Determining the Characteristics of Preferred Virtual Faces Using an Avatar Generator

Valentin Schwind[a,b], **Katrin Wolf**[b], **Niels Henze**[b], **Oliver Korn**[b]
[a]Stuttgart Media University
Stuttgart, Germany
schwindv@hdm-stuttgart.de
[b]VIS, University of Stuttgart
Stuttgart, Germany
{firstname.lastname}@vis.uni-stuttgart.de

ABSTRACT
Video game developers continuously increase the degree of details and realism in games to create more human-like characters. But increasing the human-likeness becomes a problem in regard to the Uncanny Valley phenomenon that predicts negative feelings of people towards artificial entities. We developed an avatar creation system to examine preferences towards parametrized faces and explore in regard to the Uncanny Valley phenomenon how people design faces that they like or reject. Based on the 3D model of the Caucasian average face, 420 participants generate 1341 faces of positively and negatively associated concepts of both gender. The results show that some characteristics associated with the Uncanny Valley are used to create villains or repulsive faces. Heroic faces get attractive features but are rarely and little stylized. A voluntarily designed face is very similar to the heroine. This indicates that there is a tendency of users to design feminine and attractive but still credible faces.

Author Keywords
Virtual Faces; Avatar Generation; Player Preferences; Uncanny Valley; Video Games

ACM Classification Keywords
J.4: Computer Application: Social and Behavioral Science; K.8.0: General: Games

INTRODUCTION
Success of video games often relies on the appearance and credibility of their virtual characters. As Rollings and Adams write, "a player will not play a game if its character does not interest the player or is not believable" [25]. This might be correct for most of today's online and role-playing video games (RPGs), which depend on continuously technical improvements in realistic computer graphics. Video games developers continuously increase the degree of details to create more credible human-like characters and to improve immersion in virtual worlds.

CHI PLAY 2015, October 03 - 07, 2015, London, United Kingdom

ACM 978-1-4503-3466-2/15/10$15.00
DOI: http://dx.doi.org/10.1145/2793107.2793116

Figure 1. Random selection of 15 user generated faces.

But increasing the human-likeness of virtual characters becomes a problem in regard to a counter-intuitive phenomenon, called the Uncanny Valley [22]. This effect is already known in robotics or computer animated movies and forecasts a rejection towards too human-like entities.

The Uncanny Valley hypothesis suggested by the roboticist Masahiro Mori in 1970 [22] predicts negative feelings towards figures or prosthetics that are not quite human-like. The term emerges from the graph in Figure 2 which illustrates the relationship between familiarity and human-like appearance. The more human-like characteristics a figure has, the more likely it will be accepted. However, at a certain point the similarity to humans causes a reverse effect. The affinity rapidly changes into repulsion. The figure appears in a negative way to its human observer and falls into the valley, and only an indistinguishable real human is fully accepted by observers again.

Prediction errors increase negative feelings [21] and lead to a concept that does not match with the visual sensation. Neuroimaging confirms this effect and show brain activity in regions, which has been associated with violations of prediction [26, 31]. That mismatch can lead to an interrupt in feeling empathy [20], which also means a loss of immersion in games and identification with the main character [3]. Previous research confirmed, that less human-like designs trigger less negative feelings [8].

However, the effect was primarily examined using photos, videos of robots or computer animations. If people actually prefer credible and human-like characters, the question arises which appearance and amount of details would they choose, when given a free design choice. Due to negative

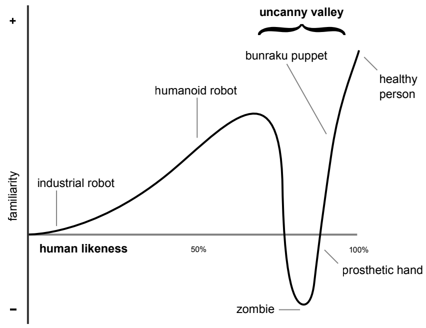


Figure 2. Simplified version of Moris original graph. Modified illustration by MacDorman [17]

emotional reactions towards very realistic human figures, we assume that avatars with unpleasant or uncanny associations consciously get realistic but abnormal characteristics. We suppose, that positively rated faces, get attractive features. But it is still unclear which and how strong these features are or in which degree a face is stylized.

An important feature of current online or role-playing games is able to give insight into these questions: customization of game characters using avatar generators. With a variety of parameters, users are able to personalize their virtual appearance. This kind of personalization can be used to answer the question, how gamers control the appearance and which settings they prefer. However, a study using an interactive manipulation of a virtual human's face to investigate the people's idea was not conducted so far. This might be owed to the fact that avatar creation modules in games generally allow copious adjustments, but are limited by in-game user interfaces, based on predefined models in certain styles, or include elements, which are unsuitable for research purposes.

In this paper we present the results of an online study using a new web based avatar generator called *faceMaker*. Participants were asked for their preferred parametric changes of six different face types. Thus, we investigate, whether users either create realistic characters with parameters, that leads into the Uncanny Valley, or search for alternate designs. Contributions of this papers are: (1) We identify important and preferred parameters. (2) We present the common concepts of stereotypical, preferred and rejected faces, and (3) give advice for design principles to avoid uncanny effects for games and interactive applications. These findings can also be very helpful for researchers as well as designers to create more credible and accepted game characters and could improve new avatar creation systems.

RELATED WORK

Previous work in character generation concentrates on the examination of preferences, attitudes, and identification of players. Exploration and case studies also report the outcome in games and decision making processes of game developers.

Nowak and Rauh [23] evaluated avatars in terms of androgyny, credibility, homophily, attraction, and likeability. The study indicates that people are more likely to perceive an avatar as credible when it was human (instead of an object or animal). Anthropomorphism and lower androgyny of avatars increase attractiveness, credibility, and homophily. Larsson and Nern [16] explored certain attitudes towards sexualized appearances of female stereotypes in games: females and males preferred exaggerated sexual bodies for their personal avatar over normal bodies. Chung et al. [4] suggested that this creation could lead a stronger sense of self-presence and more identification to the created character. The presented studies were conducted using image renderings or pre-defined models of characters. No individual customizations were allowed to the users.

