A Validation Study of the Human Perception of Virtual Characters' Features

Georgy L. J. de Freitas
Dep. of Computer Science
UERN
Natal-RN, Brazil
georgyfreitas@alu.uern.br

André Gustavo Pereira da Silva
Dep. of Computer Science
UERN
Natal-RN, Brasil
andregustavo@uern.br

Raul Benites Paradeda
Dep. of Computer Science
UERN
Natal-RN, Brazil
raulparadeda@uern.br

Abstract—**Several digital games use virtual characters to provide further immersion and interactivity to the player. In specific tasks, these characters influence players' decision-making. However, the player's perception of a virtual character is not always the one designed for that character. For example, when a character is designed to cause fear in the player but does not. In this sense, this paper describes a quantitative study validation of the appearance of virtual characters to be used in an interactive narrative game. The participants informed their perceptions regarding the temper, clothes, and behavior of 40 characters designed for this study. From the results, the player characteristics, such as gender and age, can influence the character's perception. Another result observed was that differences in visual aspects (e.g., well or poorly dressed) are better perceived when compared to differences between behavioral characteristics (e.g., honest and dishonest).**

Index Terms—*Non-Player Character*, **Virtual Character, Appearance Validation**

I. INTRODUCTION

Currently, virtual characters are increasingly present in people's daily lives, whether to perform a virtual service or in educational programs such as Duolingo[1]. In digital games, virtual characters are essential to convey greater immersion and interactivity for the player. In these environments, these characters are called NPC (*Non-Player Character*). The NPC is a resource used to perform various tasks, such as influencing the player's decision-making to follow a specific path, giving a game element to the player, and increasing immersion, among others. The strategies used to influence the player can be the most diverse, from a simple gaze to a persuasive dialogue or assertive behavior. [8].

Designers and developers have pre-defined perceptions when creating the game's characters and their aspects. So, when creating a character, designers follow existing patterns, often inspired by movies, series, or even real life. However, even using any of these patterns, it is necessary to validate the appearance of these characters. After all, how will they know if a character will cause the player the desired perception?

Based on that, we carried out a quantitative study to validate characters that will be used in an interactive narrative game developed in the future. In this study, we collected data that can point out the player's perception of the characters presented.

Among these characters, we will select those that will integrate the game, considering whether the participants' perception is similar to that for which the characters were designed. In this sense, we created forty characters, where each volunteer should answer about their perception concerning the following three characteristics: temper (mean or kind), clothing (poorly dressed or well dressed), and behavior (dishonest or honest). We believe that these characteristics suit the game's purpose, where the NPCs will be used to indicate paths and persuade the player to make a decision.

II. RESEARCH QUESTIONS

The present study aims to answer some research questions (RQ) to validate the appearances of the developed NPCs. These RQs are:

- RQ1: Do participants of different genders perceive the virtual character's features of NPCs differently?
- RQ2: Do participants from different generations perceive the virtual character's features of NPCs differently?
- RQ3: Are some of the virtual character's features (temper, clothes, and behavior) better perceived than others?
- RQ4: Is the perception designed for the NPCs the same perception obtained by the participants?

III. RELATED WORKS

A. Difference in Perception between Men and Women

Several kinds of research have already proven differences in the preferences of men and women [1]. There are also differences in cognitive performance, personality, social behavior, and psychological well-being [5]. For example, in the study of [3], the authors investigated sex differences in male and female body shape perception and found a significant difference in the categorization of attractiveness between women and men.

In this way, we consider it essential to identify if the perception of NPC characters is different or the same among participants of different genders. This aspect can lead us to find clues that there is a correlation between the influence one wants to attribute to the NPC and the player's gender.

B. Difference in Perception Between Age Generations

The diversification of users is crucial because it can bring exciting opportunities and challenges [4], which can generate

---

[1]Duolingo. URL: https://www.duolingo.com/. Accessed on 08/31/22.

different points of view, solutions, and approaches to a given problem.

Based on that, people belonging to different generations may have different perceptions [9], mainly in the case of virtual agents who are interacting with them. In this sense, [12] defined categories of virtual agents that were analyzed by users of different generations, having observed that older people prefer to be served by more realistic agents, while younger people prefer to interact with more comical agents.

Given the results obtained by the works mentioned above, the importance of investigating whether different generations present different perceptions concerning the characters designed for this study is perceptible.

