To Stylize or not to Stylize? The Effect of Shape and Material Stylization on the Perception of Computer-Generated Faces

Eduard Zell[1] Carlos Aliaga[2] Adrian Jarabo[2] Katja Zibrek[3]
Diego Gutierrez[2] Rachel McDonnell[3] Mario Botsch[1]

[1]Bielefeld University [2]Universidad de Zaragoza [3]Trinity College Dublin

Figure 1: *We explore the influence of shape and material stylization on the perception of computer-generated faces, using stylized versions of captured faces of two real people. In particular, we evaluate how any combination of shape and material style affect the perceived realism, appeal, and expressivity of a character performing several basic facial expressions. From most abstract on the left to the most realistic versions on the right, different expressions of four of the five levels of stylization used in the study are shown in the image.*

Abstract

Virtual characters contribute strongly to the entire visuals of 3D animated films. However, designing believable characters remains a challenging task. Artists rely on stylization to increase appeal or expressivity, exaggerating or softening specific features. In this paper we analyze two of the most influential factors that define how a character looks: shape and material. With the help of artists, we design a set of carefully crafted stimuli consisting of different stylization levels for both parameters, and analyze how different combinations affect the perceived realism, appeal, eeriness, and familiarity of the characters. Moreover, we additionally investigate how this affects the perceived intensity of different facial expressions (sadness, anger, happiness, and surprise). Our experiments reveal that shape is the dominant factor when rating realism and expression intensity, while material is the key component for appeal. Furthermore our results show that realism alone is a bad predictor for appeal, eeriness, or attractiveness.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: Character Design, Shape and Material, Stylization

1 Introduction

Over the last years, advances in the field of computer graphics have allowed the entertainment industry to create very realistic virtual humans [Alexander et al. 2010; Jimenez et al. 2015]. However, depicting convincing and believable characters continues to be a difficult task. The reasons why people like the appearance of a particular character may be influenced by many psychological factors, such as the familiarity of the characters to the observers [Dill et al. 2012], the possible adaptation of the society to cartoon faces [Chen et al. 2010], the viewers' level of expertise in computer graphics [Fan et al. 2012], or their increasing ability to notice the tricks and techniques employed [Tinwell and Grimshaw 2009]. In this context, it is generally accepted that stylized versions are often preferred over realism [Geller 2008], but the reasons for this are still unclear. Moreover, stylization allows the artist to explore possibilities beyond what is found in the real world (such as oversized eyes for cartoons, for instance) to enhance the appeal or expressiveness of the characters, particularly in the case of cartoons [McCloud 1993] or illustrations [Gooch et al. 2004].

The challenge is to understand and translate the knowledge of artists into feasible guidelines for generating appealing 3D virtual characters. Some works have investigated the role of few of the many variables involved in the creation of 3D faces, such as the influence of rendering style [McDonnell et al. 2012], anthropomorphism [Seyama and Nagayama 2007], or applying different 2D filters on the images [Wallraven et al. 2007].

In our work, we focus instead on two of the main aspects that primarily define the stylization of a 3D character: shape and material (including texture and optical properties). Due to the high dimensional nature of the problem, experiments were performed in two rounds. We first analyze which of the many sub-dimensions of both shape and material affect the appearance of the character the most. To this end, we define three different levels of stylization along shape and material for a single male character: a realistic head obtained by state-of-the-art 3D-scanning and two stylized versions designed by artists. Moreover, each level includes five different facial expressions: anger, happiness, sadness, surprise, and neutral. We then create all combinations along these dimensions (shape, material, expression) and analyze the perceived realism, appeal, eeriness, and familiarity of each character by means of perceptual studies. We also analyze the effect of decoupling the material dimension into its main components, testing two different shaders, three illumination methods, and three progressively blurred albedo textures. Results and acquired knowledge from these tests are then used to guide a second round of experiments, where we deeply explore the space with more samples along the factors found as most important in the previous studies. For this, we substantially increase the stimuli to two characters (male and female), five stylization levels (of both shape and material), and five expressions. We then analyze the most significant scales of the previous experiments (realism, appeal), and also evaluate how the combination of each of these dimensions affects the expressivity of the characters.

