin color analyzes of VHs. For example, use of nonbinary names for genderless VHs in possible gender attribution analyses, use of VR in interaction with genderless VHs, skin color bias in the analysis of a continuum between different skin colors using the same VH model, etc.

For future work, our intention is to continue this research trajectory, investigating how individuals perceive VHs and fostering discussions aimed at creating more inclusive virtual environments. By leveraging advanced rendering and animation algorithms, we aim to enhance the realism of VH visualizations, thereby facilitating more authentic interactions and experiences within virtual spaces. Other potential avenues for the future are: *i)* Cross-cultural Studies - Explore how perceptions of VHs vary across different cultural contexts; *ii)* Variations between Gender and Skin Color Differences: Investigating how variations in skin color and gender across a continuum of the same VH model can impact user experience and interaction; *iii)* Emotional Expressiveness: Explore how subtle differences in facial expressions, body language, and vocal intonation influence individuals' emotional responses to VHs with different genders, skin colors, etc; *iv* Ethical Considerations: Investigate ethical considerations related to the design and use of VHs. Explore topics such as privacy, consent, identity, and representation, and examine how ethical considerations influence perceptions of VH technology; *v)* Long-term Effects: Explore the long-term effects of interacting with VHs on individuals' attitudes, behaviors, and perceptions. Investigate whether repeated exposure to VHs leads to desensitization, habituation, or changes in social cognition (including gender diversity, skin color, etc).

These future research directions can help advance our understanding of how individuals perceive and interact with VHs, informing the design of more engaging, immersive, and socially impactful virtual environments.

References

[ADF+21] Araujo, V.; Dalmoro, B.; Favaretto, R.; Vilanova, F.; Costa, A.; Musse, S. R. “How much do we perceive geometric features, personalities and emotions in avatars?” In: Proceedings of the 38th Computer Graphics International Conference (CGI), 2021, pp. 548–567. (Citado na página 164.)

[ADG+22] Araujo, V.; Dalmoro, B.; Geiss, R. C.; Pinho, M. S.; Musse, S. “How do we perceive characters? an analysis of human perception in still images, animations and vr scenarios”. In: Proceedings of the 21th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2022, pp. 295–300. (Citado na página 165.)

[ADM21] Araujo, V.; Dalmoro, B.; Musse, S. R. “Analysis of charisma, comfort and realism in cg characters from a gender perspective”, *The Visual Computer*, vol. 37, Jan-Feb 2021, pp. 1–14. (Citado nas páginas 43 and 164.)

[AMDM21] Araujo, V.; Melgare, J.; Dalmoro, B.; Musse, S. R. “Is the perceived comfort with cg characters increasing with their novelty”, *IEEE Computer Graphics and Applications*, vol. 42, Sep 2021, pp. 32–46. (Citado nas páginas 16, 80, and 163.)

[AMRM19] Abbruzzese, L.; Magnani, N.; Robertson, I. H.; Mancuso, M. “Age and gender differences in emotion recognition”, *Frontiers in Psychology*, vol. 10, Oct 2019, pp. 2371. (Citado nas páginas 30, 42, 44, 89, and 138.)

[AOR22] Armando, M.; Ochs, M.; Régner, I. “The impact of pedagogical agents’ gender on academic learning: A systematic review”, *Frontiers in Artificial Intelligence*, vol. 5, Jun 2022, pp. 1–23. (Citado nas páginas 89, 97, 111, 134, 136, 140, and 143.)

[ASCM22] Araujo, V.; Schaffer, D.; Costa, A. B.; Musse, S. R. “Towards virtual humans without gender stereotyped visual features”. In: Proceedings of the ACM SIGGRAPH Asia, In Technical Communications Section, 2022, pp. 1–4. (Citado na página 165.)

[AWdS+21] Andreotti, L.; Weber, M. L.; da Silva, T. L.; de Andrade Araujo, V. F.; Musse, S. R. “Perception of charisma, comfort, micro and macro expressions in computer graphics characters”. In: Proceedings of the 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2021, pp. 107–116. (Citado nas páginas 42 and 164.)

[Bas78] Bassili, J. N. “Facial motion in the perception of faces and of emotional expression”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4–3, Aug 1978, pp. 373. (Citado na página 42.)

[BB17] Bailey, J. D.; Blackmore, K. L. “Gender and the perception of emotions in avatars”. In: Proceedings of the Australasian Computer Science Week Multiconference, 2017, pp. 1–8. (Citado nas páginas 43 and 111.)

[BBB+19] Bedder, R. L.; Bush, D.; Banakou, D.; Peck, T.; Slater, M.; Burgess, N. “A mechanistic account of bodily resonance and implicit bias”, *Cognition*, vol. 184, Mar 2019, pp. 1–10. (Citado na página 41.)

[BBN+20] Banakou, D.; Beacco, A.; Neyret, S.; Blasco-Oliver, M.; Seinfeld, S.; Slater, M. “Virtual body ownership and its consequences for implicit racial bias are dependent on social context”, *Royal Society*, vol. 7–12, Dec 2020, pp. 201848. (Citado na página 41.)

[BH05] Beaupré, M. G.; Hess, U. “Cross-cultural emotion recognition among canadian ethnic groups”, *Journal of Cross-Cultural Psychology*, vol. 36–3, May 2005, pp. 355–370. (Citado na página 39.)

[BHS16] Banakou, D.; Hanumanthu, P. D.; Slater, M. “Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias”, *Frontiers in Human Neuroscience*, vol. 10–1, Nov 2016, pp. 601–601. (Citado na página 41.)

