The Uncanny Valley and the Importance of Eye Contact

Valentin Schwind[1,2], Solveigh Jäger[2]

Institute for Visualization and Interactive Systems, University of Stuttgart[1]
Institute for Games, Stuttgart Media University[2]

Abstract

The Uncanny Valley hypothesis describes the negative emotional response of human observers that is evoked by artificial figures or prostheses with a human-like appearance. Many studies have pointed out the meaning of facial features, but did not further investigate the importance of eye contact and its role in decision making about artificial faces. In this study we recorded the number and duration of fixations of participants (N = 53) and recorded gaze movements and fixations on different areas of interest, as well as the response time when a participant judged a face as non-human. In a subsequent questionnaire, we grasped subjective ratings. In our analysis we found correlations between the likeability and the duration of eye fixations on the eye area. The gaze sequences show that artificial faces were visually processed similar to the real ones and mostly remained not assessed as artificial as long as the eye regions were not considered.

1 Introduction

Artificial systems are often designed according to human expectations and simulate human-like appearance to improve ways of interaction or communication. But a certain phenomenon causes that observers of such realistic figures reject a very human-like artificial representation. The Uncanny Valley hypothesis suggested by the roboticist Masahiro Mori (Mori 1970) describes the negative emotional reaction of a human towards figures or prosthetics that are not quite human-like. The term emerges from the function curve which illustrates the relationship between affinity and human-like appearance (Figure 1). The more human-like characteristics a figure has, the more likely it will be accepted. Nevertheless, at a certain point the similarity to humans causes a reverse effect, and the affinity rapidly changes to aversion, eeriness, or repulsion. The figure appears uncanny to its human observer and falls in the aforementioned valley. Only a distinctively real human is fully accepted by observers. (MacDorman 2006, MacDorman et al. 2009). Robotic engineers, puppet designers or computer

animators seek to improve the subjective appearance and behavior of their figures in an intended way. The objective of this research is to understand the cognitive proceedings in the course of the observation of human-like figures. This knowledge might help character designers to elaborate anthropomorphic features and to design a volitional acceptance of artificial figures. Difficulties arise since humans perceive the ambiguity in categories when they cannot clearly distinguish between a human and an object. Therefore the current research focuses on the examination of differences in visual processing of ambiguous faces, which allegedly caused difficulties in categorization.

The present paper focuses on differences in eye movement behavior depending on the type of figures that are observed. The question is how eye contact is relevant for categorization of negatively rated characters and whether there are fundamental differences in observation schemes compared to a situation when photos of ordinary people are observed. Answers to that could give a better understanding of human perception with artificial entities and could provide hints how to improve interactions with human-like systems. Our approach to studying human visual perception using eye tracking. This technique has had a profound influence on our understanding of the mechanism of recognition of faces and has been established as an important method of collecting empirical measurements (Just & Carpenter 1980).

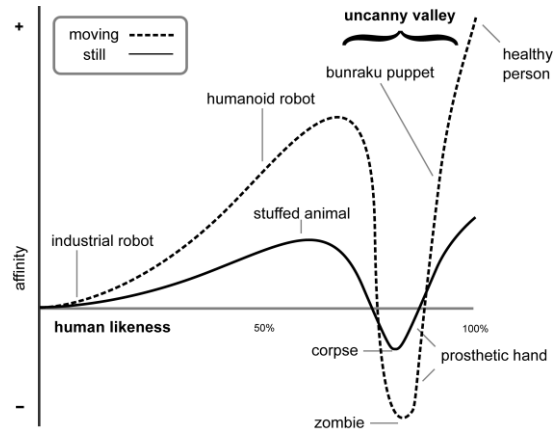

Figure 1: Simplified diagram (illustration from MacDorman 2005) of Mori's graph shows the generally positive feelings of an observer towards artificial entities with a certain degree of human-likeness.