Our experiment design is inspired and justified by the current trends in feature animation, which have recently used different combinations of stylized shapes and materials to depict 3D characters. Examples include highly stylized shapes and textures in Pixar's *Toy Story* movies, or the somewhat less stylized shapes but photorealistic materials in *The Adventures of Tintin*. Furthermore, in a review of recent advances in facial appearance capture, Klehm and colleagues [2015] mention the need for deeper insights into human perception of facial appearance. They note the complexity and the importance of focusing on important features, which we address by carefully isolating the effects of the parameters being studied in each test. We use static pictures as stimuli, as it has been found that much of the information that people use to evaluate virtual characters is available in a still image [McDonnell et al. 2012].

Our main findings are:

- Shape is the key attribute for perceived realism. Stylized materials decrease the perceived level of realism for realistic shapes; however, realistic materials do not increase realism of stylized shapes.
- On the other hand, appeal, eeriness, and attractiveness are mainly affected by the stylization level of material, and not shape; realistic materials reduce appeal in general. Within the materials, the albedo texture is the dominant factor.
- The degree of realism is a bad predictor for appeal or eeriness.
- The perceived intensity of expressions decreases with realism of shape, but is nearly independent of material stylization.
- Our results are consistent across all tested expressions. Only the anger expression has been constantly perceived as less appealing and more eerie.

To our knowledge, this is the first work attempting to evaluate how the combination of different levels of stylization in shape and material affect the perception of a virtual character across different expressions. Conclusions from this study are restricted to the stimuli specifically created for these experiments, but can help to provide useful insights on how to create believable characters. To foster further research, we will make our stimuli and data publicly available at: `http://graphics.uni-bielefeld.de/publications/sigasia2015/`

2 Related work

Creating computer-generated (CG) characters is a challenging task, since even small imperfections can trigger negative responses [Seyama and Nagayama 2007]. Many studies have investigated this effect in the context of the theory of the uncanny valley [Mori et al. 2012], but the variables involved in the perception of virtual faces are still mostly unknown. For instance, Dill et al. [2012] evaluated CG characters on still images and videos, concluding that people often prefer familiar faces, while other studies analyzed how knowledge of computer graphics affects 3D character perception [Tinwell et al. 2011b; Fan et al. 2012]. For brevity, we refer the reader to the excellent discussion on this topic by Tinwell [2014], and focus on works closer to our particular goal.

Stylization Some of the first attempts to measure the likability of stylized and realistic characters was performed by morphing pictures [Hanson 2005; MacDorman 2006; Schneider et al. 2007; Dill et al. 2012]. Schneider et al. [2007] studied the effect of stylization on characters in Japanese video games, and found that it increased perceived attractiveness. All these studies used different characters, including confounding factors such as changing lighting and background. In contrast, we investigate the effects of stylization on the same character under identical conditions. Other works have focused on changing certain features and modifying proportions in the shape of digital faces. It has been shown that uncanniness emerges when abnormal features of the face become apparent for highly realistic characters [Seyama and Nagayama 2007; Burleigh et al. 2013]. Green et al. [2008] concluded that there is less tolerance to deviations from original proportions in the cases where faces are more attractive and human-like. Different from these works, we investigate the effects of global stylizations as commonly adopted by the animation industry.

Wallraven et al. [2007] studied the perceived realism, recognition, sincerity, and aesthetics of real and computer-generated facial expressions using 2D filters to provide brush, cartoon, and illustration styles. They concluded that realistic depictions improve subjective certainty about the conveyed expression. Later, they evaluated the perceptual realism of computer-generated faces under progressively blurred normal vectors and textures, finding no effect with their setup [Wallraven et al. 2008]. In contrast to them, we do not employ Gaussian blurring for producing abstract stimuli, but instead use stylized models produced by artists, in order to better match the character styles used in industry.