[BL16] Bailey, A. H.; LaFrance, M. “Anonymously male: Social media avatar icons are implicitly male and resistant to change”, *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 10–4, Jan-Dec 2016, pp. 1–14. (Citado na página 136.)

[BN10] Balas, B.; Nelson, C. A. “The role of face shape and pigmentation in other-race face perception: An electrophysiological study”, *Neuropsychologia*, vol. 48–2, Jan 2010, pp. 498–506. (Citado na página 40.)

[Bro20] Brown, R. “The social identity approach: Appraising the tajfellian legacy”, *British Journal of Social Psychology*, vol. 59–1, Nov 2020, pp. 5–25. (Citado nas páginas 29, 44, and 84.)

[CC76] Condry, J.; Condry, S. “Sex differences: A study of the eye of the beholder”, *Child Development*, vol. 4, Sep 1976, pp. 812–819. (Citado nas páginas 32, 45, 47, 89, 91, 93, 96, 135, and 136.)

[CCH83] Culp, R. E.; Cook, A. S.; Housley, P. C. “A comparison of observed and reported adult-infant interactions: Effects of perceived sex”, *Sex Roles*, vol. 9, Apr 1983, pp. 475–479. (Citado nas páginas 33, 45, 47, and 137.)

[CHK07] Chaminade, T.; Hodgins, J.; Kawato, M. “Anthropomorphism influences perception of computer-animated characters’ actions”, *Social Cognitive and Affective Neuroscience*, vol. 2–3, Sep 2007, pp. 206–216. (Citado na página 38.)

[CPJ+13] Cheetham, M.; Pavlovic, I.; Jordan, N.; Suter, P.; Jancke, L. “Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data”, *Frontiers in Psychology*, vol. 4, Mar 2013, pp. 108. (Citado na página 31.)

[dAACM23] de Andrade Araujo, V. F.; Costa, A. B.; Musse, S. R. “Evaluating the uncanny valley effect in dark colored skin virtual humans”. In: Proceedings of the 36th Conference on Graphics, Patterns and Images (SIBGRAPI), 2023, pp. 1–6. (Citado na página 166.)

[dAASCM21] de Andrade Araujo, V. F.; Schaffer, D. H. M.; Costa, A. B.; Musse, S. R. “Perception of computer graphics characters in groups with skin color diversity”. In: Proceedings of the 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2021, pp. 1069–1072. (Citado na página 165.)

[DFH+12] Dill, V.; Flach, L. M.; Hocevar, R.; Lykawka, C.; Musse, S. R.; Pinho, M. S. “Evaluation of the uncanny valley in cg characters”.

In: Proceedings of the 12th International Conference on Intelligent Virtual Agents, 2012, pp. 511–513. (Citado nas páginas 32, 39, 49, 75, and 78.)

[DK22] Durupinar, F.; Kim, J. “Facial emotion recognition of virtual humans with different genders, races, and ages”. In: Proceedings of the ACM Symposium on Applied Perception, 2022, pp. 1–10. (Citado na página 43.)

[DMND+21] Dal Molin, G. P.; Nomura, F. M.; Dalmoro, B. M.; Victor, F. d. A.; Musse, S. R. “Can we estimate the perceived comfort of virtual human faces using visual cues?” In: Proceedings of the 15th IEEE International Conference on Semantic Computing (ICSC), 2021, pp. 366–369. (Citado na página 163.)

[Dra11] Draude, C. “Intermediaries: reflections on virtual humans, gender, and the uncanny valley”, *AI & Society*, vol. 26–4, Jan 2011, pp. 319–327. (Citado nas páginas 30, 44, 73, 84, 87, 88, 89, 111, 133, and 143.)

[DSKP22] Dodik, A.; Sellán, S.; Kim, T.; Phillips, A. “Sex and gender in the computer graphics research literature”. In: Proceedings of the ACM SIGGRAPH, In the Talks Section, 2022, pp. 1–2. (Citado na página 89.)

[EA02] Elfenbein, H. A.; Ambady, N. “Is there an in-group advantage in emotion recognition?”, *Psychological Bulletin*, vol. 128, Mar 2002, pp. 243—249. (Citado nas páginas 29, 44, 84, 89, 134, 138, and 143.)

[EA03a] Elfenbein, H. A.; Ambady, N. “Universals and cultural differences in recognizing emotions”, *Current Directions in Psychological Science*, vol. 12–5, Oct 2003, pp. 159–164. (Citado na página 41.)

[EA03b] Elfenbein, H. A.; Ambady, N. “When familiarity breeds accuracy: cultural exposure and facial emotion recognition”, *Journal of Personality and Social Psychology*, vol. 85–2, Aug 2003, pp. 276. (Citado na página 41.)

[EF78] Ekman, P.; Friesen, W. V. “Facial action coding system (facs): a technique for the measurement of facial action”, *Environmental*

Psychology and Nonverbal Behavior, vol. 5–3, Jan-Dec 1978, pp. 56–75. (Citado nas páginas 42 and 122.)

[EHEM13] Ennis, C.; Hoyet, L.; Egges, A.; McDonnell, R. “Emotion capture: Emotionally expressive characters for games”. In: Proceedings of the 6th Motion on Games Conference, ACM (Editor), 2013, pp. 53–60. (Citado nas páginas 42, 111, and 138.)

