

Research Article

Valentin Schwind*, Solveigh Jäger

The Uncanny Valley and the Importance of Eye Contact

DOI 10.1515/icom-2016-0001

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.

Keywords: Uncanny Valley, Eye Tracking, Human-likeness, Subjective Ratings, Eye Contact

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 [29] 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, and 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. [1, 24, 25]. Mori also clarified the role of motion: the effect is thought to be stronger if figures are in motion or interacting. Robotic engineers, puppet designers or computer animators seek to improve the subjective appearance and behavior of their figures in an intended way. In uncanny research the phenomenon and its impact has produced inconsistent findings so far [2, 5, 6, 12, 32, 35, 36] 1970. The objective 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.

A common approach to studying human visual perception of digital imaginary figures is recording eye movement using eye tracking devices. 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 [7, 19, 33]. Polygonal areas of interest (AOI) provide important information on the number and duration of fixations within predefined areas on two-dimensional stimuli. The disadvantage of this method is that variations can arise due to different patterns of AOI boundaries and thus possibly lead to inconsistent results. For this reason some studies reject such templates completely and focus on the differences between the stimuli [37]. It is noteworthy to mention that this method only works for figures with identical proportions of facial features. A wide range of stylized, drawn, uncanny, and real faces suggested by Mori's graph would not allow drawing direct comparison of gaze sequences.

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

*Corresponding author: Valentin Schwind, HCI Laboratory, Institute for Visualization and Interactive Systems (VIS), Department of Computer Science, University of Stuttgart, Stuttgart, Germany, e-mail: valentin.schwind@vis.uni-stuttgart.de

Solveigh Jäger, Games Laboratory, Institute for Games, Department of Computer Science, Stuttgart Media University, Stuttgart, Germany

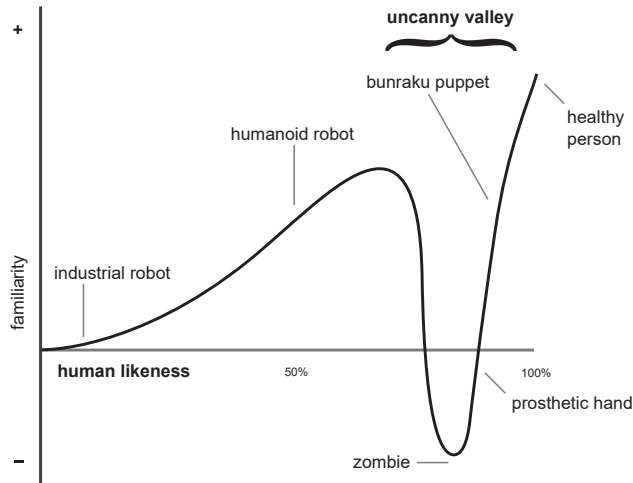


Figure 1: Simplified diagram (illustration from MacDorman 2005) of Mori's graph shows the generally positive feelings of an observer towards artificial entities with a certain degree of human-likeness.

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 [19].

2 Related Work

Approaches to define the dimension of human likeness (DHL) with a linear morph continuo [5, 6] 1970 lead to prolonged response times during a categorization task with ambiguous stimuli. Regardless of whether an object is human-like or not, difficulties in categorization lead to negative ratings or feelings [40]. Green et al. [13] 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 [4]. Average faces generally perceived as attractive can hardly be compared with faces of less human-like characters. For example, cartoon-faces can show a high acceptance rate whereas computer-generated faces with even a higher similarity to an average human face are often perceived as uncanny. This study highlights

the differences in visual perception of different types of faces (i. e. ambiguous and clearly attributable). These also include drawn, stylized, artificial, and lifelike faces.

Mori associated the phenomenon with our aversion towards corpses. MacDorman [23] argues that eerie anthropomorphic figures trigger the subconscious fear of death. This is a highly discussed topic regarding the Uncanny Valley theory [16, 28, 36]. A study on monkeys shows that this phenomenon also occurs in primates [34]. This leads to the assumption that there is an evolution-related cognitive connection between the uncanny valley and the fear of death. For this reason here, we include pictures that depict deceased persons standing upright with opened eyes. This kind of post-mortem photograph was particularly common in the 19th century [26], and the practice lives on in some cultures today. Although these photographs are supposed to create an impression of a lifetime snapshot, they produce a particularly gruesome effect on the viewer, as our study shows.

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 [15, 17, 18, 20, 31, 38]. Farah et al. [9–11] 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 [17, 18]. Previous studies on artificial figures confirm this assumption [6, 12, 22, 25].

Therefore, main focus of this study is dedicated particularly to fixations on the eye regions and different gaze patterns on faces of different artificial human representations.

3 Materials 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. There were 11 pairs of images with the same or a very similar person and the same posture. 4 images show a pair: a person and his or her double. All depicted figures have a neutral facial expression and an upright posture.