Content and design decisions of game developers are also related and were investigated by researchers. A content analysis conducted by Downs and Smith in 2009 [6] of 489 game characters in 60 top-selling video games says that players had more opportunities to select a non-human character (robots or anthropomorphized beings) than female characters. The study also determines that females are shown more nude, with an unrealistic body shape, and inappropriate clothing. A case study conducted by John [13] delivers insights of design decisions (and conflicts) between graphic designers of a game company: While the male character has to be the main figure and a photo-realistic ”average guy” with anatomically correct proportions, the only female character was ”designed explicitly not realistic to adjust her to heterosexual male fantasies. [13]”. In our paper, we will investigate whether people who acts as designers themselves, also would decide as the game developers.

Researchers also investigate the process of avatar creation itself: Heeter [11] observed teens in designing games and describe that girls allow a very high level of avatar customizations while boys rather used pre-defined characters. An exploratory study by Hussain et al. [12] found that 57% of online gamers decide to play a different gendered character. A study by Rice et al. [24] indicates similar interests in customization of an avatar between different age groups. Older players prefer higher attractiveness and show a higher homophily than younger people. Ducheneaut et al. [7] compare avatar creation systems of three different virtual environments. Their study show that user emphasize the capability to change body shapes, hair styles, and hair colors. Thus, we assume that hair is an important feature in perceiving a face. But it remains unclear, which hair color is a significant predictor of a positive perception of a virtual face.

The related work show, that further investigations are necessary to get more knowledge about facial preferences of users towards avatars and game characters and whether and how people would avoid the Uncanny Valley.

Uncanny Valley

As previously mentioned, the Uncanny Valley is a phenomenon discovered and described in the field of robotics [22]. Meanwhile, researchers have found that the phenomenon can be applied to many other disciplines related to anthropomorphic figures [2]. Early empirical hints for the Uncanny Valley with human-like game characters were found by Schneider et al. [27]. MacDorman et al. [18] list some explanations of the Uncanny Valley and design principles for bridging the valley: High polygon counts with smooth facial proportions and less photo-realistic textures. But their study only investigates manipulations of eyes and skin from one pre-rendered CG male face and is unrelated to impacts of other facial characteristics or unrealistic stylization. Results of further study by Tinwell et al. [30] show that not only adults but children (9-11 year old) also do experience an uncanny feeling in too human-like virtual characters.

Summary

Researchers have demonstrated that artificial figures may trigger the Uncanny Valley in games. But there are no findings about the context of facial preferences and the effects associated with the Uncanny Valley phenomenon. Thus, we decided to examine preferences of users towards generated faces and how they were designed to be liked or rejected.

INTERACTIVE FACE GENERATION

To determine facial preferences, we decided to use an avatar generation system that both allows as well as records interactive customizations of a human young adult face. Body, clothes, very old or young ages, hair styles (including skinheads), scars, tattoos, and equipments rely on the context of a game or the environment and were not included in our examination of facial preferences.

For reliable investigations of the face the following prerequisites must be met: (1) A neutral human with a minimum of characteristics, that could be preferred or rejected, (2) uniform and outbalanced mesh distance in order to consistently deform facial changes, and (3) must originates from the population and mean age of the surveyed target group (European or Western-oriented countries). The system to change the human face model should meet the following criteria: (1) Reach a large number of participants as possible under similar conditions as gamers usually build avatars, (2) contemporary and interactive rendering engine, (3) should be controlled by a simple, unified user interface.

The Average Face

In order to meet all requirements of an editable human face, we decided to create a 3D model based on the Caucasian average face. Humans of that population are investigated in our target group and regularly appear in modern video games. Furthermore, the Caucasian average face is quite well known from foregoing anthropometric or attractiveness research using digitally processed faces [10]. Due to the lack of any individual characteristics of a human average face, participants can unhampered blend between facial changes and need not compensate unwanted or unfamiliar tendencies. However, previous generated images of average faces emerges from certain research questions and do not clearly refer to a large and reproducible group of faces from our target population. Due to gender related differences among female and male average face, we decided to develop both average faces separately.

Parameter Identification and Classification

To develop an avatar creation system, which both evaluates general issues as well as the importance of facial features, sets of 5 general and 32 facial parameters were identified.

General Parameters

Two general questions inevitably emerges because of using the average face, three further general questions emerges from the related work:

1. Participants should change gender related facial differences with one parameter. As previously mentioned, people are more likely to lower androgyny [23] but there is little knowledge about the degree of masculinity and femininity of virtual faces. To investigate the balance between facial stereotypes, we introduce a continuous *face gender* morphing between both models, starting from an androgynous center.

2. Further studies of the average face discovered that it is more attractive than individual faces from which it is composed [10, 14]. Bumps of skin and facial asymmetry disappear by overlapping faces, which is associated with physical attractiveness caused by mate choice in societies with higher parasite loads [9]. But this effect in image-processing is contrary to the assumption that more realistic or detailed characters should be used in games and have an influence on the perception of our model. Thus, we investigate whether and when the concept of a human face is consciously preferred and introduce the investigation of asymmetric *face details* on the skin using a simple multiplied blending of a realistic phototexture above the regular skin map of the average face.

3. Characters from the Uncanny Valley are often associated with dead or zombies [29]. These are often depicted with relatively bright skin. In order to control the preferred skin type we add the *skin color* parameter, which linearly blends between dark and bright human skin, starting at the original map texture of the average skin tone.

4. To investigate the research question, whether realism or stylization is preferred in avatar creation, we developed a representative cartoon model of the average face with exaggerated and unrealistic but common used facial proportions. In order to implement a common way of stylization, two modeling artists aligned the model with styles of current animated family movies and comic-like game titles (see Figure 8). The morph state was controlled by the *face style* parameter, starting from the average face without stylization.

5. The related work emphasizes the importance of hair. To investigate if rather darker or brighter hair color can positively or negatively influence the perception of a virtual face, we introduce the *hair color* parameter, which enables to blend among natural hair colors from black to bright blonde, starting from the average medium blonde. As previously explained certain hair styles were not included in our examination.