C. Appearance Validation

Appearance validation of virtual agents and NPCs is standard in the gaming industry. After all, how to identify that a specific character refers to a particular sensation (fear, trust, hope, etc.)? For example, how to know that a given virtual agent for a bank business does not cause weirdness in the customer? For this, it is necessary to validate the characteristics and behaviors defined for these characters to verify that people perceive them in the way the developer foresaw.

In the study described in [7], the authors aimed to conduct an NPC's version evaluations. This character provided guidance and companionship to players, who had to indicate which version they preferred for a given task. According to the results, most players preferred colorful, naturalistic, funny, and friendly NPCs. In another work, [2] investigated the influence of appearance and personality traits of NPCs on gameplay and player decision-making. The results provided a better understanding of the perception of virtual characters in games and information about factors underlying aggressive behavior in video games.

As mentioned before, the characters used in the present study will be used in an interactive narrative game if they have their appearance validated; that is, the characters drawn will be used in the game as NPCs if the perception of the participants of this study about each character is the perception desired by its creators. For example, if a character was designed to be perceived as mean, poorly dressed, and dishonest and was not perceived in this way by the participants, this character will not be used in the future game.

IV. METHODOLOGY

A. Participants

A convenient sample of 99 participants was invited via social media to fill out an online form with images of the NPCs designed for the study. Of these participants, 35 are male, 62 are female, and two are unidentified. Regarding the age group, two participants claimed to be under 18 years old (Generation Alpha), 25 participants reported being between 18 and 25 years old (Generation Z), 36 participants were between 26 and 35 (Generation Y or Millennial), 24 reported having between 36 to 45 years old (Generation Y or Millennial), 6 participants

reported being between 46 and 55 years old (Generation X) and 6 participants over 55 years old (Generation Boomer).

B. Materials

To find the answers for this research, we carried out a quantitative study where a questionnaire was applied to identify the profile of the study participants and their perception of the characters drawn. For each of the characters drawn, we asked the participant's perception of the temper (mean or kind), dress (badly dressed or well dressed) and behavior (dishonest or honest) of each character. The questionnaire used was hosted on the Google Forms platform.

C. Manipulation

In the questionnaire, the user's perception of each characteristic of the NPC can vary within a bipolar Likert scale from 1 to 5, where 1 represents the most significant perception for the values mean, poorly dressed, and dishonest, while value 5 identifies the most perceived perception to the values kind, well dressed and honest. For example, if a participant scored 1 for temper and 5 for clothing, it meant that the participant perceived the NPC as mean and well dressed, respectively. The average value (3) of the Likert scale for each feature means that the participant did not clearly perceive what was being measured.

To achieve the objective of this study, 40 NPCs were designed, representing each of the following combinations: kind/well dressed, mean/well dressed, kind/poorly dressed, and finally, mean/poorly dressed. According to the designer's perception, each character was designed to belong to a specific category[2]. Besides, to measure whether the participants' perception of the NPCs is the same as that of the designer, the NPCs were arranged randomly and not sequentially concerning each category they belong to. The character presentation was grouped by each combination described above until all NPCs drawn were presented. For each character, we elaborated three questions related to the mentioned characteristics, where the Likert scale identifies the participant's perception of the presented image, as can be seen in Fig. 1.

D. Data analysis

The data obtained did not deviate from a normal distribution according to the Shapiro-Wilk [11] normality test ($p = .000$) performed using the SPSS (v.28) program. In this way, we performed two tests. The first test was to verify the differences in the perception of the NPCs' evaluated characteristics (temper, clothes, and behavior) according to the participant's gender and age group, using the Kruskal-Wallis test. The second test verifies the statistical differences within each group of evaluated characteristics (for example, if the participants perceived the image of NPC 1 differently concerning their clothing) through the Kolmogorov Smirnov test. In the tests, 5% of statistical significance was considered, that is, values of $p <= 0.05$.

---

[2]All the NPCs designed were at: https://osf.io/59uzj/. Accessed at: 09/05/22.

Fig. 1. Example of NPC and questions that were presented.