[Ekm92] Ekman, P. “An argument for basic emotions”, *Cognition and Emotion*, vol. 6, May 1992, pp. 169–200. (Citado nas páginas 42 and 116.)

[FCW21] Fang, Z.; Cai, L.; Wang, G. “Metahuman creator the starting point of the metaverse”. In: Proceedings of the International Symposium on Computer Technology and Information Science (ISCTIS), 2021, pp. 154–157. (Citado na página 29.)

[FdMM+12] Flach, L. M.; de Moura, R. H.; Musse, S. R.; Dill, V.; Pinho, M. S.; Lykawka, C. “Evaluation of the uncanny valley in cg characters”. In: Proceedings of the 11th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2012, pp. 108–116. (Citado nas páginas 15, 32, 39, 47, 49, 50, 51, 52, 53, 58, 59, 62, 75, 78, and 141.)

[GAHH19] Garza, M.; Akleman, E.; Harris, S.; House, F. “Emotional silence: Are emotive expressions of 3d animated female characters designed to fit stereotypes”, *Women’s Studies International Forum*, vol. 76, Sep-Oct 2019, pp. 102252. (Citado nas páginas 88 and 134.)

[GBBM07] Guadagno, R. E.; Blascovich, J.; Bailenson, J. N.; McCall, C. “Virtual humans and persuasion: The effects of agency and behavioral realism”, *Media Psychology*, vol. 10–1, Dec 2007, pp. 1–22. (Citado nas páginas 44, 84, 89, 134, 135, 138, and 143.)

[GFGW+23] Ghosh, R.; Feijoo-Garcia, P. G.; Wrenn, C.; Stuart, J.; Lok, B. “Evaluating face gender cues in virtual humans within and beyond the gender binary”, *Frontiers in Virtual Reality*, vol. 4, Aug 2023, pp. 1251420. (Citado nas páginas 43, 44, 89, 97, 111, 114, 121, 138, and 140.)

[Gom19] Gomes, L. F. E. “Ser pardo: o limbo identitário-racial brasileiro e a reivindicação da identidade”, *Cadernos de Gênero e Diversidade*, vol. 5–1, May 2019, pp. 66–78. (Citado na página 79.)

[Gre89] Greeno, C. G. “Gender differences in children’s proximity to adults”, Tese de doutorado, Departament of Psychology of Stanford University, 1989, 210p. (Citado nas páginas 33, 46, 101, and 137.)

[HAJK04] Hess, U.; Adams Jr, R. B.; Kleck, R. E. “Facial appearance, gender, and emotion expression”, *Emotion*, vol. 4, Dec 2004, pp. 378. (Citado nas páginas 44, 89, 114, 121, and 138.)

[Hal66] Hall, E. T. “The hidden dimension”. Anchor, 1966, vol. 609, 246p. (Citado nas páginas 20, 101, 108, and 109.)

[HCKH16] Hyde, J.; Carter, E. J.; Kiesler, S.; Hodgins, J. K. “Evaluating animated characters: Facial motion magnitude influences personality perceptions”, *ACM Transactions on Applied Perception (TAP)*, vol. 13–2, Feb 2016, pp. 8. (Citado na página 39.)

[HHM10] Hall, J. K.; Hutton, S. B.; Morgan, M. J. “Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition?”, *Cognition & Emotion*, vol. 24–4, May 2010, pp. 629–637. (Citado na página 31.)

[HJO+10] Hodgins, J.; Jörg, S.; O’Sullivan, C.; Park, S. I.; Mahler, M. “The saliency of anomalies in animated human characters”, *ACM Transactions on Applied Perception (TAP)*, vol. 7–4, Jul 2010, pp. 22. (Citado na página 39.)

[HM10] Ho, C.-C.; MacDorman, K. F. “Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices”, *Computers in Human Behavior*, vol. 26–6, Nov 2010, pp. 1508–1518. (Citado na página 139.)

[HMA+17] Hsieh, R.; Mochizuki, Y.; Asano, T.; Higashida, M.; Shirai, A. “"real baby-real family" vr entertainment baby interaction system”. In: Proceedings of the ACM SIGGRAPH, In the Emerging Technologies Section, ACM (Editor), 2017, pp. 1–2. (Citado na página 46.)

[HMG10] Hehman, E.; Mania, E. W.; Gaertner, S. L. “Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect”, *Journal of Experimental Social Psychology*, vol. 46–2, Mar 2010, pp. 445–448. (Citado na página 41.)

[HSS17] Hasler, B. S.; Spanlang, B.; Slater, M. “Virtual race transformation reverses racial in-group bias”, *PloS One*, vol. 12–4, Apr 2017, pp. 1–20. (Citado nas páginas 30 and 41.)

[Joh73] Johansson, G. “Visual perception of biological motion and a model for its analysis”, *Perception & Psychophysics*, vol. 14–2, Jun 1973, pp. 201–211. (Citado na página 42.)

[KABM20] Kolling, C.; Araujo, V.; Barros, R. C.; Musse, S. R. “How does computer animation affect our perception of emotions in video summarization?” In: Proceedings of the International Symposium on Visual Computing, 2020, pp. 374–385. (Citado na página 163.)

[KAVM23] Kolling, C.; Araujo, V.; Veloso, A.; Musse, S. R. “Mitigating bias in facial analysis systems by incorporating label diversity”, *Computers & Graphics*, vol. 116, Nov 2023, pp. 173–184. (Citado na página 165.)

