How do true eyes really move in video sequences?

Willian Dias, Lucas Felix, Julia Melgare, Victor Araujo, Paulo Knob, Soraia Musse
School of Technology, Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul
Email: {willian.dias,lucas.salaverry,julia.melgare,victor.flavio,paulo.knob}@acad.pucrs.br,
soraia.musse@pucrs.br

***Abstract*—Realistic animation of characters is an important topic in our contemporary world, that includes application as games and movies. Uncanny Valley is one aspect which try to measure how the human perception works related to virtual character's faces. In this work, we investigate the relevance of the eyes in people communication and perception. We use OpenPose to detect data from the eyes of a real human face and then use it to animate a virtual character's eyes. We performed an evaluation regarding the impact of such animations when compared to randomly modeled eyes and animated eyes from games and movies.**

***Keywords*-uncanny valley; virtual character's; eyes; perception;**

I. Introduction

In recent years, advances in Computer Graphics (CG) have allowed the entertainment industry to create very realistic virtual humans [1] in terms of animation of their bodies and faces [2]. Increasingly, real actors have been replaced by CG characters and often this substitution is not even perceived by the public, but there are still some perceived oddities, such as the movement of the mouth and eyes [3]. Human perception is also a theme present in many researches in CG [4] and can be very important in the evolution of virtual humans and realistic faces [1], [2]. According to Mori [5], robots made to appear too similar to real humans can fall into the "Uncanny Valley", where a too high degree of human realism evokes an eerie feeling in the viewer. The effect of Uncanny Valley hypothesis on CG characters has become increasingly influential in scientific studies [6], [7], [8] and although the concept of Uncanny Valley in Computer Animation is very popular, some questions are still unanswered and we are interested about a specific one: *i)* What is the relationship between emotion expressiveness perceived by people and the effect of Uncanny Valley? *ii)* Is human perception of comfort conveyed by CG characters affected by their eye movement?

In this work, our goal is to evaluate the importance of eye movement and the emotions transmitted by them through communication and human perception. For this, we collect trajectory data from real eyes movement using OpenPose[1] [9], use this trajectory to animate eyes created with CG, and then evaluate previously mentioned questions through questionnaires about CG characters and CG eyes.

The remainder of this paper is organized as follows: Section II mentions some related work presented in literature, and Section III describes the proposed methodology. Results are detailed and discussed in Section IV, while Section V addresses final comments and future work.

[1]https://github.com/CMU-Perceptual-Computing-Lab/openpose

II. Related Work

This section discusses some work related to the analysis of Uncanny Valley effects caused by CG characters. The Uncanny Valley is a theory created by roboticist Masahiro Mori [5] who analyzes the emotional reaction of humans to artificial beings. According to this theory, if robots have a high degree of realism close to humans, they may fall into the "Uncanny Valley," which can cause an eerie impression on the viewer. Thus, when robots show signs of life, such as movement or conversation, it tends to generate high peaks and valleys, changing the shape of the Uncanny Valley.

From Mori's work, several other works have used the Uncanny Valley to measure the discomfort of artificial characters (robots, characters created with computer graphics, among others). For example, in the work of Katsyri et al. [7], the authors reinterpreted the original Uncanny Valley hypothesis and revisited empirical evidence for theoretically motivated Uncanny Valley hypotheses. Thus, this work helped to understand and form the axes of the Uncanny Valley graphs (x-axis being the human likeness, and the y-axis being the comfort) in this present paper.

The effect of the Uncanny Valley hypothesis has been investigated in computer graphics on human perception of 3D models. Flach et al. [10] investigated the Uncanny Valley theory to evaluate their 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, to obtain answers to the following research questions: "Does the Uncanny valley exist in CG characters?" and "Does adding movement to these characters change the shape of the Uncanny Valley curve, like Mori suggested?".

III. METHODOLOGY

The methodology of this paper is organized in three phases as described in three next sections.

A. Characters from Games and Movies

The first stage of this work is the selection of the characters from games and movies to be analyzed. All characters are illustrated in Figure 1 and each character is accompanied by a legend of its origin, which may be a movie, game, or an animation we found on the Internet from various origins.