Facial Parameters

In addition to the general parameters, we classified the facial sections: eyes, eyebrows, nose, outer face, head shape, mouth, and make-up. 9 State-of-the-Art RPGs were examined for getting an impression of prevalently used facial parameters: *Mass Effect*, *The Elder Scrolls: Oblivion & Skyrim*, *The Sims II & III*, *World of Warcraft*, *Destiny*, *Dragon Age 1 & 2*. Similar parameters were merged, obviously irrelevant (e.g. elf ears) were not taken into account. For usability and analysis issues, we declare a maximum number of 32. Table 1 lists all implemented parameters, their impact $(-/+)$, and default value $(*)$.

parameter	$-$	**values**	$+$	**type**
face gender	female	androgynous*	male	t m
face style	realistic*		cartoon	m
face details	none	half*	full	t
skin color	black	average*	white	t
hair color	black	brunette med.blonde* red	bright blonde	t
eyes color	black	brown* amber blue lt. blue green	grey	t
eyes shape	droopy	down round oval* almond up	asian	m
eyes opening	narrow	average*	wide	m
eyes size	small	average*	big	m
eyes height	up	average*	down	m
eyes distance	narrow	average*	wide	m
eyes orbit	bulgy	average*	covernous	m
eyes rotation	in	average*	out	m
eyebrows color	black	brunette med.blonde* red	bright blonde	t
eyebrows shape	pointed	straight average* round	hooked	m
eyebrows strength	thin	average*	thick	t
nose shape	snub	average*	hooked	m
nose length	short	average*	long	m
nose width	thin	average*	thick	m
nose bridge	thin	average*	thick	m
nose cartilage	round	average*	flat	m
forehead size	down	average*	up	m
ear size	small	average*	big	m
throat size	thin	average*	thick	m
jaw shape	triangle	average*	squared	m
jaw length	long	average*	short	m
chin shape	pointed	average*	cleft	m
cheeks shape	full	average*	scraggy	m
lips volume	thin	average*	full	m
lips size ratio	upper lip	average*	lower lip	m
mouth shape	down	average*	up	m
mouth width	wide	average*	narrow	m
mouth height	up	average*	down	m
mouth depth	backwards	average*	forwards	m
make-up eyes shadow	none*		full	t
make-up lipstick	none*		full	t
make-up rouge	none*		full	t

* default value, t = texture blending, m = mesh morphing

Table 1. Identified and implemented linear parameter scales (from top to bottom): common parameters, eyes, eyebrows, nose, mouth, outer face, cheeks/jaw, and make-up

Modeling of Face and Morphings

Face repositories of 117 male and 151 female Caucasian people from 18–40 years were downloaded from the online Face Database[1] of the Park Aging Mind Laboratory (PAL) [19] and 3d.sk[2]. Both average faces (see Figure 3) were generated using the average face prototyping function of PsychoMorph[3], developed by Tiddeman et al. [28]. These templates were used as inputs of the PhotoFit feature to generate 3D models with the FaceGen[4] software. Some features of the generated 3D models were not completely useful for our purposes. Thus, both meshes were retopologized and retextured using Autodesk® Maya® 2014 and Mudbox® 2014 by two experienced 3D modeling and texturing artists. The results of the template compositions of the processed averaged faces and the final 3D models are depicted in Figure 3.

[1]PAL Face Database: http://agingmind.utdallas.edu/facedb
[2]3D.SK: http://www.3d.sk
[3]PsychoMorph: http://users.aber.ac.uk/bpt/jpsychomorph
[4]FaceGen: http://www.facegen.com

Figure 3. *(l)* overlay compositions of the female and male average face, *(r)* final models in 3D with skin texture (without details layer).

Deformations of the parametrized human faces were realized using vertex displacements also know as morphings or blend shapes. Figure 4 shows the eight facial domains, resulting from the classification in Table 1 and were used as vertex selections sets while modeling as well as for explanation in the help windows of the final application. Morphings were modeled by hand.

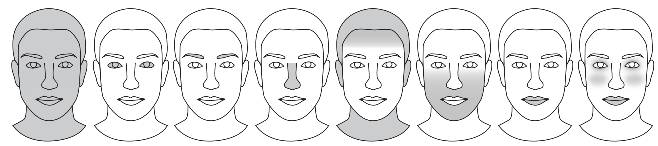


Figure 4. Facial domains used for vertex selection and help graphics of the final application (from left to right) : common, eyes, eyebrows, nose, mouth, outer face, cheeks/jaw, and make-up

Application Design

Only one available avatar creation system which allow collecting user data was found [1]. But, the core objectives of that system do not match with our requirements, because the system based on a whole human body model and do not allow facial parametric changes on the human average face. In order to realize a contemporary interactive avatar generator that meet all requirements, we develop a new application called *faceMaker*, which controls facial changes using parametric values. According to our research design we do not directly investigate which parameters participants select in the mean, but how they build a face that they like and what they do to realize this. The application design prevents the calculation of mean characteristics in the final evaluation for two reasons: (1) Multiple features (e.g. *eyes color*) on a single scale allow no reliable assumption about a average. (2) Participants sometimes will not change the default value. But, in order to understand which concepts participants have of a face, additional measures were taken: (1) objectives: participants should implement predefined or arbitrary concepts, (2) parameter changes, (3) areas of interests (4) assessments: participants evaluate their success in objective completion as well as the likeability, attractiveness, and gender affiliation of a face.

Objectives

Based on our research question of how positively or negatively associated faces, we introduce 6 objectives as independent variables: (A) An personal arbitrarily avatar face. (B) An uncanny, repulsive face. The stereotypical, positive related face of an attractive (C) heroine and (D) hero. The stereotypical face of a (E) female villain and (F) male one, which is also related to negative associations. Using these objectives clearly separates facial changes attributed to positively and negatively associated faces as well as to male and females. This design also enables an evaluation of stereotypical faces which look similar to an arbitrarily designed avatar.

Measurements
Parameter values, their usage (clicks), and mouse moves above the interface were dependent measures. Mouse movements were used to rotate and to slightly zoom the 3D camera to the corresponding region of the parameter. There were three reasons: (1) we make sure that possible changes on the face were noticed immediately, (2) all participants look at the same view during parameter changes, and (3) we learn more about areas of the face, which drawn increased attention. Furthermore, the mean displacement of all vertices and texture blendings starting from the androgynous center between the male and female average face mesh were calculated.

Self-Assessments
After generating a face, the participant had to evaluate the success in completing the objective on a 7-point Likert scale (very successful (7)/very unsuccessful (1)), likeability (very likeable (7)/very unlikeable (1)), attractiveness (very attractive (7)/very repulsive (1)), and gender affiliation (very masculine (7)/very feminine (1)). The question of success in completing the objective acts both as control function to ensure that a user complete the task correctly as well as an indicator for possible insecurities in face generation.