[KFMT15] Kätsyri, J.; Förger, K.; Mäkäräinen, M.; Takala, T. “A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness”, *Frontiers in psychology*, vol. 6, Apr 2015, pp. 390. (Citado nas páginas 29, 38, 51, 54, 74, 77, 86, 111, and 144.)

[KL13] Kang, S.-M.; Lau, A. S. “Revisiting the out-group advantage in emotion recognition in a multicultural society: Further evidence for the in-group advantage”, *Emotion*, vol. 13–2, Apr 2013, pp. 203. (Citado na página 40.)

[KL22] Kteily, N. S.; Landry, A. P. “Dehumanization: trends, insights, and challenges”, *Trends in Cognitive Sciences*, vol. 26–3, Jan 2022, pp. 222–240. (Citado na página 86.)

[KLV23] Khoo, M.; Lipson, D.; Velasquez, J. “Lighting and look dev for skin tones in disney’s" strange world"”. In: Proceedings of the Digital Production Symposium, 2023, pp. 1–5. (Citado na página 144.)

[KMJ+24] Kao, D.; Mubarrat, S. T.; Joshi, A.; Pandita, S.; Mousas, C.; Liang, H.-N.; Ratan, R. “Exploring how gender-anonymous voice avatars influence women’s performance in online computing group work”, *International Journal of Human-Computer Studies*, vol. 181, Jan 2024, pp. 103146. (Citado nas páginas 44 and 136.)

[KMT17] Kätsyri, J.; Mäkäräinen, M.; Takala, T. “Testing the ’uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli”, *International Journal of Human-Computer Studies*, vol. 97, Jan 2017, pp. 149–161. (Citado nas páginas 32 and 38.)

[KRD+22] Kim, T.; Rushmeier, H.; Dorsey, J.; Nowrouzezahrai, D.; Syed, R.; Jarosz, W.; Darke, A. “Countering racial bias in computer graphics research”. In: Proceedings of the ACM SIGGRAPH, In the Talks Section, 2022, pp. 1–2. (Citado nas páginas 31, 85, 86, 89, and 144.)

[KST+15] Krumhuber, E. G.; Swiderska, A.; Tsankova, E.; Kamble, S. V.; Kappas, A. “Real or artificial? intergroup biases in mind perception in a cross-cultural perspective”, *PloS One*, vol. 10–9, Sep 2015, pp. 1–14. (Citado nas páginas 29, 42, 44, 84, 89, 134, 138, and 143.)

[KTS+04] Kohler, C. G.; Turner, T.; Stolar, N. M.; Bilker, W. B.; Brensinger, C. M.; Gur, R. E.; Gur, R. C. “Differences in facial expressions of four universal emotions”, *Psychiatry Research*, vol. 128–3, Oct 2004, pp. 235–244. (Citado na página 115.)

[Lau17] Laue, C. “Familiar and strange: Gender, sex, and love in the uncanny valley”, *Multimodal Technologies and Interaction*, vol. 1–1, Jan 2017, pp. 2. (Citado na página 31.)

[LNB14] Lee, J.-E. R.; Nass, C. I.; Bailenson, J. N. “Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups”, *Cyberpsychology, Behavior, and Social Networking*, vol. 17–4, Apr 2014, pp. 248–254. (Citado na página 136.)

[MAM22] MOLIN, G.; ARAUJO, V.; Musse, S. R. “Estimating perceived comfort in virtual humans based on spatial and spectral entropy”. In: Proceedings of the 17th International Joint Conference on Computer

Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), 2022. (Citado na página 165.)

[MBB12] McDonnell, R.; Breidt, M.; Bülthoff, H. H. "Render me real? investigating the effect of render style on the perception of animated virtual humans", *ACM Transactions on Graphics (TOG)*, vol. 31–4, Jul 2012, pp. 1–11. (Citado na página 43.)

[MC16] MacDorman, K. F.; Chattopadhyay, D. "Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not", *Cognition*, vol. 146, Jan 2016, pp. 190–205. (Citado nas páginas 31, 38, and 144.)

[MFdAAV+21] Migon Favaretto, R.; de Andrade Araujo, V. F.; Vilanova, F.; Brandelli Costa, A.; Raupp Musse, S. "Cultural behaviors analysis in video sequences", *Machine Vision and Applications*, vol. 32–4, Jul 2021, pp. 1–24. (Citado na página 163.)

[MGCPV07] Morrison, E. R.; Gralewski, L.; Campbell, N.; Penton-Voak, I. S. "Facial movement varies by sex and is related to attractiveness", *Evolution and Human Behavior*, vol. 28–3, May 2007, pp. 186–192. (Citado na página 42.)

[MGHK09] MacDorman, K. F.; Green, R. D.; Ho, C.-C.; Koch, C. T. "Too real for comfort? uncanny responses to computer generated faces", *Computers in Human Behavior*, vol. 25–3, May 2009, pp. 695–710. (Citado nas páginas 31, 51, and 144.)

[MHA+17] Mochizuki, Y.; Hsieh, R.; Agatsuma, D.; Asano, T.; Higashida, M.; Nishikizawa, T.; Shirai, A. "Real baby-real family: Holdable tangible baby vr". In: Proceedings of the Virtual Reality International Conference-Laval Virtual, 2017, pp. 1–4. (Citado na página 46.)

[MJH+07] McDonnell, R.; Jörg, S.; Hodgins, J. K.; Newell, F.; O'Sullivan, C. "Virtual shapers & movers: form and motion affect sex perception". In: Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization, 2007, pp. 7–10. (Citado nas páginas 43, 44, 89, 111, and 140.)