Computer-generated faces can be recognized due to junctions or contours. Cheetham et al. [6] 1970 mentioned that in addition to the mouth, eyes, and nose, other areas such as hair, hairline, and head contours may be of interest. For this reason we masked faces with no oval overlay and made no additional changes to the back- and foreground. 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 our University or visitors on 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.

To get a profile of the individual view toward artificial figures, we asked the participants whether it makes sense for them to build or to simulate human-like figures. 16 gave their consent. 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. One far-sighted participant wearing glasses could not perform the test because of too strong diopter values (+12, r) that could not be calibrated by the device.

3.3 Procedure

Participants were seated upright on a fixed chair in front of a 21,7” LCD Monitor in a soundproof lab. Non-reflective whiteboards were used to prevent reflections of infrared light. The background brightness was 320 lux. 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. The subjects could move their eyes freely. After a random sequence in a group of 30 stimuli, a pause followed with an instruction slide providing information about the upcoming reaction measurement. The subject had to press the buzzer if he or she believed that a figural representation was *not a real human*. Two participants asked what *real* means. We repeated the task orally and added the note, that some human representations are *artificially produced*. 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.)

With this experimental scheme, it was not possible to measure response time when real human figures were displayed, since in this case participants simply had to wait until a trial finished. The 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. A multiple selection during a complex categorization task could have led to delays and long considerations.

The last task in this experiment was a categorization task with 4 final images. On every image there was a real human and the same person with a digital double or a doppelgänger android. The subject had to press the left or right keyboard button to identify the “fake.” This part of the experiment was not included in the regular AOI calculations or other fixation measurements of the previous stimuli. This was the final part of the eye tracking experiment.

The eye tracking device was a Tobii X2-30 Compact Edition¹ with 30 Hz. Recording and playback of the slides was carried out with OGAMA 4.3 [39] on an Intel i7 3635QM with 2,4 GHz and 32 GB RAM. After each recording, a

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

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. Then the participants were asked to indicate whether they had already known the figure before. To specify additional characteristics of the presented figures, participants could freely choose between 15 positive and 15 negative attributes randomly assorted in a multiple-choice matrix.

3.4 Analysis

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

and calculations of AOI hits were performed using SPSS V. 21 and Excel 2013. To clearly delineate areas of interest, we used a template for polygonal boundaries shown in Figure 2. The template resulted from a preliminary study on the accuracy and precision of the eye tracking device. In order to understand the significance of facial features in the processing of eerie faces, we decided to use a common facemask and thus apply similar boundaries of polygonal regions of interest on each face.

It is important to note that 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 (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.

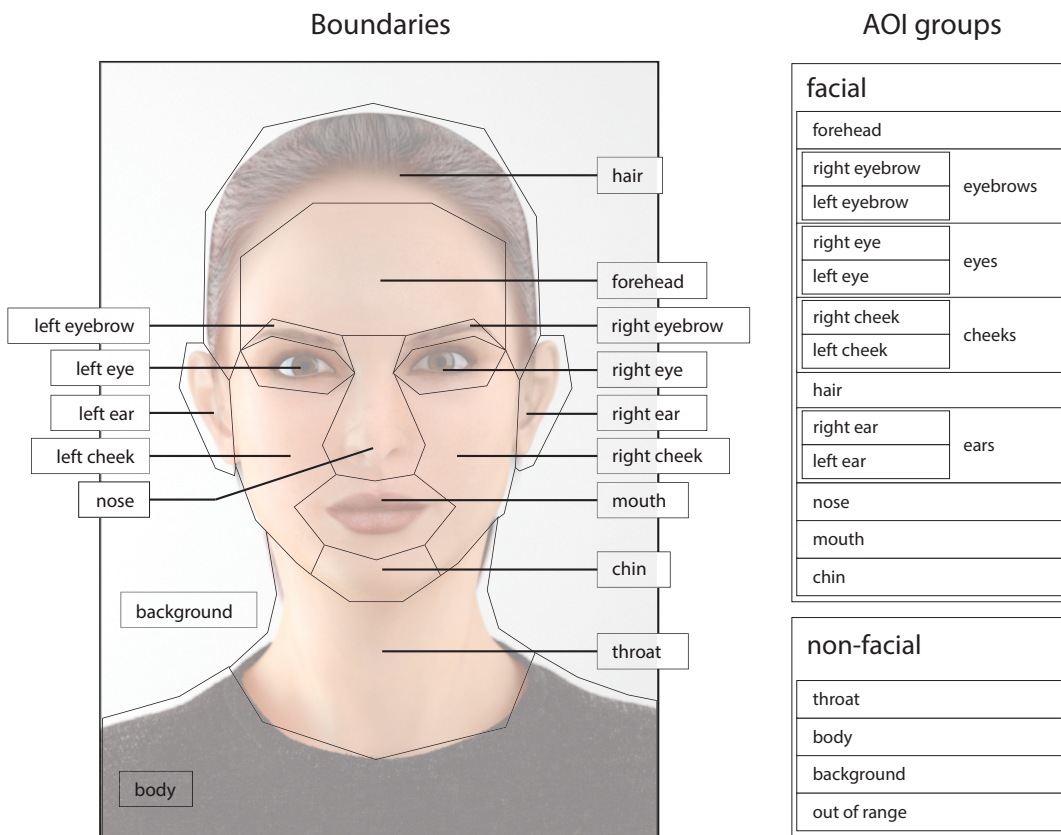


Figure 2: Boundaries of a predefined area of interest on a face. Divided and symmetrical areas were summarized into AOI groups of respective pairs (eyes, ears, cheeks, ears).