2 Related Work

Approaches using the dimension of human likeness (DHL) with a linear morph continuo (Cheetham 2011, Cheetham et al. 2013) lead to prolonged response times during a categorization task with ambiguous stimuli. Difficulties in categorization lead to negative ratings or feelings (Yamada et al. 2013). Green et al. (2008) also recognized that face proportions such as differences in jaw width or face height have an impact on the subjective perception of human-likeness. It is assumed that judgments about familiarity and the respective categorization as well as the notion of an average face are formed basing on previously

recognized faces (Bruce et al. 1991). MacDorman (2005) assume that eerie anthropomorphic figures trigger the subconscious fear of death. The Uncanny Valley hypothesis refers both to the overall impression created by a figure as well as to its prostheses. However, in the perception of human-like figures, increased attention is undoubtedly paid to the face and the eyes (Itier et al. 2007, Janik et al. 1978, Langten et al. 2000, Ro et al. 2007, Vinette et al. 2004). Farah et al. (1998) assume that face recognition is fundamentally different from object recognition. Even the simplest stroke patterns (e.g. emoticons) or templates are attributed to facial human characteristics. Of all facial features, the eye area draws the most attention and the highest number of eye fixations (40%), and thus shows the highest attention rate (Itier et al. 2007, Janik et al. 1978). Previous studies on artificial figures underscore this assumption and partially emphasize the importance of the eyes, without further investigations about the eye regions (Cheetham et al. 2013, Looser & Wheatley 2010, MacDorman et al. 2009).

3 Material and Methods

3.1 Stimuli

68 images of persons and figures were created with 3D modelling software, captured from movies and games or chosen from the internet. Authors and owners of images gave consent for the use of their images in this study. The sample composition represents a cross section of character styles that are related to the scale of human-likeness in Mori's graph and have been examined or mentioned in previous studies. Based on the type of face, the portraits were classified into these 8 categories: 13 real persons in photographs (ordinary human), 21 computer-generated (CGI), 9 cartoon figures, 5 wax sculptures, 5 geminoids (androids), 5 humans with visible impacts of cosmetic surgery, 3 deceased (post-mortem), and 3 hyper-realistic cartoon figures with a real look. All depicted figures have a neutral facial expression and an upright posture. None of the stimuli was manipulated. The only change applied to the figures was a uniform cut of the face sizes to fit a Full-HD screen resolution.

3.2 Participants

Participants (N = 53, 26 male, 27 female; aged 18 – 63 years, M = 31.7) were volunteers among students and staff of the *removed for blind review* or visitors of the campus. All stated that they have no mental or physical illness. 21 participants were wearing glasses during the examination. 45 participants were German, 2 Chinese, 1 Indian, 1 Italian, 1 Mexican, 1 Pakistani, 1 Turkish, and 1 from USA. 21 participants claimed to have no experience with computer and video games. 16 played once per week, 12 several times a week, 4 daily. 3 participants never watched featured films, 15 watched one per week, 28 several times a week, and 7 watched them daily. There were 13 active in art or humanities, 11 engineers, 6 social scientists, 15 involved in nature sciences, and 8 non-academicians. 14 participants stated that they had heard the term 'Uncanny Valley' before. Before the eye tracking test, all participants declared their agreement to a data protection and privacy policy. After the test they were asked to fill out a questionnaire.

3.3 Procedure

Participants were seated upright on a fixed chair in front of a 21,7" LCD Monitor in a soundproof lab. Every procedure took about 20 minutes. At the beginning of the eye tracking test each participant was instructed about the procedure itself (calibration, first test, session). After 30 slides had been demonstrated, a participant took a break and received further instructions about the categorization task. The viewing distance was 60 cm. At the beginning of each experiment, a 12-point-matrix calibration was conducted with every participant. To avoid fixed gaze in the same position as on the previous slide, a black screen was shown between the stimuli. All images were presented for 10 seconds. After a random sequence in a group of 30 stimuli, a pause followed with an instruction slide providing information about the upcoming reaction measurement. A subject had to press a buzzer if he or she believed that a figural representation was *not a real human*. To prevent confusion between stimuli and the instruction slide, the test only continued after the participant made an initial input. 38 further stimuli of the categories ordinary human, wax sculptures, cosmetic surgery, and CGI were shown for 10 seconds. When a real figure appeared and the subject gave no response (did not press the buzzer), it meant the subject recognized a real figure correctly. When an artificial figure was shown and the subject gave no response, it meant that the artificial figure was perceived as a real human (and thus "passed" the test.). Benefit to this method was the possibility of getting precise measurements of the affective reaction with very low latencies when unrealistic figural representations were demonstrated. The eye tracking device was a Tobii X2-30 Compact Edition[1] with 30 Hz. Recording and playback of the slides was carried out with OGAMA 4.3 (Vosskühler et al. 2008) on an PC. After each recording, a backup of the eye tracking data was copied on an external drive via batch script. After every eye tracking session, a questionnaire was handed out. On a numeric rating scale from 1 to 10, participants were asked to state subjective value for realism, human-likeness, likeability, and attractiveness of each figure.