[MJH+09] McDonnell, R.; Jörg, S.; Hodgins, J. K.; Newell, F.; O'sullivan, C. "Evaluating the effect of motion and body shape on the perceived sex of virtual characters", *ACM Transactions on Applied Perception (TAP)*, vol. 5–4, Feb 2009, pp. 1–14. (Citado nas páginas 44, 89, 111, and 140.)

[MKKL17] McDuff, D.; Kodra, E.; Kaliouby, R. e.; LaFrance, M. "A large-scale analysis of sex differences in facial expressions", *PloS One*, vol. 12–4, Apr 2017, pp. 1–11. (Citado nas páginas 44 and 138.)

[MM16] Mustafa, M.; Magnor, M. "Eeg based analysis of the perception of computer-generated faces". In: Proceedings of the 13th European Conference on Visual Media Production (CVMP), 2016, pp. 4. (Citado na página 29.)

[MMA+23] Musse, S. R.; Molin, G. P. D.; Araujo, V. F. d. A.; Schaffer, D. H. M.; Brandelli, A. C. "Perceptual analysis of computer graphics characters in digital entertainment". In: *Grand Research Challenges in Games and Entertainment Computing in Brazil - GranDGamesBR 2020–2030*, Santos, R. P. d.; Hounsell, M. d. S. (Editors), Springer, 2023, chap. 10, pp. 207–232. (Citado na página 166.)

[MMK12] Mori, M.; MacDorman, K. F.; Kageki, N. "The uncanny valley [from the field]", *IEEE Robotics & Automation Magazine*, vol. 19–2, Jun 2012, pp. 98–100. (Citado nas páginas 29, 38, 60, 61, 75, 83, 85, 111, 135, 139, and 144.)

[MMLD20] Morales-Martinez, G.; Latreille, P.; Denny, P. "Nationality and gender biases in multicultural online learning environments: The effects of anonymity". In: Proceedings of the Conference on Human Factors in Computing Systems (CHI), 2020, pp. 1–14. (Citado nas páginas 44 and 136.)

[MPDdAA21] Musse, S. R.; Pinho, G.; Dalmoro, B.; de Andrade Araujo, V. F. "Grandgamesbr: Perceptual analysis of computer graphics characters in digital entertainment". In: Proceedings of the 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2021, pp. 1029–1032. (Citado na página 164.)

[MQM22] Melgaré, J. K.; Queiroz, R. B.; Musse, S. R. "Investigating emotion style in human faces using clustering methods", *IEEE Computer*

Graphics and Applications, vol. 42–4, Nov 2022, pp. 52–71. (Citado nas páginas 120 and 122.)

[MRG+23] Montanha, R.; Raupp, G.; Gonzalez, V.; Partichelli, Y.; Bins, A.; Ferreira, M.; Araujo, V.; Musse, S. “Revisiting micro and macro expressions in computer graphics characters”. In: Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment (SBGames), 2023, pp. 38–45. (Citado na página 166.)

[MZCD21] McDonnell, R.; Zibrek, K.; Carrigan, E.; Dahyot, R. “Model for predicting perception of facial action unit activation using virtual humans”, *Computers & Graphics*, vol. 100–1, Nov 2021, pp. 81–92. (Citado na página 42.)

[NSE+21] Nardino, L.; Schaffer, D.; Elsner, F.; Krzmienski, E.; Araujo, V. F. D. A.; Silva, G. F.; Cassol, V. J.; Favaretto, R. M.; Musse, S. R. “Perception of personality traits in crowds of virtual humans”. In: Proceedings of the 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), 2021, pp. 117–125. (Citado na página 164.)

[NVDSN09] Niculescu, A.; Van Der Sluis, F.; Nijholt, A. “Feminity, masculinity and androgyny: how humans perceive the gender of anthropomorphic agents”. In: Proceedings of 13th International Conference on Human-Computer Interaction, Springer Verlag, Heidelberg, 2009, pp. 628–632. (Citado nas páginas 111 and 140.)

[NY20] Nag, P.; Yalçın, Ö. N. “Gender stereotypes in virtual agents”. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, 2020, pp. 1–8. (Citado nas páginas 30, 44, 111, and 140.)

[OAL+23] Obremski, D.; Akuffo, O. B.; Lücke, L.; Semineth, M.; Tomiczek, S.; Weichert, H.-F.; Lugrin, B. “Reducing racial bias by interacting with virtual agents: An intervention in virtual reality”. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, 2023, pp. 747–755. (Citado na página 42.)

[PCC+10] Petrásová, A.; Czanner, S.; Chalmers, A.; Farrer, J. V.; Wolke, D. “The playability evaluation of virtual baby feeding application”. In:

Proceedings of the 2nd International Conference on Games and Virtual Worlds for Serious Applications, 2010, pp. 95–100. (Citado na página 46.)

[PCH+10] Petrasova, A.; Czanner, G.; Happa, J.; Czanner, S.; Wolke, D.; Chalmers, A. "Assessing a virtual baby feeding training system". In: Proceedings of the 7th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa, 2010, pp. 37–44. (Citado na página 46.)

[RPM17] Ruhland, K.; Prasad, M.; McDonnell, R. "Data-driven approach to synthesizing facial animation using motion capture", *IEEE Computer Graphics and Applications*, vol. 37–4, Aug 2017, pp. 30–41. (Citado na página 39.)