Apparatus
The online WebGL application *faceMaker* was developed using the Javascript library Three.js[5], PHP and MySQL. Physical based skin rendering using the Beckmann shader was applied for the face model. Three directional lights and a slight ambient light (all neutral white) were used for lighting. The key light casts shadow maps. The camera could be rotated within a half circle in front of the face. The face consists of the head model, eyes, and eye lashes. For a neutral transition from neck to the décolleté, a gray T-shirt was added. The view was in fullscreen of the browser window (see Figure 5) with a neutral, dark grayed background.

Figure 5. Start screen of the final *faceMaker* application: 37 slider were organized in 8 randomly arranged containers. Default values begins between the average male and female face as well as 50% skin details.

All facial parameters were linearly controlled using horizontal sliders. Every slider has a width of 200 pixels and was labeled with the facial parameter at its left. At the right a reset button was placed to return to the default value. All parameters were gray, color sliders were equipped with a colored scale. For every new user, all sliders within a domain group as well as the groups on the interface were shuffled an distributed on the left and right column (see Figure 5). If a slider possessed multiple values, the current value name was displayed using tooltips below the slider.

[5]ThreeJS WebGL Engine: http://threejs.org

STUDY
The online study was conducted using a foregoing questionnaire and the *faceMaker* application. The testing phase lasted for six weeks with four beta testers. The study phase lasted three weeks. In this time, a developer monitors the application and answered to technical questions and compatibility issues via e-mail. No changes of the application were made during the main study.

Procedure
The procedure was devided into five steps: (1) Every session starts with demographic questions about gender, age, origin, and in consummating games and movies. Terms of use have to be accepted. (2) Objectives and instructions appear in a window, which have to be closed to edit the average face using all 37 sliders. Reset buttons could restore the default value. (3) Before submitting, a participant had to fulfill the four self-assessments about the created face. (4) After that, the application goes back to step 2 and was repeated until all six faces were completed. (5) If all six objective were finished, the user could view and download renderings of the submitted faces. Each objective could be processed only once. Repeats were prevented using cookies. The order was determine using the objective with the fewest participants and then successively processed in a session using a balanced Latin square (6x6). A participant could continue the study up to 7 days, in this case all timers were stopped when leaving.

Participants
We recruited 431 participants from 7 to 75 years ($M = 29.52$, $SD = 11.42$) using mailing lists, social networks, and advertisements. 150 participants (35.7%) declared that they play video games more than once a week, 75 (17.9%) play more than once a month, 114 (27.1%) play infrequently, 81 (19.3%) say that they do not play video games. 204 participants (48.6%) specified that they watch movies or series more than once a week, 173 (41.2%) watch them more than once a month, 38 (9.0%) watch infrequently, 5 (1.2%) never. 267 participants (63.6%) have their roots in Germany, 48 (11.4%) from Poland, 14 (3.3%) from Italy, 10 (2.4%) from the U.S., and 72 (17.1%) from other countries in the western-oriented world.

RESULTS
1403 faces were received. 62 faces were dismissed due to following reasons: cookies were deleted during session (4), no parameter changes (32), creation time of a single face was up to 6 hours (4), participants with incompatible OS language (22). Finally, 1341 valid faces were generated by

420 participants (204 males, 210 females, 6 n.a.) were evaluated. 145 (34.5%) participants complete all 6 objectives ($M = 3.19$, $SD = 2.19$).

Self-Assessments and Common Parameters

To examine facial parameters which influence the subjective perception of a face we use a regression analysis, as it is used in previous analyses of physical attractiveness (c.f. [5]). How well facial preferences can depict likeability, attractiveness, and gender affiliation in relation to the average face, three multiple linear regression analyses were conducted using the enter method.

The regression equations with the facial parameters as predictors was significant for likeability ($R^2 = .463$, $R^2_{Adj.} = .447$, $SE = 1.560$, $F(37, 1303) = 30.313$, $p < .001$, $d = 2.050$), attractiveness ($R^2 = .456$, $R^2_{Adj.} = .441$, $SE = 1.457$, $F(37, 1303) = 29.518$, $p < .001$, $d = 2.047$), and gender affiliation ($R^2 = .630$, $R^2_{Adj.} = .620$, $SE = 1.246$, $F(37, 1303) = 60.049$, $p < .001$, $d = 2.027$). The scatterplots (not illustrated) of standardized residuals indicated that the data met the assumptions of homogeneity of variance, linearity, and homoscedasticity for all three regression analyses. No auto-correlations d were found. No effect was expected in the model of success in objective completion, which was confirmed by a fourth test ($R^2 = .059$, $R^2_{Adj.} = .032$, $SE = 1.403$, $F(37, 1303) = 1.823$, $p < .001$). We therefore assume, there is a reliable model on the relative importance among facial parameters and final concepts of likeability, attractiveness, and gender affiliation. Table 2 lists all β-coefficients for each facial parameter. Related to the average face, the coefficient can be considered as a measure of the impact on a subjective impressions.

We firstly considered the results for the common parameters: Weak partial correlations among other facial parameters, excluding the parameters with the high β-coefficients (*face details*, *skin color*, and *face gender*), ranged from $-.166$ to .138 for likeability, $-.166$ to .085 for attractiveness, and $-.073$ to .0.93 for gender affiliation indicate that relevant judgments about the importance of the three predictors could be made.

Negative values are inversely related. The direction of that relation can be extracted from the direction of the facial parameter specified in Table 1. For example, both negative β-coefficients for *face gender* indicates that female faces with an increased femininity are more likeable and more attractive.

Gender affiliation and face gender correlates positive because indicator and parameter used an increasing value for masculinity. The influence of the *face gender* parameter on the assessment of gender affiliation can clearly be attributed to the given task. The weighting is higher because participants initially changed its value according to the four gender related objectives. Spearman's ρ reveal a strong significant correlation of *face gender* and gender affiliation ($r_s(1341) = .727$, $p < .001$). We therefore assume that the participants were able to generate their ideas of male or female faces. However, *face details* and *skin color*, voluntarily configurable parameters without initial relations to any objective, are strongly and frequently be used to design faces more likeable (or unlikeable) and more attractive (or repulsive).