3.4 Analyses

The recorded raw data were aggregated to eye fixations. A fixation was calculated by the maximum distance of 20 pixels (0.45°) and the minimal number of at least 3 samples (100 ms) of successive gaze positions. Lost data through eye blinking or fixation outside the screen were discarded. The first fixation of a stimuli recording was not deleted. AOIs, events, and the fixation table were exported separately for further analysis. The analysis were performed using SPSS and Excel. To clearly delineate areas of interest, we used a template for polygonal boundaries shown in Figure 2**.** The AOI sizes had to be different due to the presented stimuli. However, we assume that we can make reliable statements about the proportion of attention within the specified AOI boundaries and with a relative calculation of facial fixations and dwell time. In order to be able to compare several samples with absolute fixation times, we provided identical AOI for 12 stimuli pairs – usually that of the same person or character

[1] Technical specifications of the eye tracker can be found here: http://www.tobii.com/en/eye tracking-research/global/products/hardware/tobii-x2-30-eye-tracker/

(cartoon vs. hyper-real, computer-generated vs. real, etc.). Fixations outside the predefined regions, in accordance with Figure 2**,** were treated separately (neck, background, out of display, for example). The AOI boundaries were invisible for the participants.

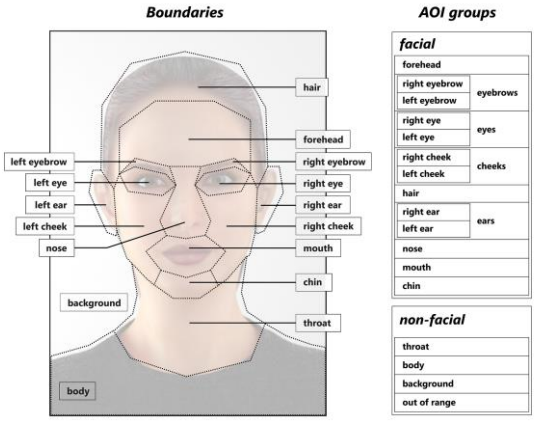

Figure 2: Boundaries of a predefined area of interest on a face. Divided and symmetrical areas were summarized into AOI groups of respective pairs (eyes, ears, cheeks, ears).

4 Results

4.1 Questionnaire

The average rating of the subjective assessment of the human-likeness results in a uniform grading of the categories. The results sorted by human-likeness: 1. Hyper-realistic cartoon ($M = 1.667$, $SD = 1.062$), 2. Cartoon ($M = 2.560$, $SD = 1.984$), 3. Cosmetic surgery ($M = 5.571$, $SD = 2.866$), 4. Robots ($M = 6.548$, $SD = 2.847$), 5. Post Mortem ($M = 6.976$, $SD = 2.540$), 6. CGI ($M = 7.857$, $SD = 1.905$), 7. Wax sculpture ($M = 8.786$, $SD = 1.337$), 8. Ordinary human ($M = 9.119$, $SD = 1.419$). The linear distribution on our scale indicates that the participants were able to differentiate between the stimuli. The average subjective ratings of human-likeness and realism per trial show a strong positive correlation between human-likeness and realism ($r = 0.966$, $p < 0.001$, $N = 64$, $CI_{-} = 0.011$, $CI_{+} = 0.008$). Both realism and human-likeness result in the same categorization sequence. Attractiveness and likeability also show a strong positive correlation ($r = 0.921$, $p < 0.001$, $N = 55$, $CI_{-} = 0.024$, $CI_{+} = 0.018$).