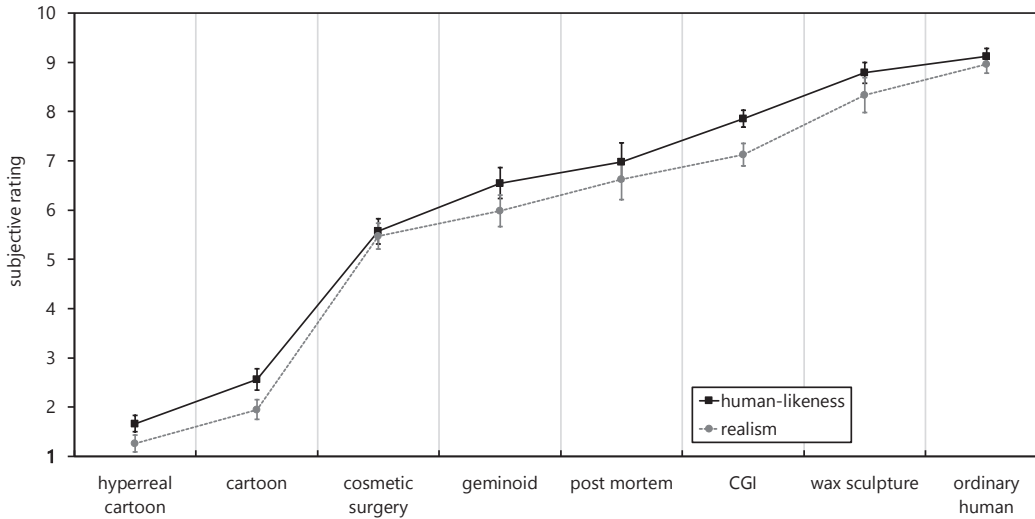


Figure 3: Human-likeness sorted by increasing subjective assessments of the participants.

4 Results

4.1 Questionnaire

Subjective assessments design the sorting of the categories in graphs and descriptions of the human-like axis in further diagrams of the study. The resulting ratings follow the order of 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 sorting of human-likeness organized the subsequent evaluation of the categories without a linear morpho continuum. 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 (see Figure 3).

Attractiveness and likeability also show a strong positive correlation ($r = 0.921$, $p < 0.001$, $N = 55$, $CI_- = 0.024$, $CI_+ = 0.018$). Ratings show differences between familiar and unfamiliar faces, especially in case of cartoon and hyper-realistic cartoon figures. Cartoon figures were rated more positively when they were familiar to the observers. An independent t-test shows a significant difference in the ratings of known ($M = 8.200$, $SD = 2.203$) and unknown ($M = 4.562$, $SD = 2.257$) cartoon figures: $t(80) = 7.233$, $p < 0.001$.

The examination of a multiple choice matrix with 30 different characteristics as a supplement of the subjective scale ratings provides additional insights: at first glance, the participants were certain about the category they could attribute to the observed figural representations. Deceased, for example, were categorized as artificial by only 6.98% of participants and as ordinary humans by 5.81%. In contrast, cartoon characters (46.51%), hyper-real cartoons (53.49%), and humans with visible impacts of cosmetic surgery (51.16%) were all often rated as artificial. Participants rated certain images as repulsive: hyper-real faces (30.23%), images in the cosmetic surgery group (34.11%), and the deceased (60.47%). Interestingly, 44.19% of the participants assessed neutral faces of the dead as “aggressive.” The same holds true for hyper-real cartoons (34.88%) and for cosmetic surgeries (27.91%). This possibly indicates that the deceased were immediately perceived as a threat despite their neutral facial expression. This phenomenon is illuminated in detail in the analysis of the eye tracking data.