MacDorman et al. [2009] showed participants several images of virtual faces, combining different textures (from realistic to simple lines) with *geometric* levels of detail (i.e. decreasing polygon counts). Results suggested that decreasing photo-realism can make the face look less eerie and more attractive. In our work, *shape* refers to the global, high-level features of the face, not to technical aspects such as polygon count. Closer to our goal, the recent study by McDonnell and colleagues [2012] found that rendering style affects the appeal and trustworthiness of the characters. Additionally, a character rendered in an appealing style can be perceived to have more positive personality traits [Zibrek and McDonnell 2014]. Recent studies focusing on neurocognitive mechanisms attribute negative appeal ratings to the difficulty of categorizing images in a particular category, resulting in competing visual-category representations during recognition [Ferrey et al. 2015]. Negative effects for such images occurs to the extent that selecting one interpretation over the other requires inhibition of the visual-category information associated with the non-selected interpretation. Following the conclusions from these studies, stylization affects pleasantness ratings and furthermore, some combinations of visual elements might result in negative effect. Therefore, we study the effects of combining different levels of stylization for shape and material, which are the two key parameters governing visual appearance.

Skin Appearance Taking into account previous work related to the perception of human skin appearance helps understanding effects of material stylizations. Many studies about attractiveness of human faces merged different photographs to achieve average appearance. There was speculation that this technique impacts ratings of attractiveness not just because it averages the shape, but also because it removes blemishes and other skin irregularities [Alley and Cunningham 1991]. Several studies confirmed that texture changes do result in a significantly more attractive face [Benson and Perrett 1992; Little and Hancock 2002]. Publications in the cosmetics domain also help explain the observed effects on appeal: Fink and colleagues [2006] created textures from photographs of

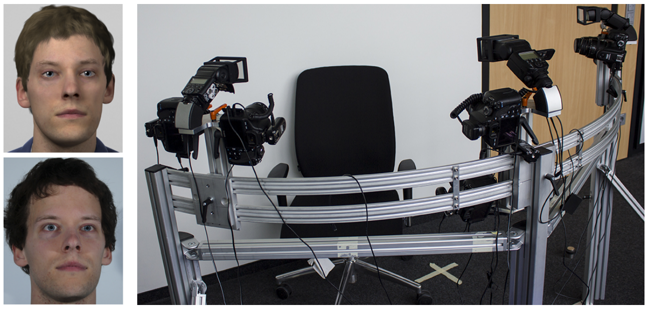


Figure 2: *Our face scanning setup (right) and comparison between photographs and virtual reconstructions of our actor (left).*

women of different age and evaluated these textures on a single female virtual character. Renderings with pure skin have been rated as younger and more attractive than renderings with strong variations in skin pigmentation. This observation was confirmed in a follow-up study [Fink and Matts 2008], which showed that blurring the skin texture can increase attractiveness. Similar suggestions can be found in many photograph retouching books (e.g. [Nitzsche and Rose 2011]).

Expression We are also taking into account the influence of the particular expression. Brain studies show that some areas in the brain respond differently to certain expressions of emotion, specifically the amygdala, which tends to activate while viewing fearful and angry faces, as opposed to happy, surprised, and sad faces [Calder 1996]. Since the amygdala region is activated in response to danger, it is believed that negative emotional expressions, such as anger and fear, trigger a defense response in the perceiver. Another example comes from studying the "uncanny valley" effect on CG characters, where modified expressions of emotion with negative valence (e.g. anger, sadness) increased the perceived uncanniness of the character [Tinwell et al. 2011a]. Additionally, given different hypotheses that iconic representation of faces increase the expressibility and the recognizability of expressions [McCloud 1993], we further analyze our stylization domain by evaluating whether different levels of stylization in shape and material, including mismatches between them, affect these scales.