4.2 Response times and fixation sequences

To get reliable answers to the research questions and to facilitate the analysis of the eye tracking data in relation to the identification task, it was necessary to clarify whether significant changes in gaze behavior occurred during the reaction test. In this part of the eye tracking test, the subject was asked to press an input button in case the presented figure was *not a real*

human. We found no significant differences between the relative fixation times on a face between this states and therefore assume that there is no significant difference in the ratio of facial attention and dwell time between the tests ($p > 0.05$). The same analysis of fixations before and after the identification of a virtual figure was conducted within identification tests and led to similar results.

The relative dwell time on facial features varies between the stimuli categories. Figure 3 illustrates a decline of dwell time on the eye regions of figures. Depending on the category, attention was paid to the surrounding features instead (e.g. cheeks). A difference of 1.1 % between the dwell time on the eyes of ordinary people ($M = 35.19\%$, $SD = 21.79\%$) and CGI characters ($M = 33.29\%$, $SD = 20.23\%$) is relatively low and exceeds no general significance level, $t(1800) = 1.776$, $p = 0.076$). The largest and most significant difference in facial fixation time (15.98%) was recorded between the eye regions of ordinary people ($M = 35.19\%$, $SD = 21.79\%$) and post-mortem photographs ($M = 19.21\%$, $SD = 19.64\%$), with $t(846) = 8.417$, $p < 0.001$. Possible reasons for this differences are not further investigated in this paper, but could be caused due to mortality fears proposed by MacDorman (2005).

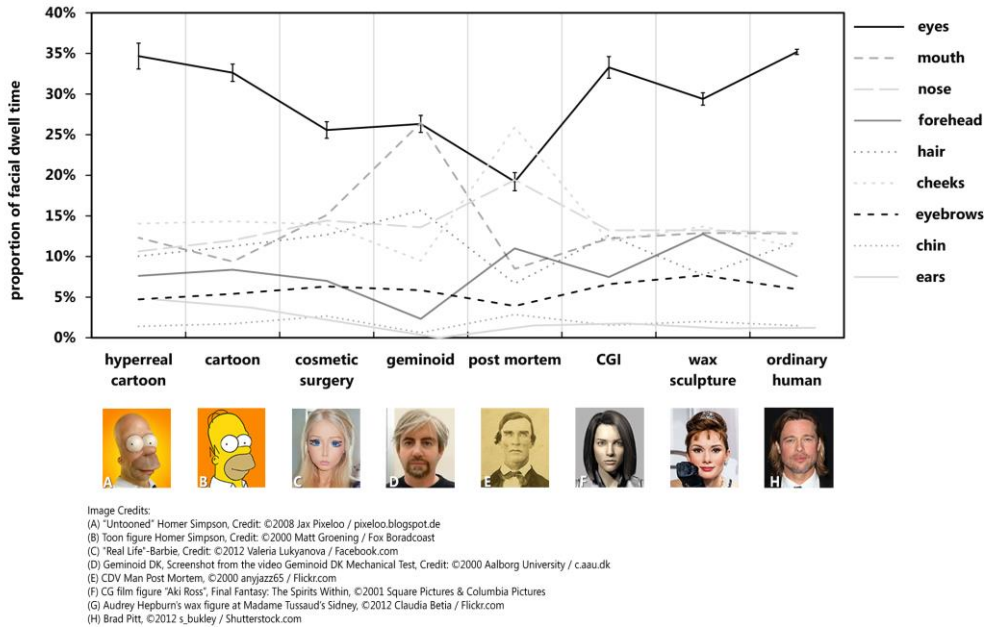

Image Credits:
(A) "Untooned" Homer Simpson, Credit: ©2008 Jax Pixeloo / pixeloo.blogspot.de
(B) Toon figure Homer Simpson, Credit: ©2000 Matt Groening / Fox Boradcoast
(C) "Real Life"-Barbie, Credit: ©2012 Valeria Lukyanova / Facebook.com
(D) Geminoid DK, Screenshot from the video Geminoid DK Mechanical Test, Credit: ©2000 Aalborg University / c.aau.dk
(E) CDV Man Post Mortem, ©2000 anyjazz65 / Flickr.com
(F) CG film figure "Aki Ross", Final Fantasy: The Spirits Within, ©2001 Square Pictures & Columbia Pictures
(G) Audrey Hepburn's wax figure at Madame Tussaud's Sidney, ©2012 Claudia Betia / Flickr.com
(H) Brad Pitt, ©2012 s_bukley / Shutterstock.com

Figure 3: Relative dwell time on facial regions. The curves show the ratios of the average dwell time on various facial regions. The error bars on the eyes curve show the standard error.