4.2 Response Times and Fixation Sequences

To answer our research questions and to facilitate the analysis of the eye tracking data, 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 differences between the relative fixation

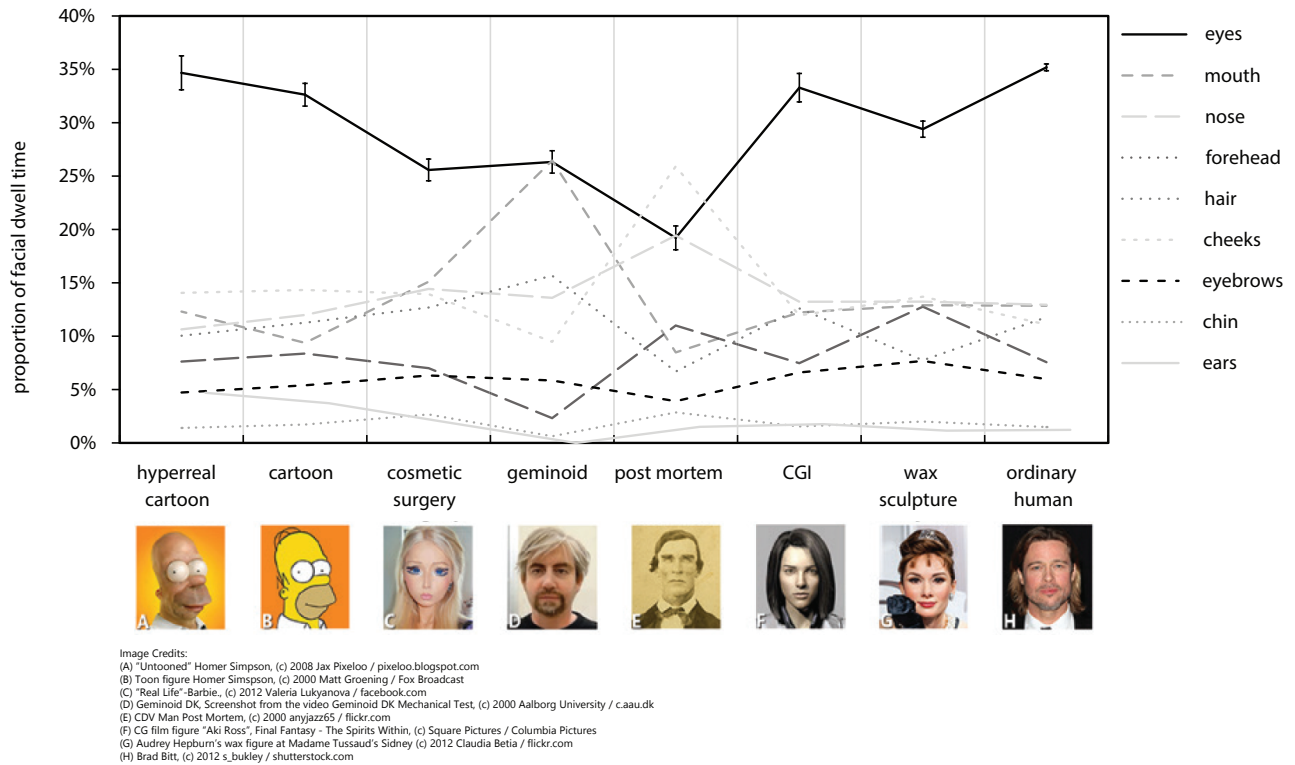


Figure 4: Relative dwell time on facial regions. The curves show the ratios of the average dwell time on various facial regions. The error bars on the eyes curve show the standard error.

times on a face. An equivalent test confirmed this result: we performed a paired t-test, which showed no significant differences within the relative distribution between the dwell time in rated and unrated trials ($p > 0.05$), except for the forehead region ($p = 0.007$). The same analysis of fixations before and after the identification of a virtual figure was conducted within identification tests. This analysis provided similar results. During the reaction test, no significant differences between the ratios of facial attention were recorded.

To examine whether the task had an impact on the overall number and duration of fixations, we considered the average number of fixations and duration before and during a reaction measurement. An independent t-test showed that the average duration of a fixation before ($M = 354.87$, $SD = 370.72$) and during a reaction test ($M = 350.49$, $SD = 358.53$) differed with a mean of 4.389 ms and was minimally elevated (higher attention) with no significant difference ($\alpha = 0.05$) between the two $t(77032) = 1.716$, $p = 0.086$. We assume that the reaction test had apparently no significant effect on the number and duration of fixations. Since the subjects were not informed that a check for fakes of the figures will be part of the test, we can exclude the possibility that this was of

any influence on their gaze behavior. However, we can assume that participants knew that it was part of the experiment. Such conscious analysis of artificial figures may take place in reality as well. We therefore suspect that subjects searched for the truth of whether a character was genuine or not, and thus displayed no concrete changes in gaze behavior.