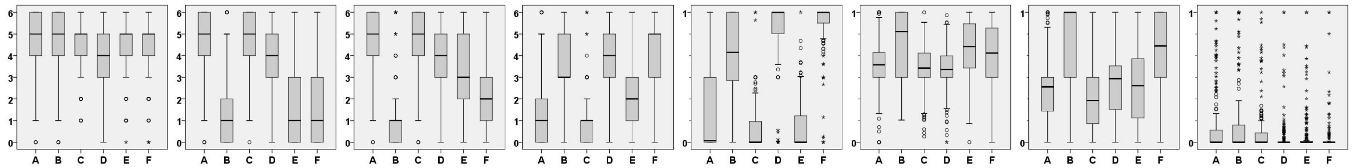
predictors	**likeability β**	**attract. β**	**gender aff. β**
face gender	−.166**	−.169**	.615**†
face style	−.036	−.102**	−.062**
face details	−.249**†	−.242**†	.075**
skin color	−.167**	−.183**†	−.023
hair color	.001	−.026	−.006
eye color	−.025	.018	.003
eye shape	.048*	.077**	−.034
eye opening	.064*	−.014	−.058*
eye size	.025	.051*	.011
eye height	.005	−.002	.000
eye distance	−.006	.008	−.025
eye depth	−.009	.016	−.017
eye rotation	.034	−.006	.044*
eyebrows color	.062*	.072*	.004
eyebrows shape	−.078**	−.041	.002
eyebrows strength	−.002	−.006	.052*
nose shape	−.106**	−.161**	.030
nose length	−.018	.009	.004
nose width	−.017	−.012	.041
nose bridge	.033	.006	.003
nose cartilage	.003	.030	−.034
forehead size	−.009	.023	−.008
throat size	−.013	.013	.093**
ear size	−.061*	−.105**	.030
cheeks shape	−.034	.001	.041*
jaw shape	.048	.016	.009
chin shape	.015	.030	.031
jaw length	−.019	−.065*	.005
lips volume	.131**†	.085**	−.073**
lips size ratio	−.014	−.021	.011
mouth shape	.138**	.086**	−.006
mouth width	.066*	.065*	.002
mouth height	−.041	−.037	.018
mouth depth	.001	.059*	−.010
make-up eye shadow	−.106**	−.044	.021
make-up lipstick	−.161**	−.131**	−.001
make-up rouge	.016	.006	−.036

* $p < .05$, ** $p < .001$, † partial correlation $r_p < -.2$, $r_p > .2$

Table 2. Standardized β-coefficients show the strength of relationships between predictors and assessment scores of likeability, attractiveness, and gender affiliation related to the average face.

To determine possible relationships between the basic parameters *face details*, *skin color*, *face gender* and both self-assessments a further series of Spearman's rank-order correlations were conducted. A two-tailed test of significance indicated that there was a positive correlation between likeability and attractiveness ($r_s(1341) = .743, p < .001$), but we assume that the parameters are not entirely redundant. Inverse relationships between *face details* and likeability ($r_s(1341) = -.400$, $p < .001$) as well as between *face details* and attractiveness ($r_s(1285) = -.424$, $p < .001$) were found. Both negative correlations of *skin color* towards likeability ($r_s(1341) = -.325$, $p < .001$) as well as attractiveness ($r_s(1343) = -.300$, $p < .001$) were weaker.

The box plots from Figure 6b and 6c show the similarity of the two assessments as well as the contrary relationships among them and the parameters *face details* and *skin color* shown in Figures 6f and 6g. Due to lack of normal distribution a series of Kruskal-Wallis tests were conducted to find



(a) success (b) likeability (c) attractiveness (d) gender aff. (e) face gender (f) skin color (g) face details (h) face style

Figure 6. Box plots of self-assessments and facial parameters between the objectives (A) arbitrary face, (B) repulsive face, (C) female hero, (D) male hero, (E) female villain, (F) male villain

	self-assessment		facial parameter		
test	like-ability	attrac-tiveness	face details	skin color	face gender
A-B	576.4**	−630.0**	−442.6**	−289.0**	−335.2**
A-C	68.8	−75.8	98.3	60.8	101.0
A-D	68.1	96.2	−45.5	69.2	−525.9**
A-E	480.2**	273.9**	−29.4	−239.1**	85.3
A-F	507.0**	415.1**	−298.6**	−159.6**	−553.3**
B-C	−645.2**	−705.8**	540.8**	349.8**	436.1**
B-D	−508.3**	−533.9**	397.1**	358.2**	−190.7**
B-E	−96.3	−356.1**	413.1**	49.9	420.5**
B-F	−69.4	−214.9**	143.9*	129.4*	−218.1**
C-D	136.9*	171.9**	−143.7*	8.3	−626.8**
C-E	548.9**	349.7**	−127.7*	−299.9**	−15.6
C-F	575.8**	490.8**	−396.9**	−220.5**	−654.2**
D-E	412.0**	177.7**	16.0	−308.3**	611.1**
D-F	438.8**	318.9**	−253.2**	−228.8**	−27.4
E-F	26.8	141.2**	−269.2**	49.4	−638.6**

* $p < .05$, ** $p < .001$

Table 3. Rank differences (H) among all objectives: pairwise Kruskal-Wallis post-hoc tests with asymptotic significances (2-sided). (A) arbitrary face, (B) repulsive face, (C) female hero, (D) male hero, (E) female villain, (F) male villain

contrasts among these tendencies. The tests showed significant differences of the distributions of likability ($\chi^2(5, N = 1341) = 644.832$), attractiveness ($\chi^2(5, N = 1341) = 552.770$), *face details* ($\chi^2(5, N = 1341) = 320.970$) and *skin color* ($\chi^2(5, N = 1341) = 184.985$) across all objectives (with all $p < .001$). Table 3 lists the post-hoc rank differences of all pairwise comparison tests.

Significant rank differences were found by pairwise comparisons between the ratings of success through a independent-samples Kruskal-Wallis test ($\chi^2(5, N = 1341) = 60.365$, $p < .001$) in the category of male heroes towards arbitrary ($H = 245.4$, $p < .001$), repulsive ($H = -185.3$, $p < .001$), female hero ($H = -138.2$, $p < .001$), and female villain ($H = -141.8$, $p < .001$) faces. In spite of quite positive ratings of success (see Figure 6a), we therefore assume that it was relatively more difficult for users to generate faces of male heroes.