The choice of characters was limited to movies, games, series, among others, to a maximum of five years ago. Thus, as proposed by many work in the literature, such as Flach et al. [10], we evaluated the human likeness criterion. It contributes with the choosing order of the characters in the horizontal axis of the Uncanny Valley graphics, i.e, listing from left (small value in this axis) to right (higher values) the less realistic characters to more realistic. The vertical axis is responsible to, based on human perceptions, quantify the uncanny feeling or comfort perceived by people. To ensure the characteristic of human likeness, we chose some characters that could represent a human being in a more realistic way, as shown in the case in Figure 1(f). These cases attempt to present simulated virtual humans with high levels of realism. In addition, we selected characters present in movies, series and games, as we can see in Figure 1(a, b and c), which also represent a high level of realism. Therefore, we need characters that escape from realism, counteracting the others cited earlier, such as the cartoon characters shown in Figure 1(d, e and g). This counterpoint is needed to form the horizontal axis of the Uncanny Valley quoted above and goes from low to high realism.

B. Data Extraction and Eye Animation

In addition to the games and movies characters, we modeled a pair of 3D cartoon eyes at Unity (see Figure 1 (h and i)). These eyes have been animated taken into account a real expression of real eyes and also animated using a random saccade behavior as in [11]. The data used in the virtual eye's animation was obtained using the OpenPose[2] [9] software over a video data set. Various facial points (eyes, mouth, ...) can be extracted, obtaining as result a text file for each frame of the video. Each point has an identification number, the image coordinates and a assertion rate. For the eye's animation in the present work, the coordinates were the only extracted information during the file parsing. For our purpose, only a single video from single emotion (angry) was used, due to the fact that this video sequence presented an emotion easily perceived by the authors.

As the data was extracted, a conversion from image coordinates to world (Unity) coordinates was needed. For this purpose, the distance between the pupil position relative to the eye's horizontal extremities was used. An equivalent approach was also used for the vertical distance. The eye's rotation was then computed by projecting the motion vector to a specific pupil angle. For both horizontal and vertical movements, the rotation limits for the world coordinates were empirically tested, but took into account the ones presented in Andrist et al. [12]. Finally, for a smoother transition, a small interpolation was made between the frames.

For the randomly animated eyes, some of the ideas presented in Lee et al.[11] were used. The eye's rotation was done using the magnitude equation presented in the paper, which was responsible for the random coordinates. Differently from the data animated eyes, the rotation from the randomly ones was done always considering the initial (central) rotation of the eyes, not the last one.

[2]https://github.com/CMU-Perceptual-Computing-Lab/openpose

C. The Questionnaires

We formulated two surveys to conduct our research. First, as we wanted to analyze the effects of Uncanny Valley on human perceptions of CG characters, we use the same two questions to evaluate the comfort (y-axis) and human likeness (x-axis) that were used in work of Flach et al. [10]. These questions are listed in Table I named Q2 and Q1 questions. Before answering the survey, participants received no explanation of the original intent of the research. This was done to avoid any type of influences on participants' responses. All participants were asked if they agreed to cede their answers and personal information to the survey. They answer questions as age, gender, schooling and level of familiarity with Computer Graphics. In addition, the questionnaires were made available through social media.

Table I
QUESTIONS REGARDING HUMAN PERCEPTION APPLIED TO THE PARTICIPANTS.

Question	Possible answers
Q1: If created with CG, how realistic does it seem?	*a*) Very realistic *b*) Moderately realistic *c*) Unrealistic *d*) Don't know
Q2: Do you feel some discomfort (strangeness) looking to this character?	*a*) Yes *b*) No *c*) Don't know
Q3: Do you perceive realism (as a real person) in the emotion conveyed by the eyes?	*a*) Yes *b*) No *c*) Don't know
Q4: Do you think the eye movements of Computer Graphics originated from the movements of a real person?	*a*) **Yes, in video 1** *b*) Yes, in video 2 *c*) In both videos *d*) In neither video *e*) Don't know

The first questionnaire was used to evaluate perception of characters from (a) to (g) in Figure 1. The process is following described: image and video of each character were shown before asking questions Q1 and Q2 as referred in Table I (we did not use images h and i in this evaluation). It aims to compare the level of comfort (Q2) people feel when

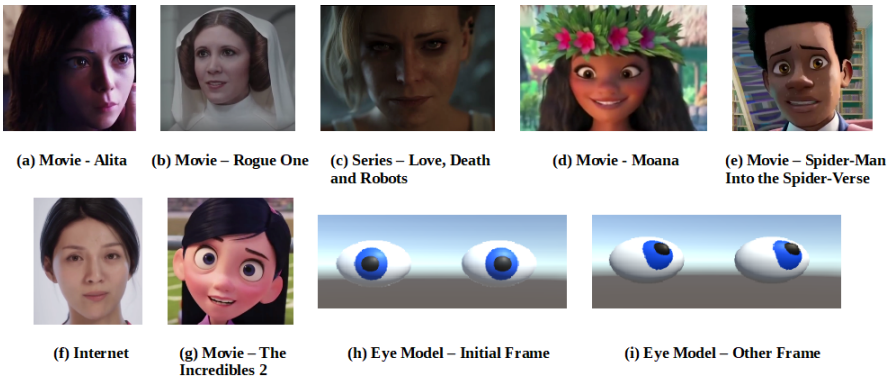

Figure 1. All the CG characters and 3D eye model (h - initial frame and i - other frame) used in this work.