However, the relative dwell time on facial features varies between the categories. Figure 4 illustrates a significant decline of fixations and dwell time on the eye regions of figures that were rated as less attractive and had an average human-likeness. Depending on the category, attention was paid to the surrounding features instead. In figures after cosmetic surgery, fixations shifted to other areas of the face, whereas in geminoids the attention tends to be given to the mouth region. 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 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 ($M = 19.21\%$, $SD = 19.64\%$), $t(846) = 8.417$,

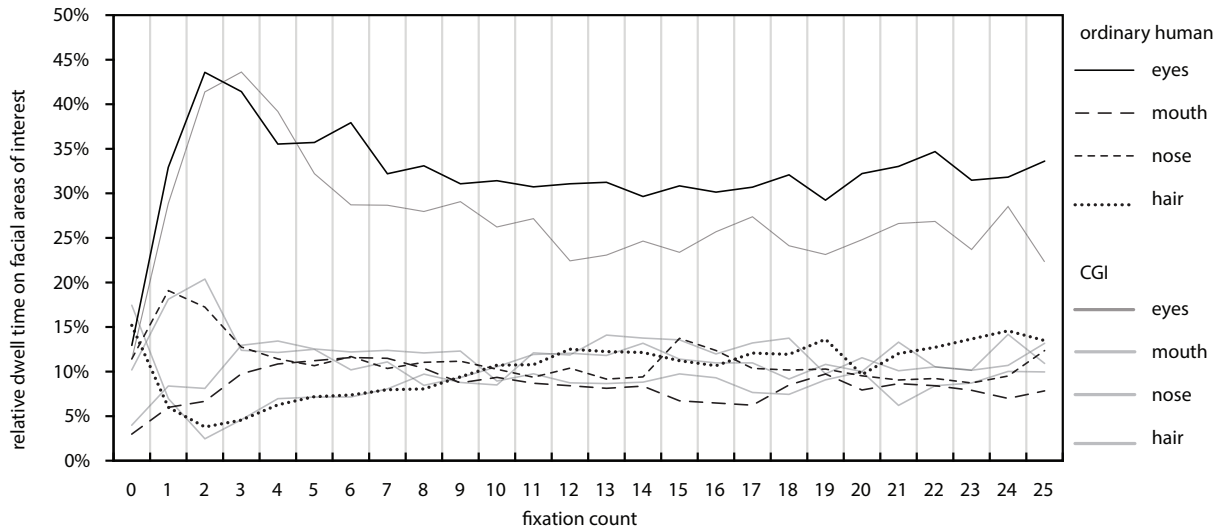


Figure 5: AOI hits of the first 25 fixations. The diagram shows the distribution of attention in four selected categories: eyes, mouth, nose, and hair. In the 2nd fixation (3) after the emergence of the image (1), almost 40 % of all fixations landed on the eye region.

$p < 0.001$. The subjects looked at the underlying cheek region instead. Similar observation regarding menacing facial expressions is done by Moss et al. [27]. Their study pointed particularly to differences between genders. A one-way ANOVA of all fixations within our study shows no significant differences between the occurrences of male and female fixations exceptionally for mouth ($M_f = 0.101$, $M_m = 0.084$ [$F(1,51) = 5871$, $p = 0.190$]) and background regions ($M_f = 0.130$, $M_m = 0.164$ [$F(1,51) = 8585$, $p = 0.005$]). That indicates that women fixated more frequently on mouth regions and men on background structures. As mentioned above, images of the dead were often rated in the questionnaire as aggressive and repulsive. No significant difference (0.19 %) was observed between the average facial fixation time in the case of ordinary humans and CGI figures ($M = 33.38\%$, $SD = 20.32\%$), $t(256,630) = 1.757$, $p = 0.076$. The analysis indicates that the most and longest fixations were recorded in the case of photographs of ordinary humans. Our evaluation demonstrates that eyes receive the most attention in all categories of human-like figures.

Figure 5 shows the relative dwell time on the four most important facial areas of interest during the first 25 fixations. At the very beginning of the trial, the eye regions attract attention like a magnet. The fixation sequences in all categories prove that. Up to the 2nd and 3rd fixation, almost 45 % of all cases 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 relative attention paid to the eye regions of CGI figures is slightly lower than that of photos showing ordinary people. This explains a small but measurable difference in the dwell times between the two categories in Figure 5. The difference is particularly obvious between the eye regions and can be explained by the fact that other facial regions of CGI figures need to be examined more than real people. The ambiguity provoked by the CGI faces results in fewer fixations on the eye region and thus additional facial features had to be considered to make a decision. The proportion of attention paid to these additional characteristics differs depending on stimuli (see “Direct Comparisons”). 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.

The graphs in Figure 6 show the occurrence of gazes resting on the eye region during the whole trial sequence within a time interval of 100 ms. The curves illustrate the relative occurrence of gazes focusing on the eye region of CGI figures, cosmetic surgical patients, wax sculptures, and ordinary humans. Noticeably, a considerable increase of attention paid to this region is registered within the first 400 ms. 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 4th and 5th 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 depending on the degree of realism, this judgment generally applies to all categories, including photographs depicting ordinary humans.