According to the image of the character below, point your perception of him (with *
1 strongly agreeing with the first characteristic and 5 with the second):

	1	2	3	4	5
Mean(1) / Kind(5)	○	○	○	○	○
Poorly dressed(1) / Well dressed(5)	○	○	○	○	○
Dishonest(1) / Honest(5)	○	○	○	○	○

V. RESULTS

RQ1 Firstly, the average values indicated on the Likert scale were calculated for each of the three characteristics evaluated to answer the first research question. For the character temper (mean or kind), the mean reported by the participants was equal to 3.1 with a standard deviation of 0.51. For clothing (poor dressed or well dressed), the mean was 3.2, and the standard deviation was 0.54. Finally, for the behavior characteristic (dishonest or honest), the mean was 3.12 with a standard deviation of 0.55. Then, to answer the RQ1, that is, to verify if there is a significant difference between the perceptions of the participants of different genders, the Kruskal-Wallis H test was applied. The result showed that there was no statistically significant difference in the score regarding the perception of temper ($X^2(2) = 0.196, p = .907$), clothing ($X^2(2) = 2,083, p = .353$) and behavior ($X^2(2) = 2010, p = .366$) among the different genders indicated by the participants.

RQ2 To answer whether participants from different generations perceive the characteristics of NPCs differently, the Kruskal-Wallis H test was applied to the calculated mean for each characteristic measured. In addition, participants were grouped by generation according to the range selected for their age. Moreover, participants who reported being between 26 and 35 years old or between 36 and 45 years old are part of Generation Y. The Kruskal-Wallis H test showed that there was no statistically significant difference in the scores regarding the perception of temper ($X^2(2) = 2,367, p = .669$) and behavior ($X^2(2) = 4,692, p = .320$) between the different age generations of the participants. As for the clothing characteristic, the test found a statistically significant difference ($X^2(2) = 9,938, p = .041$). When performing a more in-depth test to determine which of the generations presented statistical differences in the perception of the clothing characteristic, the Mann-Whitney U test identified that the difference was between generations Z and Millennials ($U = 87, p = .034$).

RQ3 The one-sample Kolmogorov-Smirnov test was applied to the average scores perceived by the participants for each of the evaluated characteristics. This statistical test was applied to identify whether the sample has a specific distribution within each characteristic. This test can provide evidence that perceptions may or may not have been perceived differently. The Kolmogorov-Smirnov test indicated that the character temper ($D(99) = .098, p = .021$) and behavior ($D(99) = .091, p = .043$) did not follow a normal distribution, while the clothing characteristic ($D(99) = .065, p = .200$) followed a normal distribution.

RQ4 The chi-square test to a sample was applied to the answers of each evaluated characteristic of the 40 NPCs to answer the last research question. From the statistical test, it was possible to perceive that the NPC1 had a significant difference in the perceptions of temper ($p = .07$) and behavior ($p = .07$), for NPC6 ($p = .06$) and NPC27 ($p = .06$) the difference in perception was in the characteristic of clothing, and the NPC34 ($p = .24$) and NPC38 ($p = .14$) the significant difference was presented only for the character temper. All other NPCs showed no significant differences ($p <= .05$)[3].

VI. DISCUSSION

A. Perception differences between genders

The results showed that, regardless of gender, people similarly perceived the characters; that is, participants presented the same perception for most NPCs, e.g., most participants perceived the third NPC as well dressed, kind and honest. This result differs from [6], in which the authors found differences regarding the perception of the appearance of a virtual animated agent between the participants' gender. In that study, male participants were more influenced in a decision-making task by the agent's appearance than females. However, the authors were evaluating whether the agent was more or less attractive and its persuasiveness. Our study evaluated three distinct characteristics based on the participant's perception of a virtual image. Plus, our results showed that the drawn NPCs would be perceived similarly to whoever interacts with them, regardless of gender, which can be essential for a game scenario.

[3]The detailed results are available at: https://osf.io/59uzj/. File: Data_validation_perception.pdf. Accessed at: 09/06/22

B. Difference in perception between age generations

Of the three characteristics evaluated in this study in each NPC, only the clothing presented different perceptions in relation to the age generation of the participant. We observed differences in perception between Gen Z teenagers (ages 18-25) and Millennial adults (ages 26-35). The difference in perception between generations was also observed in the work of [9]. The authors identified differences in several aspects evaluated, such as motivation, feelings (optimism and perception of learning), perception of the place of study, and technology knowledge. This difference corroborates the work of [10], who found evidence that reinforces this difference. In our study, this behavior may have happened because the characters drawn had clothes from older times, given that they were validated to be used in a game with a medieval narrative. Thus, some younger participants may have perceived the old dress as poorly dressed, while older participants did not. On the other hand, participants did not perceive differences in the temper and behavior of NPCs. Thus, we conclude that, regardless of the participant's generation, these two characteristics of the NPC will be perceived similarly.