Facial Preferences

In the following we report results of the regression analyses, which are related to facial and significant parameters within in the facial sections.

Eyes and Eyebrows

The *eye shape* is highly significant for both attractiveness ($\beta = .077$**) as well as likeability ($\beta = .048$*). The β-coefficients are slightly positive which means that more almond, upturned, or asian shaped eyes in relation to the average face are more likeable and attractive than downturned or droopy eyes. Interestingly the complete eye was slightly downturned ($\beta = .044$*) to shape more masculine faces. *Eye opening* controls the distance between the eye lids. The β-coefficients for likeability ($\beta = .064$*) is positive and gender affiliation ($\beta = -.058$*) negative, which means that slightly opened eyes are more attractive and less masculine. A higher β-coefficient for *eye size* ($\beta = .051$*) indicates, that slightly increased eyes size are more attractive. Likeability ($\beta = .062$*) and attractiveness ($\beta = .072$*) raise using brighter eyebrows than with the average brightness. Straight and pointed *eyebrows shapes* ($\beta = -.078$**) lead significantly to less likeability, stronger eyebrows ($\beta = .052$*) to more masculine faces.

Nose, Outer Face, and Head Shape

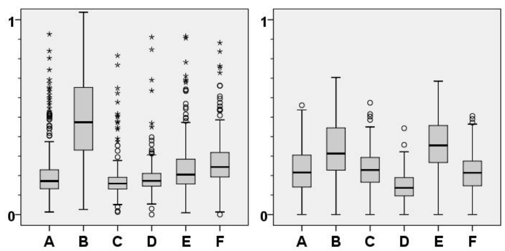
Only one facial parameter in the nose group significantly lead to more likeability ($\beta = -.106$**) and attractiveness ($\beta = -.161$**): the nose shape. When viewing negatively associated faces we already noticed the often used hooked nose, while positive associated face were sometimes equipped with straight but mainly with snub noses. This effect can be observed at both gender. Wider throats ($\beta = .093$**) and scraggy cheeks ($\beta = .041$*) were used to generate more masculine faces. Ear sizes were reduced to generate more attractive ($\beta = -.105$**) or likeable ($\beta = -.061$**) faces. The jaw length was also reduced for more attractive faces ($\beta = -.065$*).

Mouth and Make-Up

A dynamic facial parameters is the volume of the lips, which was strongly increased to generate more likeable ($\beta = .131$**) and attractive ($\beta = .086$**) faces, but were reduced to generate more masculine faces ($\beta = -.073$*). Mouths were shaped upturned ($\beta = .138$**, $\beta = .086$**) and slightly tighter ($\beta = .066$**, $\beta = .065$**) in relation to the average to generate more attractive and likeable faces. Slightly protruding lips lead to more attractiveness using mouth depth ($\beta = .059$**). Texture blendings of make-up layers show a different effect of the actually sense of that feature. More eye shadows ($\beta = -.106$**) lead to lower likeability, lip stick to lower likeability ($\beta = -.161$**) as well as attractiveness ($\beta = -.131$**). Both features are often used for villain faces (not illustrated).

Stereotypes and Differences to the Averages

Due to different vertex displacement between the facial morphings, it was not possible to compare the objectives and the original average faces with using the mean values of the parameters. Thus, we calculate the mean displacement of all vertices as well as texture blendings within the mesh to understand, how far a face differ from the androgynous average face (see Figure 7). All vertex displacements are normalized to their tenth starting from the androgynous center of both average faces. This means, a displacement of 0.1 is equal to a mesh blending to both the average male as well as the female face respectively.



(a) vertex displacement (b) texture blending

Figure 7. Mean differences of vertex displacements and textures transparencies between the objectives. A texture blending of 0.1 is equal to a transparency of 10%. Error bars show 95% CI.

Further independent Kruskal-Wallis tests shows significant rank differences between the vertex deformations ($\chi^2(5, N = 1341) = 446.1$, $p < .001$) and texture blendings ($\chi^2(5, N = 1341) = 368.4$, $p < .001$) of stereotypical faces. Pairwise comparisons of vertex displacements reveal adjusted significances for all test pairs excluding F-A ($p = 1.000$), F-C ($p = 1.000$), B-E ($p = .093$), and A-C ($p = .469$). Mean texture transparencies differ between all test pairs excluding D-A ($p = .495$), F-C ($p = .173$), and A-C ($p = 1.000$). Both tests underscore the similarity between the average arbitrary face and the heroine shown in Figure 10a and 10c.

Realism versus Stylization

One research questions was whether and how participants used a stylized cartoon face and how they control the visibility of details if they had a choice. Thus, each participant had the chance to stylize a face (or not) and to completely remove (or add) texture details. However, in 71.0% of all cases the face was absolutely not morphed by the parameter *face style*. In contrast, only 11.8% of all faces received no *skin details*. In the cases when stylization was applied, participants choose a very low degree of cartoon-like proportions (see Figure 6h).

To understand why some participants choose this facial proportions, we examine occurrences between the objectives and discover an increase for the heroine face between a morphing range 0.0% to 33.0%. This causes the following effect: The face appears almost realistic, but seems younger through childlike proportions (see Figure 8) – attributes that are perceived as particularly attractive [15]. Light *skin details* give the impression of a healthy but natural and realistic skin. Also noticeable at this point is, that villains were rarely stylized but few people also use strong stylization (from 66–100%) to generate repulsive faces.

Times

The mean trial time in generating a face is about 8 minutes ($M = 8.67$, $SD = 19.09$) and slightly varies between the objectives. Participants took most time for the arbitrary