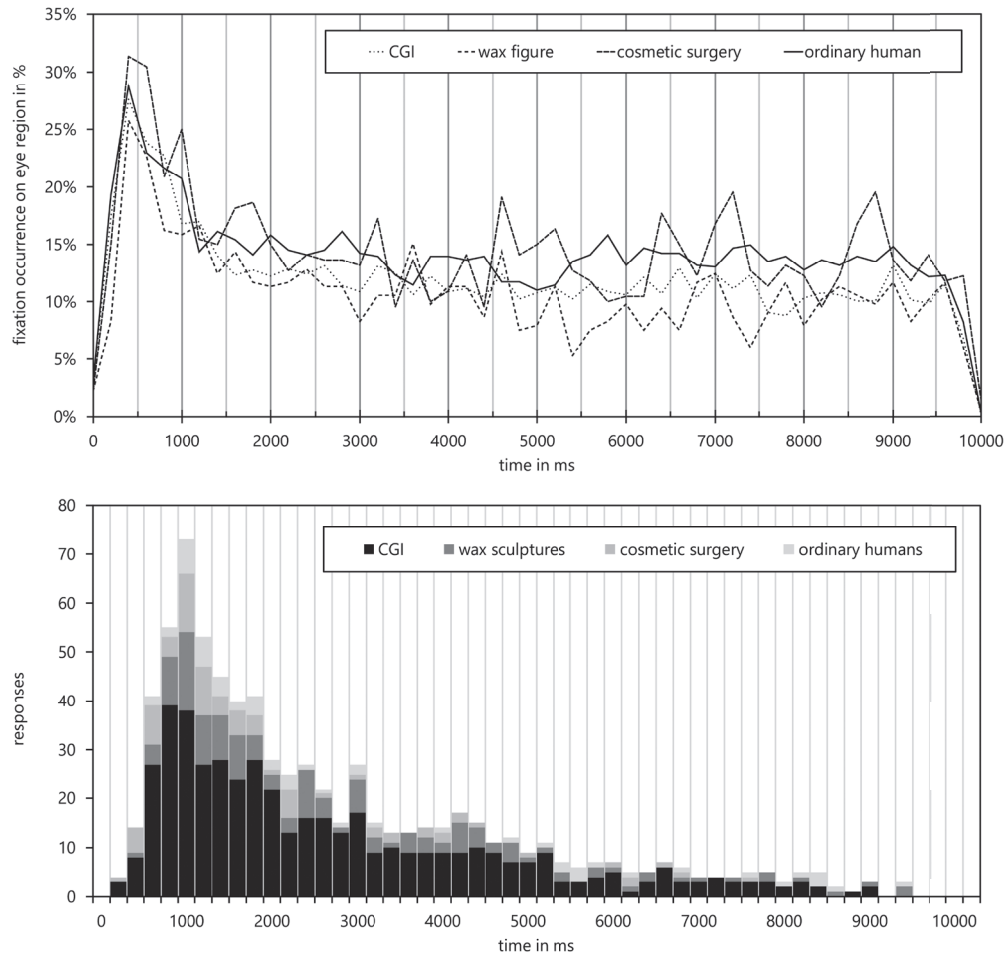


Figure 6: Temporal gaze sequence. The diagram shows the relative occurrence of fixations on the eye regions.

When a CGI figure was demonstrated, 39 participants responded within 800 to 1000 ms. 38 responded within 1000–1200 ms. When wax sculptures (16) and cosmetic surgery patients (12) were shown, most responses also followed 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. As mentioned above, some real faces of ordinary humans and cosmetic surgery patients were rated as unreal. Overall, 87% of responses in the case of images of ordinary humans and only 36% of responses in the case of cosmetic surgery patients got a correct rating. Wax figures ended up with 50%, whereas CGI figures with 66%. This explains the results of the questionnaire, according to which CGI figures were considered unrealistic and less human-like than the wax figures: they were recognized more often. The high error rate in the case of images of faces extremely altered by cosmetic surgery is obvious due to their strong deviation from a regular human face.

A certain amount of time passes between the first eye contact and the response. This suggests that eyes might be a very 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 importance of eye region becomes particularly evident when we count 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. 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.

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. Attractiveness correlates weakly with the response time ($r = 0.383$, $p = 0.004$, $N = 55$, $CI_- = 0.123$, $CI_+ = 0.108$). The moderate correlation between likeability and the fixation time on the eye region ($r = 0.435$, $p = 0.001$, $N = 55$, $CI_- = 0.118$, $CI_+ = 0.102$) is the strongest correlation between the fixation times on facial areas and subjective ratings. Other significant correlations ($p < 0.05$, $r > 0.2$) between response times and subjective ratings or fixation time on facial regions could not be found.