3 Stimuli Creation

Our initial experiments required the design of three levels of stylization of the same character. Additionally, for each stylization level we modeled four of the universal facial emotions: anger, happiness, sadness, and surprise [Ekman 1972], plus a neutral expression. We discarded disgust and fear because their status as basic expression was questioned recently [Jack et al. 2014] and they are harder to identify by observers.

Our realistic characters are based on real people of about average attractiveness without ethnic bias to the group of participants. To generate the realistic models we replicated the multiview-stereo face scanner of Beeler et al. [2010], which reconstructs high-resolution textured point clouds from the photographs of six cameras arranged as pairs around a person (Figure 2). Since all photographs are taken simultaneously, the scanning process is instantaneous and therefore well suited for capturing different facial expressions. Each pose representing one emotion was captured several times, and the most convincing one was selected by a group of about twenty people of different cultural backgrounds, while referring back to Ekman's guidelines. For the stylized shapes, we did not intend for the artist to precisely match the emotional intensity across the shapes, but rather to create expressions that resembled the expressions of our scanned actors to the best of their ability (e.g., teeth showing slightly in a happy smile) given the available facial features.

Since the scanner only captures the frontal part of the face and fails to faithfully reconstruct eyes and hair, we fit a template head model to the measured point cloud using a non-rigid registration approach similar to Weise et al. [2011]. Regions of missing data are therefore filled in by the template model, which additionally provides a 2D parameterization of the model. This parameterization is used for texture mapping, with texture images being generated automatically from the photographs. The hair style, the eyes, and the teeth were manually sculpted and adjusted to fit the scanned model. Figure 2 shows one example of our reconstructed models.

While a realistic character can be obtained from 3D scans of a real person, no automatic solution exists to generate increasingly stylized versions. Therefore professional 3D artists produced the required stylized shapes and materials from our realistic characters, taking inspiration from commercial animation films (see examples in Figure 3). For our first set of experiments (Section 5) we used three stylization levels for shape and material (see Figure 4). The extended stimuli for the later experiments (Section 6) used two more stylization levels (see Figure 8, left).

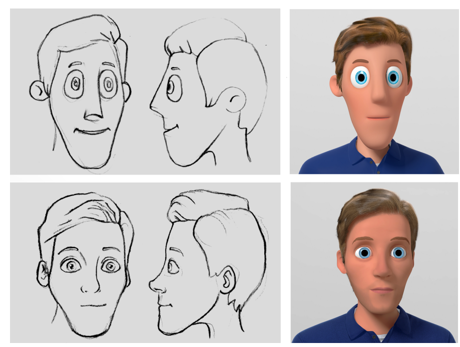


Figure 3: *Two of the stylizations created for the study, showing the sketches provided to the artists on the left and their resulting stylized 3D models on the right. The designs are inspired by the films* Cloudy with a Chance of Meatballs *(top) and* Toy Story *(bottom).*

We are interested in analyzing the effect and interaction of shape, material, and textures. Therefore, we transferred all material properties of the baseline characters to the other character shapes (Figure 4). The inter-surface mapping for the texture transfer was computed based on a dense correspondence map established using the non-rigid registration technique of Zell and Botsch [2013].

Rendering of all stimuli was performed using Mental Ray, with each character being placed in front of a light gray background. The lighting setup consists of a key light and a rim light, and photon mapping is used for global illumination. For the skin all three characters use the same multi-layer skin shader with subsurface scattering, with diffuse albedo specified by a high-resolution texture map. The shader parameters vary between the models in order to closely resemble the targeted render styles (Figure 3).

4 Experiment Design

The appearance of virtual humans is a function defined over a huge multi-dimensional space. While it is generally recognized that *shape* and *material* are the main contributors to the overall appearance of virtual characters, these two might be affected by several sub-dimensions. For example, material is the combination of shader, shader parameters, and textures, each of which having a potentially different influence on appearance. This makes the experiment design an extremely difficult task, given the large number of variables to explore.