At the begining of a trial, the eye regions attract attention first (see Figure 4). Up to the 2nd and 3rd fixation, almost 45% of all cases with computer-generated or ordinary humans hit the eye regions. Also, nose and mouth are targeted increasingly at the expense of other features. For example, hair, chin, and cheeks play a subordinate role in the 1st fixations. After the initial visual contact (the 4th or 5th fixation), eyes fall in importance. At this point we can recognize

a difference in the gaze behavior depending on the category. The ambiguity provoked by the CGI faces results in fewer fixations on the eye region. In the case of artificial faces, the distribution of fixations moves to other features at the expense of the relative fixation count on the eye region. However, in trials with both CGI and real figures, the most attention was paid to the eyes.

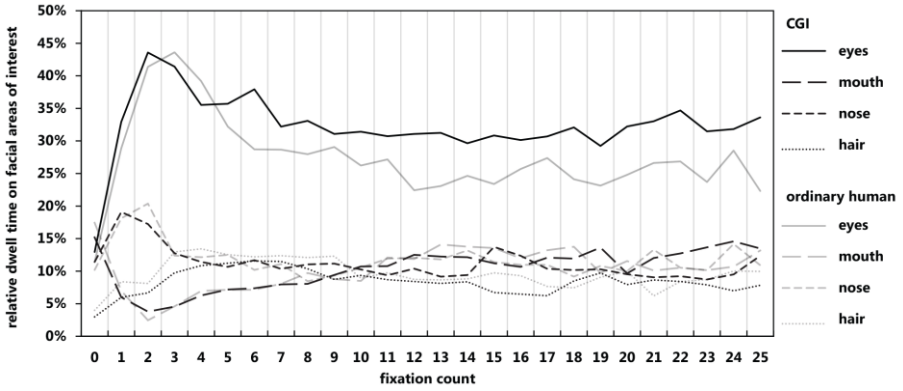

Figure 4: The diagram shows the distribution of attention in four selected categories: eyes, mouth, nose, and hair.

The upper graph in Figure 5 show the distribution of gazes on the eye region during the whole trial sequence within a time interval of 100 ms. Noticeably, a considerable increase of attention paid to this region is registered within the first 400 ms also with figures that differ from the human norm (cosmetic surgery). The increase reaches its peak after 400–600 ms. After that, the attention decreases and settles at a stable level after about 1200–1400 ms (the time of the 4^{th} and 5^{th} fixation) until the end of a trial. Fixations on the eye regions reach a stable level in a time range in which most decisions have been made. With minor deviations, this judgment generally applies to all categories, including photographs depicting ordinary humans. When a CGI figure was demonstrated, 39 participants responded within 800 to 1000 ms. 38 responded within 1000–1200 ms. A period of 400–600 ms elapsed between the highest amount of fixations on the eye regions of CGI figures and the average response time of the participants. Frequently, a certain amount of time passes between the first eye contact and the response. This suggests that eyes might be the most important feature for the assessment, but it does not clarify whether they are an explicit trigger for the reaction itself. It is possible that in the last fixations, other regions attract attention before a reaction occurs. To find strong differences between the respective regions, we conducted a one-way ANOVA of the relative count of fixations. No significant differences could be found between the 4 fixations [eye region: $F(3,208) = 1.477$, $p = 0.335$, all other: $p > 1.34$]. Only trials with 4 fixations before a reaction were included. The proportion of attention in the last 4 fixations before a response was given shows no significant changes. The immense importance of the eyes in decision making is illustrated by the following results: in 78.45% of all trials (1580 of 2014 samples), participants fixed their gaze on the eye region at least once before they made a decision. If we consider

areas adjacent to the eye region (eyebrows, cheeks, nose), these were 88.98% (1792 of 2015 samples). The mouth region was fixated in 59.15% and the nose in 69.98% of all cases at least once before the decision was made (hair: 47.61%, forehead: 46.05%, chin: 18.18%, ears: 11.12%, respectively).