[SCD+22] Simeone, A. L.; Cools, R.; Depuydt, S.; Gomes, J. M.; Goris, P.; Grocott, J.; Esteves, A.; Gerling, K. "Immersive speculative enactments: bringing future scenarios and technology to life using virtual reality". In: Proceedings of the Conference on Human Factors in Computing Systems (CHI), 2022, pp. 1–20. (Citado na página 46.)

[Sch18] Schwind, V. "Implications of the uncanny valley of avatars and virtual characters for human-computer interaction", Tese de doutorado, Institut für Visualisierung und Interaktive Systeme der Universität Stuttgart, 2018, 304p. (Citado na página 39.)

[Sco07] Scott, J. W. "Gender as a useful category of historical analysis". Routledge, 2007, 21p. (Citado nas páginas 45 and 88.)

[SH77] Shively, M. G.; Hall, M. A. "Departures from sex role stereotypes of appearance and behavior and violations of civil liberties", *Journal of Homosexuality*, vol. 2–4, Oct 1977, pp. 331–335. (Citado na página 114.)

[SKZ75] Seavey, C. A.; Katz, P. A.; Zalk, S. R. "Baby x: The effect of gender labels on adult responses to infants", *Sex Roles*, vol. 1, Jun 1975, pp. 103–109. (Citado nas páginas 33, 46, 47, 96, 100, 101, and 137.)

[SL78] Smith, C.; Lloyd, B. "Maternal behavior and perceived sex of infant: Revisited", *Child Development*, vol. 49, Dec 1978, pp. 1263–1265. (Citado nas páginas 33, 45, and 96.)

[SL80] Sidorowicz, L. S.; Lunney, G. S. “Baby x revisited”, *Sex Roles*, vol. 6, Feb 1980, pp. 67–73. (Citado nas páginas 33, 46, 47, 96, 100, 101, and 137.)

[SN07] Seyama, J.; Nagayama, R. S. “The uncanny valley: Effect of realism on the impression of artificial human faces”, *Presence: Teleoperators and Virtual Environments*, vol. 16–4, Aug 2007, pp. 337–351. (Citado na página 43.)

[SOS+22] Saneyoshi, A.; Okubo, M.; Suzuki, H.; Oyama, T.; Laeng, B. “The other-race effect in the uncanny valley”, *International Journal of Human-Computer Studies*, vol. 166, Oct 2022, pp. 102871. (Citado nas páginas 30 and 41.)

[SRLZ14] Shahid, M.; Rossholm, A.; Lövström, B.; Zepernick, H.-J. “No-reference image and video quality assessment: a classification and review of recent approaches”, *EURASIP Journal on Image and Video Processing*, vol. 2014–1, Aug 2014, pp. 40. (Citado na página 30.)

[TGG21] Tang, K.; Gerling, K.; Geurts, L. “Virtual feed: A simulated breastfeeding experience in virtual reality”. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI), In the Extended Abstracts Section, 2021, pp. 1–4. (Citado na página 46.)

[TGG22] Tang, K.; Gerling, K.; Geurts, L. “Virtual feed: design and evaluation of a virtual reality simulation addressing the lived experience of breastfeeding”. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI), 2022, pp. 1–17. (Citado na página 46.)

[TGW11] Tinwell, A.; Grimshaw, M.; Williams, A. “The uncanny wall”, *International Journal of Arts and Technology*, vol. 4–3, 2011, pp. 326–341. (Citado na página 30.)

[TNC13] Tinwell, A.; Nabi, D. A.; Charlton, J. P. “Perception of psychopathy and the uncanny valley in virtual characters”, *Computers in Human Behavior*, vol. 29–4, Jul 2013, pp. 1617–1625. (Citado nas páginas 42, 43, and 84.)

[TT04] Tajfel, H.; Turner, J. C. “The social identity theory of intergroup behavior”. In: *Political Psychology: Key Readings*, Press, P. (Editor), Psychology Press, 2004, chap. 16, pp. 276–293. (Citado nas páginas 29 and 84.)

[Tur09] Turing, A. M. “Computing machinery and intelligence”. In: *Parsing the Turing Test*, Springer (Editor), Springer, 2009, chap. 3, pp. 23–65. (Citado nas páginas 87, 89, and 133.)

[Wal10] Wallach Scott, J. “Gender: still a useful category of analysis?”, *Diogenes*, vol. 57–1, Oct 2010, pp. 7–14. (Citado nas páginas 45 and 88.)

[WSD76] Will, J. A.; Self, P. A.; Datan, N. “Maternal behavior and perceived sex of infant”, *American Journal of Orthopsychiatry*, vol. 46–1, Jan 1976, pp. 135. (Citado nas páginas 33, 45, 47, 96, and 137.)

[WT03] Walker, P. M.; Tanaka, J. W. “An encoding advantage for own-race versus other-race faces”, *Perception*, vol. 32–9, Sep 2003, pp. 1117–1125. (Citado na página 40.)

[WZC+20] Wisessing, P.; Zibrek, K.; Cunningham, D. W.; Dingliana, J.; McDonnell, R. “Enlighten me: Importance of brightness and shadow for character emotion and appeal”, *ACM Transactions on Graphics (TOG)*, vol. 39–3, Apr 2020, pp. 1–12. (Citado na página 38.)

[ZAJ+15] Zell, E.; Aliaga, C.; Jarabo, A.; Zibrek, K.; Gutierrez, D.; McDonnell, R.; Botsch, M. “To stylize or not to stylize?: the effect of shape and material stylization on the perception of computer-generated faces”, *ACM Transactions on Graphics (TOG)*, vol. 34–6, Nov 2015, pp. 184. (Citado nas páginas 29 and 37.)