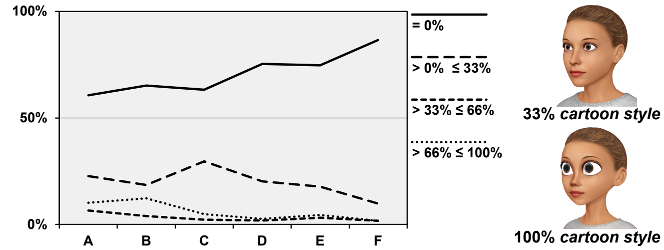
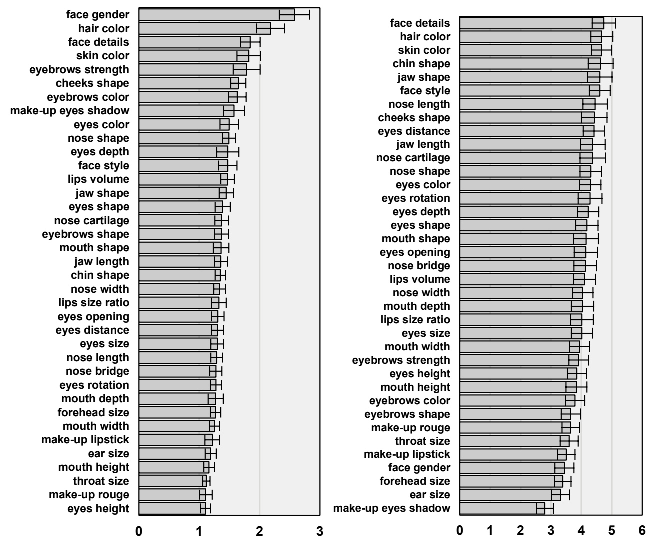


Figure 8. User stylizations in percent: Heroines get most stylized morphings from a blending range of 0–33%. (A) arbitrary, (B) repulsive, (C) heroine, (D) hero, (E) female villain, (F) male villain.

face ($M = 11.15$, $SD = 27.65$), male villain ($M = 8.62$, $SD = 15.39$), and the heroine ($SD = 9.79$, $SD = 25.14$). The male hero ($M = 7.82$, $SD = 9.94$), the repulsive ($M = 7.43$, $SD = 18.18$), the female villain ($M = 7.32$, $SD = 11.49$) are below the average. A one-way ANOVA show no significant differences of the durations times between the objectives ($F(5, 1335) = 1.396$, $p = .223$).

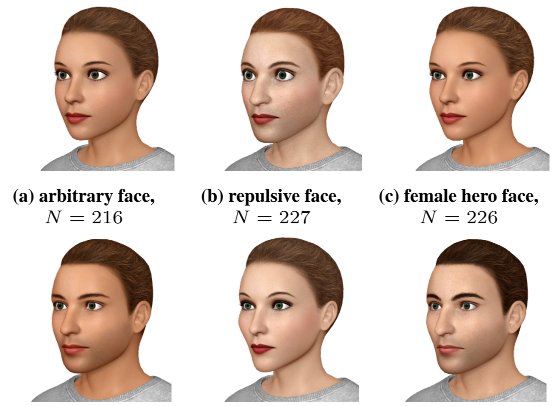


(a) average clicks (b) average mouse moves

Figure 9. Average clicks and mouse moves per user on every facial parameter. Mouse moves were used to rotate and to zoom the camera to the corresponding area. Error bars show 95% confidence interval.

Mouse Clicks and Moves

The average count of slider changes per user is $M = 1.438$ ($SD = 1.438$) and of mouse movements is $M = 4.069$ ($SD = 3.714$). Figure 9 shows a sorted list of both measurements. Comparisons with predictions from the linear regression analyses show that some important values were also changed frequently used (e.g. *face gender*, *face details*, and *skin color*). This does not apply at all values. A particularly no reliable predictor for likability or attractiveness is *hair color*. However, mouse activities indicate, that the hair color is quite important in designing avatars. *Eyes color*, *eyes depth*, and *eyes shadow*, and *eyebrows strength* as well as head deforming parameters *cheeks shape*, *nose shape*, and *face style* were no reliable predictors for likeability and attractiveness but frequently manipulated.



(a) arbitrary face, $N = 216$ (b) repulsive face, $N = 227$ (c) female hero face, $N = 226$ (d) male hero face, $N = 223$ (e) female villain face, $N = 225$ (f) male villain face, $N = 224$

Figure 10. Average stereotypical faces according to the six objectives.

CONCLUSION

This work puts the negative effects of the Uncanny Valley and facial properties on virtual human faces into context. The basis of this investigation is the average face. Our results indicate that smooth skin, natural skin color, and human proportions are the most relevant factors to avoid negative feelings. Users rarely choose a virtual face in an obvious cartoon-like look. This led to new findings in the Uncanny Valley and avatar creation research: If having a choice, people were more willing to create human-like instead of cartoon-like faces known from animation movies or video games. So people consciously move into the direction of familiar human-likeness and actually "just bridge" the valley using very attractive features for faces they like.

We show which facial parameter influence attractiveness or likeability. Negatively associated faces get features that deviate from the human norm. Villains, for example, get strong make-up, and striking mostly hooked noses than the other faces (c.f. Figure 10e). Female villains get distinctive cheek bones and lip stick. Male villains have straight eyebrows and distinctive jaws. Repulsive faces were exaggerated with unnatural violations against the human average. But negative concepts of human faces get generally unnatural bright skin (associated with an uncanny zombie-look) and a very strong overlay of realistic skin details.

In contrast, positively associated faces get smooth skin, realistic proportions, and natural average skin color (of the surveyed population). We also notice that female faces are also equipped with full lips, snub nose, and slightly upturned eyes. Males get strong eyebrows, downturned eyes, a larger throat, and thin lips. Thus, some previous mentioned guidelines (smooth skin, less photo-realistic textures) in the related work to avoid the Uncanny Valley were confirmed [18] and extended. Stylization is rarely used and, if at all, only with a low value on heroines, which gives them a more youthful look. Some repulsive face get stylization in combination with strong skin details in order to leave the human average. We therefore conclude that participants consciously do not violate against human norms and rather search for both credibility as well as attractiveness to avoid negative associations. Positively associated figures are closer to the average face than negatively rated faces (e.g. repulsive faces). This is confirmed by the calculation of average vertex displacements and texture blendings.

There were very small and not significant differences between the voluntary modeled face and the female hero —also for the parameter *face gender* (c.f. Figure 10a and 10c). About 74% of the people choose a rather feminine type when designing an arbitrary face. Due to an almost identical amount of male and female participants, we assume that an attractive female face is generally preferred. Furthermore, concepts of stereotypical (hero, villains) faces are very clear and should not include any androgyny. Both facts confirmed the results of Nowak and Rauh [23].

Further conclusions from our results are interesting for game designers: Male faces get generally more details on their skin than females. Villains get more details than heroes, but less stylization. Female heroes and female villains have a slightly brighter skin than their male stereotypes, but without significant differences. The importance of parametrized hair in avatar creation was confirmed (c.f. [7]) with measurements of the mouse activity. However, hair color is no predictor for perceived likability or attractiveness. Due to very similar color distributions (not illustrated), we assume that there are outbalanced preferences for darker and for brighter hair of a virtual face.