C. Perception of the evaluated characteristics

The results showed that only the clothing characteristic followed the normal distribution of perception among the participants. On the other hand, the characteristics of temper and behavior showed an opposite performance. Thus, we can infer that the characteristic of clothing (poorly dressed or well dressed) was better observed among the participants, thus having a more accurate perception. This effect could be observed because, in this characteristic, the score attributed by the participants was mostly above or equal to 4 or below or equal to 2. The opposite happened with the characteristics of temper and behavior, where it was impossible to identify the need if the NPC is mean or kind, or dishonest or honest, as the scores were close to 3. These differences between perceptions may have happened because we believe that the clothing characteristic may have been better identified in the drawing, in addition to being more perceptible through an image than the temper and behavior of a virtual character.

D. Final Perception of NPCs

We also noticed that most characters did not present significant statistical differences related to the perception of the evaluated characteristics. Thus, it is possible to infer that there were no different perceptions for most characters; that is, the images presented did not generate doubts regarding the analyzed characteristics. Only five characters had different perceptions from the participants. For example, for a given character, some participants perceived it as mean, while others perceived it as kind. This behavior means that the perception that the developer wanted these characters to present could not be validated. Therefore, if the participants disagreed with the same perception of an NPC, it was not clearly drawn to the point of having a perception agreement. Thus, among the 40 characters evaluated, 35 had characteristics perceived similarly among the participants, according to the developers' perception.

VII. Conclusion

This study aimed to validate the human perception of the appearance of virtual characters that will be used in an interactive narrative game. This validation is essential to identify which characteristics can arouse the desired perception in players. Furthermore, this study provides evidence that, according to their profiles, such as gender and age, players may have similar or different perceptions, depending on the character characteristic to be addressed. We observed that visual characteristics, such as clothing, are better observed when compared to behavioral characteristics. For this reason, to try to activate specific behavioral perceptions in the observer, the designer needs to focus on more details to obtain this perception. Finally, it is expected to contribute to the development of better, more influential/customizable characters, where there is human-agent interaction and perceived more efficiently within the defined characteristics, whether they are to be used as NPCs, Virtual agents, or otherwise as the developer wishes.

Acknowledgment

References

[1] Rachel Croson and Uri Gneezy. Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74, 2009.

[2] Ylva Ferstl, Elena Kokkinara, and Rachel McDonnell. Facial features of non-player creatures can influence moral decisions in video games. *ACM Transactions on Applied Perception (TAP)*, 15(1):1–12, 2017.

[3] Adrian Furnham and Sally Radley. Sex differences in the perception of male and female body shapes. *Personality and Individual differences*, 10(6):653–662, 1989.

[4] Xin Gao. On the construction of university culture under the background of value diversification. In *1st International Conference on Arts, Design and Contemporary Education (ICADCE 2015)*, pages 761–763. Atlantis Press, 2015.

[5] Janet Shibley Hyde. Gender similarities and differences. *Annual review of psychology*, 65:373–398, 2014.

[6] Rabia Fatima Khan and Alistair Sutcliffe. Attractive agents are more persuasive. *International Journal of Human-Computer Interaction*, 30(2):142–150, 2014.

[7] Stephanie B Linek, Daniel Schwarz, Ganit Hirschberg, Michael Kickmeier-Rust, and Dietrich Albert. Designing the non-player character of an educational adventure-game: the role of personality, naturalism, and color. In *Proceedings of the International Technology, Education and Development Conference*, 2007.

[8] Raul Paradeda, Maria José Ferreira, Raquel Oliveira, Carlos Martinho, and Ana Paiva. The role of assertiveness in a storytelling game with persuasive robotic non-player characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 453–465, 2019.

[9] Raul Benites Paradeda and Heide Vanessa Souza Santos. Factors that negatively influence students' transition from the traditional classroom to emergency remote education (ert). *Computers and Education Open*, page 100098, 2022.

[10] Alexandru Popa and Corina Pelau. Differences in the clothing brand perception depending on generation/diferente în perceptia brandurilor de îmbracaminte în functie de generatie. *industria Textila*, 67(4):260, 2016.

[11] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[12] Carolin Straßmann and Nicole C Krämer. A categorization of virtual agent appearances and a qualitative study on age-related user preferences. In *International Conference on intelligent virtual agents*, pages 413–422. Springer, 2017.