[ZHRM13] Zibrek, K.; Hoyet, L.; Ruhland, K.; McDonnell, R. “Evaluating the effect of emotion on gender recognition in virtual humans”. In: Proceedings of the ACM Symposium on Applied Perception, 2013, pp. 45–49. (Citado nas páginas 30, 43, 48, 97, 110, 111, and 138.)

[ZHRM15] Zibrek, K.; Hoyet, L.; Ruhland, K.; Mcdonnell, R. “Exploring the effect of motion type and emotions on the perception of gender in virtual humans”, *ACM Transactions on Applied Perception (TAP)*, vol. 12–3, Jul 2015, pp. 1–20. (Citado nas páginas 43, 48, 89, 110, 111, and 138.)

[ZMAV+23] Zhao, F.; Mayer, R. E.; Adamo-Villani, N.; Mousas, C.; Choi, M.; Lam, L.; Mukanova, M.; Hauser, K. “Recognizing and relating to the race/ethnicity and gender of animated pedagogical agents”, *Journal of Educational Computing Research*, vol. 0, Nov 2023, pp. 07356331231213932. (Citado nas páginas 42 and 44.)

[ZNO+20a] Zibrek, K.; Niay, B.; Olivier, A.-H.; Hoyet, L.; Pettre, J.; McDonnell, R. “The effect of gender and attractiveness of motion on proximity in virtual reality”, *ACM Transactions on Applied Perception (TAP)*, vol. 17–4, Nov 2020, pp. 1–15. (Citado nas páginas 43 and 137.)

[ZNO+20b] Zibrek, K.; Niay, B.; Olivier, A.-H.; Hoyet, L.; Pettre, J.; McDonnell, R. “Walk this way: Evaluating the effect of perceived gender and attractiveness of motion on proximity in virtual reality”. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2020, pp. 169–170. (Citado nas páginas 43, 97, and 137.)

[ZNO+22] Zibrek, K.; Niay, B.; Olivier, A.-H.; Pettré, J.; Hoyet, L.; Mcdonnell, R. “Proximity in vr: The importance of character attractiveness and participant gender”. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2022, pp. 672–673. (Citado nas páginas 43, 97, and 137.)

[ZZM19] Zell, E.; Zibrek, K.; McDonnell, R. “Perception of virtual characters”. In: Proceedings of the ACM SIGGRAPH, In the Courses Section, 2019, pp. 1–17. (Citado nas páginas 29 and 37.)

Appendix A – PUBLICATIONS

This appendix presents the relation of publications obtained during the development of this research. Section A.1 shows a list of already published re searches, including conference and journal papers. Section A.2 shows a list of submitted papers to conferences and journals without a decision by the moment of delivery of this manuscript.

A.1 Published Research

How Does Computer Animation Affect Our Perception of Emotions in Video Summarization? [KABM20]
*Camila Kolling, **Victor Araujo**, Rodrigo C. Barros, Soraia Raupp Musse*
Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15. Springer International Publishing, 2020. p. 374-385.
DOI: `https://doi.org/10.1007/978-3-030-64559-55_29`.

Can we estimate the perceived comfort of virtual human faces using visual cues? [DMND+21]
*Greice P. Dal Molin, Felipe M. Nomura, Bruna M. Dalmoro, **Victor Araujo**, Soraia R. Musse*
In: 2021 IEEE 15th International Conference on Semantic Computing (ICSC). IEEE, 2021. p. 366-369.
DOI: `10.1109/ICSC50631.2021.00085`

Cultural behaviors analysis in video sequences [MFdAAV+21]
*Rodolfo Migon Favaretto, **Victor Araujo**, Felipe Vilanova, Angelo Brandelli Costa, Soraia Raupp Musse*
Machine Vision and Applications, v. 32, p. 1-24, 2021.
DOI: `https://doi.org/10.1007/s00138-021-01225-2`

Is the Perceived Comfort With CG Characters Increasing With Their Novelty? [AMDM21]
***Victor Araujo**, Julia Melgare, Bruna Martini Dalmoro, Soraia Raupp Musse*

IEEE Computer Graphics and Applications, v. 42, n. 1, p. 32-46, 2021.
DOI: 10.1109/MCG.2021.3090198

Analysis of charisma, comfort and realism in CG characters from a gender perspective [ADM21]
***Victor Araujo**, Bruna Dalmoro, Soraia Raupp Musse*
The Visual Computer, v. 37, n. 9-11, p. 2685-2698, 2021.
DOI: https://doi.org/10.1007/s00371-021-02214-2

How Much Do We Perceive Geometric Features, Personalities and Emotions in Avatars? [ADF+21]
***Victor Araujo**, Bruna Dalmoro, Rodolfo Favaretto, Felipe Vilanova, Angelo Costa, Soraia Raupp Musse*
In: Advances in Computer Graphics: 38th Computer Graphics International Conference, CGI 2021, Virtual Event, September 6–10, 2021, Proceedings 38. Springer International Publishing, 2021. p. 548-567.
DOI: https://doi.org/10.1007/978-3-030-89029-2_42

Perception of Charisma, Comfort, Micro and Macro Expressions in Computer Graphics Characters [AWdS+21]
*Lucas Andreotti, Morgana Luiza Weber, Tiago Luz da Silva, **Victor Araujo**, Soraia Raupp Musse*
In: 2021 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). IEEE, 2021. p. 107-116.
DOI: 10.1109/SBGames54170.2021.00022