The meaning of eye region becomes also particularly evident when counting facial locations during the last fixation before a response follows. In 37.43% of all cases (280 of 748) the last fixation was on the eye region when a response follows. Other AOIs were fixed significantly less (nose: 14.44 %, mouth: 11.23%, cheeks: 12.30%, hair: 9.89%, eyebrows: 6.95%, forehead: 5.21%, chin: 1.74%, ears: 0.80%). Thus we can conclude that eye tracking data prove eye regions to be the most important criterion in the decision making process. Eyes are also the most important facial feature regarding the proportion of attention when artificial figures are observed. Other facial features may also play a considerable role in unmasking and subconscious recognizing of artificial figures, but in most cases no decision is made without fixing a gaze on the eyes. It is clear that without eye contact, hardly any decision is made.

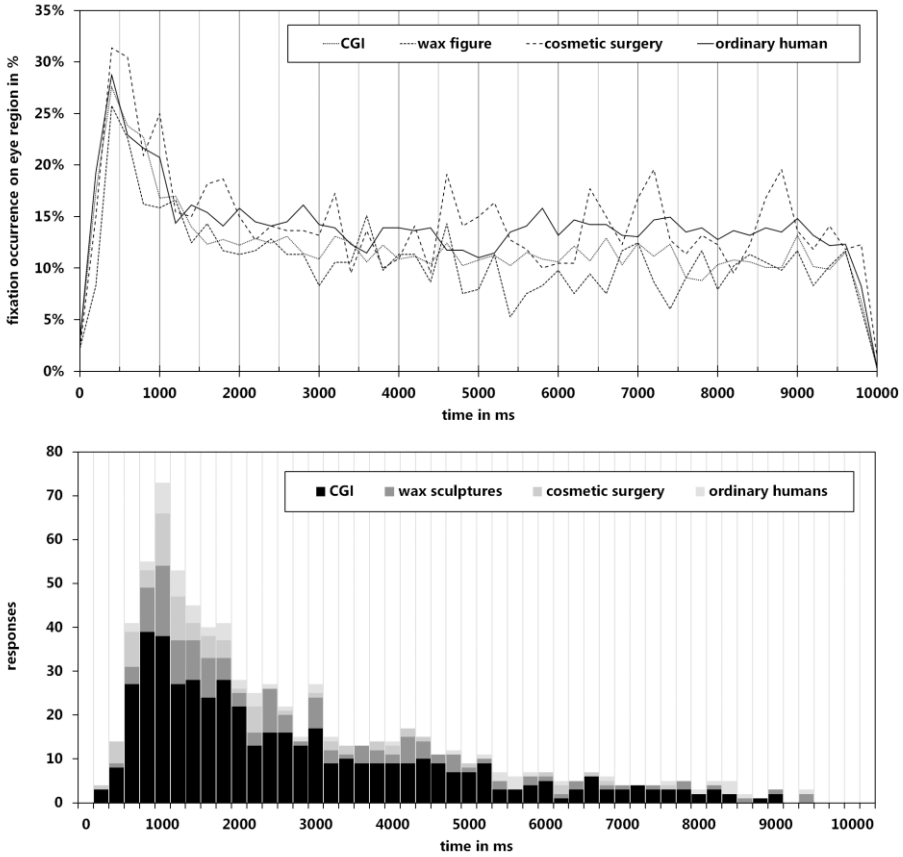

Figure 5: The upper diagram shows the relative proportions of facial fixations on the eye areas. The lower shows the frequency and timing of responses, whether a face is rated as unreal.