4.4 Direct Comparison

Calculating direct differences in gaze behavior, we could identify some treacherous areas. Compared to previously obtained findings, these were revealed less often. Our analysis shows differences in facial preferences of the observers before, during, and after the verification process. Local shifts of attention and absolute dwell time can be determined with a specific template (e. g. CGI or photo). When we draw a direct comparison between the attention areas in the case of an ordinary human image and a CGI image, we notice a significant shift of attention within facial areas when unrealistic features are observed. For example, Figure 7 illustrates the shift of attention towards the lack

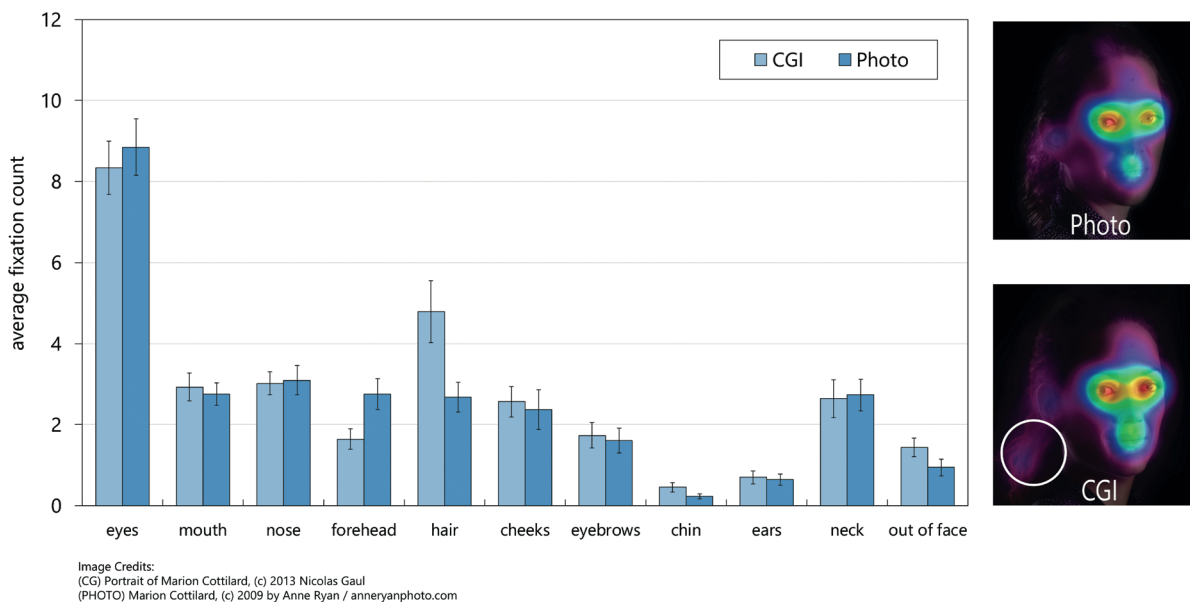


Figure 7: Direct comparison of the same person in CGI and on a photo. The diagram illustrates the relative dwell time spent on individual AOIs in direct comparison.

of glossiness in the backscattering of the hair. Noteworthy in this case is that there are fewer fixations on the forehead. This direct comparison makes clear that difficulties in the process of categorization can cause higher dwell time at the expense of other external features. In the considered case, there is a significant effect between fixations on the hair and forehead. A one-way ANOVA between the CGI representation and the photo shows a significant effect on the hair [$F(1,104) = 6.134$, $p = 0.015$] and the forehead [$F(1,104) = 5.929$, $p = 0.026$].

The direct comparison of a real person and his or her artificial double in the same image (three geminoids and one CGI character) on the final 4 stimuli show that higher attention was paid to the part of the image that displayed a non-human figure. The images were placed left and right, with the boundary in the middle. The subject had to choose the artificial figure using the keyboard cursors. In this task the images of real humans got significantly lower attention rates than their artificial counterparts. The dwell time on the doubles ($M = 3232.89$, $SD = 1401.73$) was significantly higher than that on humans ($M = 2657.27$, $SD = 1252.65$) with $t(212) = 5.171$, $p < 0.001$. The fixation count on doubles ($M = 14.773$, $SD = 5.865$) was significantly higher than on the human counterpart ($M = 14.773$, $SD = 5.865$) with $t(212) = 5.792$, $p < 0.001$. This indicates that this particular categorization task led to higher fixation times on the side of the image where a non-human was depicted (the other side showed an image of a real human). The average response time was also higher ($M = 3855.47$, $SD = 2619.32$) compared to the previous single face trials ($M = 2717.89$, $SD = 2040.52$), $t(267,777) = 5.924$, $p < 0.001$. The average rate of correct answers was higher (83.01%) than in the other trials (60.49%), without a highly significant effect $t(18) = 1.743$, $p = 0.098$.

5 Discussion

This paper shows that the visual search of eye regions 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. Our results show that this recognition occurred between 800–1200 ms after the appearance of a stimulus and at 400–800 ms after the first eye contact. A large number of subjects take its decision only when they draw attention to the eyes of a figure. Therefore, eye contact plays the same role both in decision making

and in the judgment of whether figures are artificial or real. 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.

We suppose, that future research should investigate which kind of fear human-like entities actually provoke after first eye contact and how short and long term interaction with robots or avatars changes. A further question is whether abnormality as well as ambiguity arise the same eerie feeling after eye contact. Especially the use of animated and interacting artificial characters should be the topic of further research.