observing characters in the pictures, where they are static and in the videos where they are moving. The vertical axis (Comfort) in Figure 2 is given by the percentage of the "Yes" and "No" answers of question Q2, that is, the more comfort the character provides in the human perception, the higher it is on the Uncanny Valley graph. In question Q1, people were asked how realistic the character was. As shown in the work of Katsyri et al. [7], human likeness can be varied in an almost infinite number of different ways. Question Q1 covers the realism criterion and was used to order the characters as presented in horizontal axis in Figure 2.

The second questionnaire was used to assess whether human perception of characters changes if the focus is only on the eyes of the characters. In this questionnaire there was only one process: we presented cropped videos focused on each character's eyes and ask questions Q1, Q2 and Q3 mentioned in Table I. In addition to the characters observed in questionnaire 1, we included a pair of 3D cartoon eyes animated at Unity (as illustrated in Figure 1 (h) and (i)). Character (h) was randomly animated, while (i) was animated using a real eyes movement. Question Q3 assesses the emotion expressiveness transmitted through the eyes of the games/movies characters and the 3D cartoon eyes, as perceived by the participants. Question Q4 from Table I was asked after play again the two modeled eyes (characters (h) and (i)) and aims to evaluate in which of the two pairs of eyes participants consider being originated from a real person. Q4 was the only question with a correct answer, as marked in bold in the table (first answer). Once having prepared the surveys we proceed with the performed tests with participants. Section IV details this procedure.

IV. Results

This section presents the results obtained in this paper, describing results when studying Uncanny Valley with still images and animations, and the effect of focusing only on the characters' eyes with respect to human perception. The first questionnaire (evaluating comfort in images and video w.r.t virtual human faces) was answered by 119 participants, being 58% male and 42% female. Regarding the age group, 40.3% of people is under 20 years, 37% between 20 and 30 years, 16.8% between 31 and 50 years, and 5.9% over 50 years. The second questionnaire (evaluating comfort and emotion expressiveness focused on virtual human eyes) was answered by 26 participants, 57.7% male and 42.3% female. About age, 53.8% were between 20 and 30 years old, 30.8% between 31 and 50 years, and 15.4% under 20 years. In these analysis we identify the characters as the number giving in vertical axis in Figure 2 (from 1 to 9). Such figure shows the result of the human likeness (horizontal axis) obtained with question Q1 (Table I). We can see that less realistic characters were more to the left in Figure 2, while the more realistic ones were to the right of the horizontal axis. The vertical axis is given by the percentage of people's responses to the image, video and eyes comfort perception (in yellow, red and blue lines, respectively), and the perception of eyes emotion for each character (in green line), obtained by answers from Q3. In the context of Uncanny Valley, we can see that character 6 caused the valley, in terms of both image and video comfort (yellow and red lines), as we can see in Figure 2. Average and standard deviations values for image comfort (75.75%, 20.65%) and video comfort (76.35%, 13.93%) were close, with little advantage for images. So for this character, the indication is that the movement did not influence the final perception of comfort. When we cropped the videos on the eyes of characters and asked about comfort (blue line), through results obtained by question Q2, we can see that comfort increases in almost all characters except the one on the right (9). So, it may indicate that the part of the face that can cause strangeness was not the eyes, in the studied scenarios.

The green line indicates the percentage of participants who perceive emotions in the eyes. We can see that unrealistic characters (the five on the left: characters from 1 to 5) have smaller values for emotion expressiveness, while the comfort attributes (lines blue, red and yellow) were higher [3].

[3]We did not test the characters 1 and 2 in the first questionnaire

The mean of comfort attribute was high, i.e. 81.5%, with standard deviation of 8.45% for characters 3, 4 and 5, while the same metric for characters from 6 to 9 is 83.2% with standard deviation of 17.96%. The other interesting metric is the emotion expressiveness. Characters 3, 4 and 5 presented an average of 60.9% and standard deviation of 6.88%, while characters from 6 to 9 had the highest values of realism conveyed by eye emotions (average = 93.5% and standard deviation= 20.45%). So, it is interesting to see that less realistic characters (from 1 to 5) presented high comfort but smaller values for emotion expressiveness, while the opposite happens in more realistic characters.