Limitations and Future Work

Likeability and physical attractiveness were used as indicators to learn more about features that make a virtual face more likeable or appealing. This could be extended using further indicators of interest. Shares of more realistic attributes (e.g. freckles, moles, acne, wrinkles, beards, hair, more cartoon styles) affects the impression of skin also should be further investigated. With *faceMaker* we plan to collect more data within a longer period of time to investigate gender related differences, cultural preferences, and possible differences between groups of ages. Technical limitations include the lack of volumetric hair and fine geometrical displacements instead of bump maps.

In this work, a variety of data has been collected, which can be used for further evaluation and research purposes. Therefore, we provide the face, aggregated results, and the *faceMaker* application for free disposal[6]. Due to our many positive experiences, we strongly recommend the average face as base model for design decisions in game development or in virtual face related investigations. Other questions which have arisen are whether people perceive a face differently as the user who create it and how they would rate faces they already know or look like themselves.

ACKNOWLEDGMENTS

Special thanks go to Katharina Leicht, Verena Dengler, Julia Sandrock, and Solveigh Jäger. This work was supported by the graduate program *Digital Media* of the Universities of Stuttgart and Tübingen, and the Stuttgart Media University.

[6]3D files and source code can be downloaded at: https://github.com/valentin-schwind/facemaker

REFERENCES

1. Apostolakis, K., and Daras, P. Raat - the reverie avatar authoring tool. In *Proc. 18th. Int. Conf. on Digital Signal Processing*, IEEE (2013), 1–6.
2. Brenton, H., Gillies, M., Ballin, D., and Chatting, D. The uncanny valley: does it exist? In *Proc. 19th British HCI Group Annual Conf.: workshop on human-animated character interaction* (2005).
3. Brown, E., and Cairns, P. A grounded investigation of immersion in games. In *Proc. CHI EA* (2004), 31–32.
4. Chung, D., DeBuys, B., and Nam, C. Influence of Avatar Creation on Attitude, Empathy, Presence, and Para-Social Interaction. In *Proc. 12th Int. Conf. HCI, Interaction Design and Usability* (2007), 711–720.
5. Cunningham, M. R. Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology 50*, 5 (1986), 925–935.
6. Downs, E., and Smith, S. Keeping abreast of hypersexuality: A video game character content analysis. *Sex Roles 62*, 11-12 (2010), 721–733.
7. Ducheneaut, N., Wen, M.-H., Yee, N., and Wadley, G. Body and Mind: A Study of Avatar Personlization in Three Virtual Worlds. In *Proc. CHI* (2009), 1151.
8. Flach, L. M., Dill, V., Hocevar, R., Lykawka, C., Musse, S. R., and Pinho, M. S. Evaluation of the uncanny valley in CG characters. In *Lecture Notes in Computer Science* (2012), 511–513.
9. Gangestad, S. W., and Buss, D. M. Pathogen prevalence and human mate preferences. *Ethology and Sociobiology 14* (1993), 89–96.
10. Grammer, K., and Thornhill, R. Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of Comparative Psychology 108*, 3 (1994), 233–242.
11. Heeter, C., Egidio, R., Mishra, P., Winn, B., and Winn, J. Alien Games: Do Girls Prefer Games Designed by Girls? *Games and Culture 4* (2008), 74–100.
12. Hussain, Z., and Griffiths, M. D. Gender swapping and socializing in cyberspace: An exploratory study. *CyberPsychology & Behavior 11*, 1 (2008), 47–53.
13. John, S. Un/realistically embodied: The gendered conceptions of realistic game design. In *Proc. AVI, Workshop* (2006).
14. Langlois, J. H., and Roggman, L. A. Attractive Faces Are Only Average. *Psychological Science 1*, 2 (1990), 115–121.
15. Langlois, J. H., Roggman, L. A., and Musselman, L. What is average and what is not average about attractive faces? *Psychological Science 5*, 4 (1994), 214–220.
16. Larsson, A., and Nerén, C. Gender aspects of computer avatars. *SICS Research Report* (2005).
17. MacDorman, K. F. Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it. In *Proc. CogSci Workshop*, vol. 3 (2005), 108–118.
18. MacDorman, K. F., Green, R. D., Ho, C. C., and Koch, C. T. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior 25*, 3 (2009), 695–710.
19. Minear, M., and Park, D. C. A lifespan database of adult facial stimuli. *Behavior research methods, instruments, & computers: A journal of the Psychonomic Society, Inc 36* (2004), 630–633.
20. Misselhorn, C. Empathy with inanimate objects and the uncanny valley. *Minds and Machines 19*, 3 (2009), 345–359.
21. Mitchell, W. J., Szerszen, K. a., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception 2*, 1 (2011), 10–12.
22. Mori, M., MacDorman, K. F., and Kageki, N. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE 19*, 2 (1970/2012), 98–100.
23. Nowak, K. L., and Rauh, C. The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication 32* (2005), 178–197.
24. Rice, M., Koh, R., Lui, Q., He, Q., Wan, M., Yeo, V., Ng, J., and Tan, W. P. Comparing avatar game representation preferences across three age groups. In *Proc. CHI EA* (2013), 1161–1166.
25. Rollings, A., and Adams, E. *Andrew Rollings and Ernest Adams on game design*. New Riders, 2003.
26. Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience 7*, 4 (2012), 413–422.
27. Schneider, E., Wang, Y., and Yang, S. Exploring the Uncanny Valley with Japanese Video Game Characters. In *Proc. DiGRA* (2007), 546–549.
28. Tiddeman, B., Burt, M., and Perrett, D. Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications, IEEE 21*, October (2001), 42–50.
29. Tinwell, A., Grimshaw, M., and Williams, A. Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds 2*, 1 (2010), 3–25.
30. Tinwell, A., and Sloan, R. J. S. Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior 36* (2014), 286–296.
31. Urgen, B. A., Li, A. X., Berka, C., Kutas, M., Ishiguro, H., and Saygin, A. P. Predictive coding and the Uncanny Valley hypothesis: Evidence from electrical brain activity. *Cognition: A Bridge between Robotics and Interaction.* (2015), 15–21.