Similar to previous work on rendering style [McDonnell et al. 2012], we want to analyze how different levels of stylization (e.g., shape and material) change the perception of a virtual character. Following previous work (e.g., [McDonnell et al. 2012; Ho and MacDorman 2010]) we employ (subsets of) the following scales for our experiments. The descriptions below are the ones given to the participants of the perception studies:

- *Extremely unappealing—Extremely appealing*: High appeal means that the character is one that is pleasant and you would like to watch more of. Unappealing means that you dislike to watch the character.
- *Extremely eerie—Extremely re-assuring*: Indicate if you find the character eerie, which means that they are gloomy and leave you with a sense of fear. Re-assuring means that the character restores a sense of security, confidence, calm in you.
- *Extremely abstract—Extremely realistic*: Indicate if you find the character's appearance to be highly stylized like in cartoons, or close to photo-realistic as in real pictures.
- *Extremely unfamiliar—Extremely familiar*: Indicate if you find the character's appearance familiar to you, in that you have seen something similar to it before, or if you find the character unfamiliar with an appearance that you haven't seen anything like before.
- *Extremely unattractive—Extremely attractive*: Indicate whether you find the character unattractive and ugly or beautiful and attractive.

We model these properties as Likert scales, which are popular in psychology as they allow subjective conditions such as the attitudes of participants to be measured. We chose a seven-point scale in order to give participants more response options and to allow for comparison to previous studies. The Likert scales were numbered 1–7, with a description provided on both ends of the scale.

Since both the design and the analysis of our experiments share many similarities, we describe the general setup now and later only mention deviations. The user's task and the rating scales were explained on a written document to the participants before the experiment. Afterwards all stimuli were presented in a random order and shown for 3 seconds each. The display was calibrated, 20" wide and at about 50 cm distance from the participants. The renderings have a resolution of 1024×768, corresponding to approximately $26.5\,\text{cm} \times 20.0\,\text{cm}$ on screen. After each stimulus presentation, participants were asked to rate it according to the above scales. In all experiments, the participants had normal or corrected-to-normal vision and were unaware of the final goal of the experiment. They were asked to report their 3D experience (how often they played video games, watched movies with visual effects, and how they would consider their knowledge of 3D graphics). We did not find any correlation between the reported 3D experience and the results of our tests, and thus omit this information for the rest of the paper.

For statistical analysis of each rating scale we conducted an n-way repeated measures Analysis Of Variance (rm-ANOVA). We run Maulchly's test for validating sphericity of the data, and whenever it is significant we report results with Greenhouse-Geisser correction applied and marked with an asterisk "*". Whenever main interaction effects were found, we conducted a Tukey Honestly Significant Difference (HSD) test for the comparison of means to further explore the results [Cunningham and Wallraven 2011].

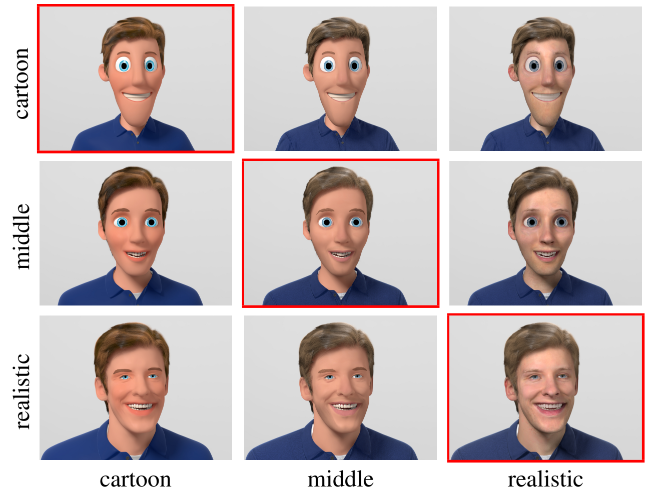


Figure 4: *Stimuli used in Experiment 1a: three levels of shape and material stylization, shown here for the happy expression. The baseline stimuli are shown on the diagonal. Their textures have been transferred to the other shapes for producing the off-diagonal stimuli of mismatching stylization levels for shape and material. Please refer to the supplemental material for the full set of stimuli used in the experiment.*