Perception of Personality Traits in Crowds of Virtual Humans [NSE+21]
*Lucas Nardino; Diogo Schaffer, Felipe Elsner, Enzo Krzmienski, **Victor Araujo**, Gabriel Fonseca Silva, Vinícius Jurinic Cassol, Rodolfo Migon Favaretto, Soraia Raupp Musse*
In: 2021 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). IEEE, 2021. p. 107-116.
DOI: 10.1109/SBGames54170.2021.00023

GranDGamesBR: Perceptual Analysis of Computer Graphics Characters in Digital Entertainment [MPDdAA21]

Soraia Raupp Musse, Greice P. Dal Molin, Bruna M. Dalmoro, ***Victor Araujo***
In: 2021 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). IEEE, 2021. p. 107-116.
DOI: `https://doi.org/10.5753/sbgames_estendido.2021.19753`

Perception of Computer Graphics Characters in Groups with Skin Color Diversity [dAASCM21]
Victor Araujo*, Diogo Hartmann Muller Schaffer, Angelo Costa, Soraia Raupp Musse*
In: 2021 20th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). IEEE, 2021. p. 107-116.
DOI: `https://doi.org/10.5753/sbgames_estendido.2021.19763`

Estimating Perceived Comfort in Virtual Humans based on Spatial and Spectral Entropy [MAM22]
Greice P. Dal Molin, ***Victor Araujo****, Soraia Raupp Musse.*
In: VISIGRAPP (4: VISAPP). 2022. p. 436-443.
DOI: `10.5220/0010831300003124`

How do we perceive Characters? An Analysis of Human Perception in Still Images, Animations and VR Scenarios [ADG+22]
Victor Araujo*, Bruna M. Dalmoro, Rafael Geiss, Márcio S. Pinho, Soraia Raupp Musse.*
In: Proceedings of Brazilian Symposium on Games and Digital Entertainment, 2022, Brasil. 2022.
DOI: `https://doi.org/10.5753/sbgames_estendido.2022.225436`

Towards Virtual Humans without Gender Stereotyped Visual Features [ASCM22]
Victor Araujo*, Diogo Schaffer, Angelo Brandelli Costa, Soraia Raupp Musse*
In: SIGGRAPH Asia 2022 Technical Communications. 2022. p. 1-4.
DOI: `https://doi.org/10.1145/3550340.3564232`

Mitigating bias in facial analysis systems by incorporating label diversity [KAVM23]
Camila Kolling, ***Victor Araujo****, Adriano Veloso, Soraia Raupp Musse*
Computers & Graphics, v. 116, p. 173-184, 2023.
DOI: `https://doi.org/10.1016/j.cag.2023.08.021`

Perceptual Analysis of Computer Graphics Characters in Digital Entertainment [MMA+23]
Soraia Raupp Musse, Greice Pinho Dal Molin, ***Victor Araujo****, Diogo Hartmann Muller Schaffer, Angelo Costa Brandelli*
In: Forum on Grand Research Challenges in Games and Entertainment. Cham: Springer Nature Switzerland, 2020. p. 207-232.
DOI: `https://doi.org/10.1007/978-3-031-27639-2`
Book Chapter

Evaluating the Uncanny Valley Effect in Dark Colored Skin Virtual Humans [dAACM23]
Victor Araujo*, Angelo Brandelli Costa, Soraia Raupp Musse*
In: 2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2023.
DOI: `https://doi.org/10.1109/SIBGRAPI59091.2023.10347145`

Revisiting Micro and Macro Expressions in Computer Graphics Characters [MRG+23]
Rubens Montanha, Giovana Raupp, Vitória Gonzalez, Yanny Partichelli, André Bins, Marcos Ferreira, ***Victor Araujo*** *and Soraia Musse*
In: 2023 22th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames).
DOI: `https://doi.org/10.1145/3631085.3631228`

A.2 Ongoing Publications

Examining the Attribution of Gender and the Perception of Emotions in Virtual Humans
Victor Araujo*, Júlia Melgare, Diogo Schaffer, Natália Pizzol, Viviane Souza, Angelo Brandelli Costa, Soraia Raupp Musse*
ACM Transactions on Applied Perception 2024
Submitted.

Evaluating the Gender Label in Virtual Babies Using a 3D Interactive Environment
Victor Araujo*, Catherine Pelachaud, Angelo Brandelli Costa, Soraia Raupp Musse*

IEEE COMPUTER GRAPHICS AND APPLICATIONS 2024
Submitted.

Perception of Micro and Macro Facial Expressions by Driven Animations in Realistic Virtual Humans
*Rubens Montanha, Giovana Raupp, Ana Schmitt, **Victor Araujo**, Soraia Raupp Musse*
Entertainment Computing 2024
Submitted.

Surveying the Evolution and Challenges in Virtual Human Applications
*Paulo Knob, Rubens Montanha, **Victor Araujo**, Greice Pinho, Gabriel Fonseca Silva, Vitor Peres, Soraia Raupp Musse*
Computers & Graphics 2024
Submitted.

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Pesquisa e Pós-Graduação
Av. Ipiranga, 6681 – Prédio 1 – Térreo
Porto Alegre – RS – Brasil
Fone: (51) 3320-3513
E-mail: propesq@pucrs.br
Site: www.pucrs.br