Acknowledgement: Thanks go to the artists, photographers, and researchers that have submitted their work for this research article. We thank the German Research Foundation (DFG) for financial support within project CO4 of SFB/Transregio 161. 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).

References

- [1] Alejandro Lopez Hernandez, J. (2010). *User Centric Media*. (P. Daras, O. M. Ibarra, O. Akan, P. Bellavista, J. Cao, F. Dressler, ... G. Coulson, Eds.) *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* (Vol. 40). Springer Berlin Heidelberg.
- [2] Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelgänger – A critical look at the Uncanny Valley.

- In *Proceedings – IEEE International Workshop on Robot and Human Interactive Communication* (Vol. 31, pp. 269–276). IEEE.
- [3] Bovet, D., & Vauclair, J. (2000, May). Picture recognition in animals and humans. In *Behavioural Brain Research*.
- [4] Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. In *Cognition*, 38, p. 109–144.
- [5] 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.
- [6] 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.
- [7] Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer.
- [8] Emery, N. J. (2000, August). The eyes have it: The neuroethology, function and evolution of social gaze. In *Neuroscience and Biobehavioral Reviews*.
- [9] Farah, M. J. (1996, April). Is face recognition “special”? Evidence from neuropsychology. In *Behavioural Brain Research*.
- [10] 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.
- [11] Farah, M. J., Wilson, K. D., Maxwell Drain, H., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. In *Vision Research*, 35 (14), p. 2089–2093.
- [12] Flach, L. M., Dill, V., Hocevar, R., Lykawka, C., Musse, S. R., & Pinho, M. S. (2012). Evaluation of the uncanny valley in CG characters. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7502 LNAI, pp. 511–513).
- [13] 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.
- [14] Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS / CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, p. 39–42.
- [15] Haxby, J. V., Hoffman, E. a., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. In *Biological Psychiatry*, 51 (1), p. 59–67.
- [16] Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. D. D. (2008). Human emotion and the uncanny valley. In *Proceedings of the 3rd international conference on Human robot interaction – HRI '08* (Vol. 1, p. 169).
- [17] 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.
- [18] 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.
- [19] Just, M. a., & Carpenter, P. a. (1980). A theory of reading: from eye fixations to comprehension. In *Psychological Review*, 87 (4), p. 329–354.
- [20] 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.
- [21] Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. In *Journal of Comparative Psychology*, 124 (3), p. 233–251.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] McDowell, D. E., & Ruby, J. (1997). *Secure the Shadow: Death and Photography in America. Contemporary Sociology* (Vol. 26). MIT Press.
- [27] Mercer Moss, F. J., Baddeley, R., & Canagarajah, N. (2012). Eye Movements to Natural Images as a Function of Sex and Personality. In *PLoS ONE*, 7 (11), p. e47870.
- [28] Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. In *Minds and Machines*, 19 (3), p. 345–359.
- [29] Mori, M. (1970). The Uncanny Valley. In *Energy*, 7 (4), p. 33–35.
- [30] Parr, L. a. (2011). The evolution of face processing in primates. In *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366 (1571), p. 1764–1777.
- [31] Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. In *Visual Cognition*, 15 (3), p. 322–348.
- [32] Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. In *Presence: Teleoperators and Virtual Environments*, 16 (4), p. 337–351.
- [33] Shebilske, W., & Fisher, D. (1983). Understanding extended discourse through the eyes: How and why. In *Eye Movements and Psychological Functions: International Views*, p. 303–314.
- [34] Steckenfinger, S. a., & Ghazanfar, A. a. (2009). Monkey visual behavior falls into the uncanny valley. In *Proceedings of the National Academy of Sciences of the United States of America*, 106 (43), p. 18362–6.
- [35] Tinwell, A., Grimshaw, M., Nabi, D. A., Williams, A., Abdel Nabi, D., Angela, T., & Mark, G. (2011). Uncanny Valley in Virtual Characters 1 Facial Expression of Emotion and Perception of the Uncanny Valley in Virtual Characters Angela Tinwell. In *Computers in Human Behavior*, 44 (0), p. 1–34.
- [36] Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. In *Journal of Gaming & Virtual Worlds*, 2 (1), p. 3–25.
- [37] van Belle, G., Ramon, M., Lefèvre, P., & Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. In *Frontiers in Psychology*, 1 (June), p. 20.

- [38] 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.
- [39] 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.
- [40] 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.



Solveigh Jäger
Stuttgart Media University, Germany
solveigh.jaeger@googlemail.com

Solveigh Jäger is game designer, 3D artist, and media author. She studies Electronic Media Master and works on research and development projects in the field of assistive and barrier-free IT.

Bionotes



Valentin Schwind
University of Stuttgart, Germany
valentin.schwind@vis.uni-stuttgart.de

Studied Media Computer Science at the Stuttgart Media University in Germany. Is now doctoral student at the University of Stuttgart and explores the Uncanny Valley in Human-Computer Interaction.