4.3 Correlations

Pairs of average response times and the mean of the respective subjective evaluation per trial revealed a strong positive correlation between the duration of a response and the specified realism of a figure ($r = 0.740$, $p < 0.001$, $N = 55$, $CI_{-} = 0.070$, $CI_{+} = 0.054$). In contrast, reaction time and human-likeness correlate moderately ($r = 0.673$, $p < 0.001$, $N = 55$, $CI_{-} = 0.083$, $CI_{+} = 0.066$). The later the response was given, the more realistic or human-like an image was rated. The perceived attractiveness of figures shows a moderate correlation with the verification period ($r = 0.435$, $p = 0.001$, $N = 55$, $CI– = 0.118$, $CI+ = 0.102$). The moderate correlation between likeability and the fixation time on the eye region ($r = 0.418$, $p = 0.002$, $N = 55$, $CI- = 0.12$, $CI+ = 0.104$) is the most strongest and significant correlation between the fixation times on facial areas and subjective ratings.

5 Discussion

This paper shows that eye contact is an essential part in gaze behavior and distinguishing between real and artificial humans. The analyses of the gaze sequences show that faces in the major part of trials have not been evaluated, as long as the eyes were not considered also with faces that differ from the human norm. Correlations analyses show, that the more realistic a figure is, the longer participants need to identify them. We found further correlations between subjective ratings of faces and the duration of gaze fixations on the facial area of the eyes. The shorter the eye contact, the more negative the subjective rating. We therefore assume, that acceptance and ways of interaction with human-like artificial figures could be particularly improved by more credible eyes and eye related areas. It is important to note that behavior of the first fixations does not significantly differ from normal and task-driven gaze behavior. An interesting hypothesis arises: gaze behavior does not change due to the task because it continuously runs. We assume that usual human shape is continuously assessed by the same perception process that also differ artificial faces from real ones. Obviously, this kind of process runs parallel to the recognition pattern which is only activated for objects when something is classified as autonomous. This visual process in face recognition has to be investigated in conducting further studies for a better understanding of the Uncanny Valley.

Acknowledgements

This work was supported by the cooperative graduate program ‘Digital Media’ of the University of Stuttgart, University of Tübingen, and the Stuttgart Media University (HdM).

Contact

Valentin Schwind, Solveigh Jäger
VIS, Universität Stuttgart / IFG, Stuttgart Media University
Pfaffenwaldring 5a, 70569 Stuttgart, Germany / Nobelstr. 10, 70569 Stuttgart, Germany
valentin.schwind@vis.uni-stuttgart.de / {schwindv,sj034}@hdm-stuttgart.de

References

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. In *Cognition*, *38*, p. 109–144.

Cheetham, M. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. In *Frontiers in Human Neuroscience*, *5*, p. 126.

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. In *Frontiers in Psychology*, *4*, p. 108.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? In *Psychological Review*, *105* (3), p. 482–498.

Green, R. D., MacDorman, K. F., Ho, C. C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. In *Computers in Human Behavior*, *24* (5), p. 2456–2474.

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. In *Neuropsychologia*, *45* (5), p. 1019–1028.

Janik, S. W., Wellens, a R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. In *Perceptual and Motor Skills*, *47*, p. 857–858.

Just, M. a, & Carpenter, P. a. (1980). A theory of reading: from eye fixations to comprehension. In *Psychological Review*, *87* (4), p. 329–354.

Langten, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. In *Trends in Cognitive Sciences*, *4* (2), p. 50–59.

Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy. How, when, and where we perceive life in a face. In *Psychological Science : A Journal of the American Psychological Society / APS*, *21* (12), p. 1854–1862.

MacDorman, K. F. (2005). Mortality salience and the uncanny valley. In *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots*, p. 399–405.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, p. 26–29.

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. In *Computers in Human Behavior*, *25* (3), p. 695–710.

Mori, M. (1970). The Uncanny Valley. In *Energy*, *7* (4), p. 33–35.

Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. In *Visual Cognition*, *15* (3), p. 322–348.

Vinette, C., Gosselin, F., & Schyns, P. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. In *Cognitive Science*, *28* (2), p. 289–301.

Vosskühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. In *Behavior Research Methods*, *40* (4), p. 1150–1162.

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. In *Japanese Psychological Research*, *55*, p. 20–32.