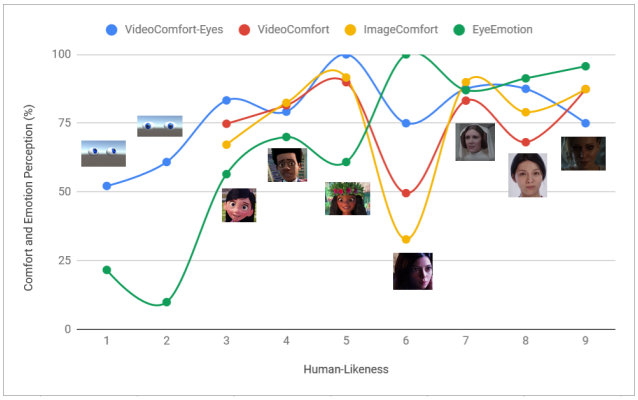

Figure 2. Attributes values (%) VS Human Likeness.

Regarding Q4, comparing only characters 1 and 2, 29.5% of participants considered that the character 1 was animated based on a real human, while 29.5% chosen character 2 (correct answer) and 41% said that both of videos were animated using real motion.

V. Final Considerations

This short paper presents a study we performed regarding comfort and expressiveness of virtual characters. Nine characters have been evaluated by participants. We tried to answer the following questions: *i)* What is the relationship between emotion expressiveness perceived by people and the effect of Uncanny Valley? *ii)* Is human perception of comfort conveyed by CG characters affected by their eye movement? Regarding the first question, observing Figure 2 (in green line), we can conclude that the more realistic the character (less feeling of Uncanny Valley), the higher values for emotion expressiveness, as shown in Section IV. Regarding the second question, we can see, from the blue line in Figure 2, that the comfort of CG character eye videos seems to have a correlation with the effect of Uncanny Valley on human perception.

References

[1] J. Jimenez, K. Zsolnai, A. Jarabo, C. Freude, T. Auzinger, X.-C. Wu, J. von der Pahlen, M. Wimmer, and D. Gutierrez, "Separable subsurface scattering," in *Computer Graphics Forum*, vol. 34, no. 6. Wiley Online Library, 2015, pp. 188–197.

[2] L. Ma and Z. Deng, "Real-time facial expression transformation for monocular rgb video," in *Computer Graphics Forum*, vol. 38, no. 1. Wiley Online Library, 2019, pp. 470–481.

[3] M. Mustafa and M. Magnor, "Eeg based analysis of the perception of computer-generated faces," in *Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016)*. ACM, 2016, p. 4.

[4] V. Araujo, R. M. Favaretto, P. Knob, S. R. Musse, F. Vilanova, and A. B. Costa, "How much do you perceive this? an analysis on perceptions of geometric features, personalities and emotions in virtual humans (extended version)," *arXiv preprint arXiv:1904.11084*, 2019.

[5] M. Mori, "Bukimi no tani [the uncanny valley]," *Energy*, vol. 7, pp. 33–35, 1970.

[6] J. Kätsyri, M. Mäkäräinen, and T. Takala, "Testing the uncanny valleyhypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli," *International Journal of Human-Computer Studies*, vol. 97, pp. 149–161, 2017.

[7] J. Kätsyri, K. Förger, M. Mäkäräinen, and T. Takala, "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, p. 390, 2015.

[8] M. Mustafa, S. Guthe, J.-P. Tauscher, M. Goesele, and M. Magnor, "How human am i?: Eeg-based evaluation of virtual characters," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 5098–5108.

[9] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *CVPR*, 2017.

[10] L. M. Flach, R. H. de Moura, S. R. Musse, V. Dill, M. S. Pinho, and C. Lykawka, "Evaluation of the uncanny valley in cg characters," in *Proceedings of the Brazilian Symposium on Computer Games and Digital Entertainmen (SBGames)(Brasília)*, 2012, pp. 108–116.

[11] S. P. Lee, J. B. Badler, and N. I. Badler, "Eyes alive," in *ACM transactions on graphics (TOG)*, vol. 21, no. 3. ACM, 2002, pp. 637–644.

[12] S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher, "Designing effective gaze mechanisms for virtual agents," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2012, pp. 705–714.