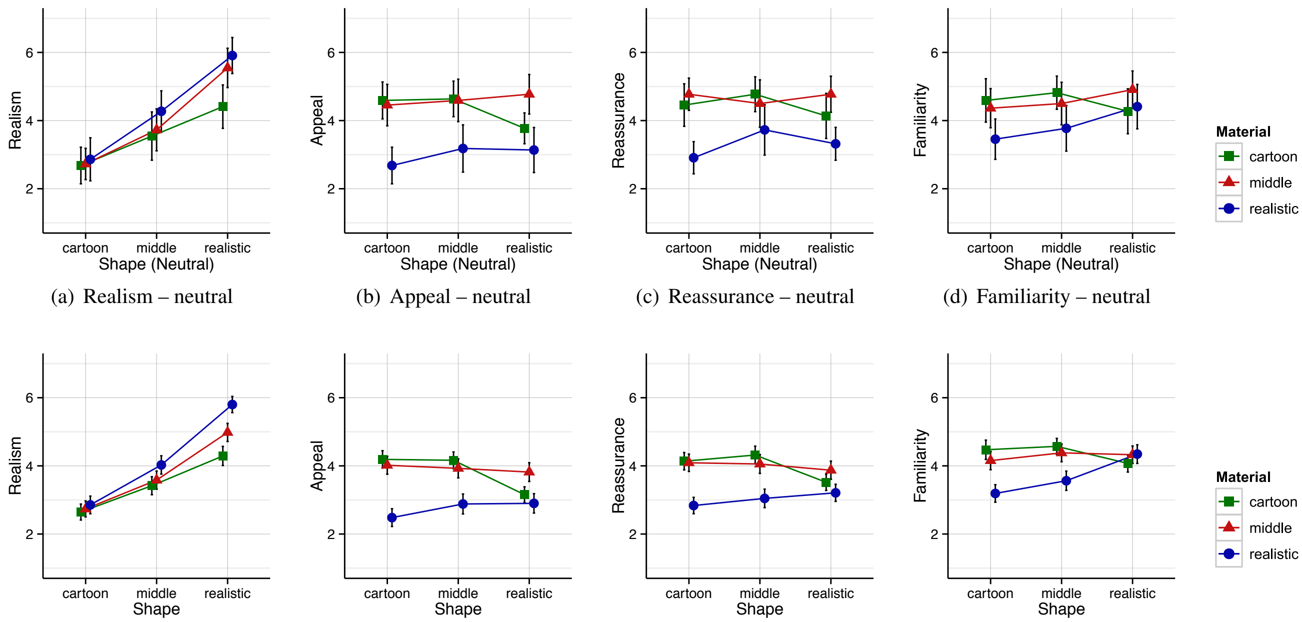
5 Importance of Shape, Material, Shading and Texture

The goal to investigate the influence of shape and material independently of the overall appearance of a CG character is motivated by differing design choices of recent animation films, ranging from cartoon shapes with cartoon materials (e.g., *Despicable Me*), to stylized shapes with realistic material (e.g., *The Adventures of Tintin*), to very realistic shapes and material (e.g., *Beowulf*). From a detailed analysis of character designs in commercial animation we identified three different recurrent stylization levels, which we denote by *cartoon*, *middle*, and *realistic*, where *Cloudy with a Chance of Meatballs* and *Toy Story* act as references for the two stylized versions, respectively.

5.1 Experiment 1a: Shape and Material

We first investigate the influence of shape and material, where we denote by material the combination of shader, shader parameters, and textures. The combination of each material with each shape style leads to a total of nine different versions of the character, times five different expressions, resulting in a set of 45 stimuli. Figure 4 shows the 3×3 stimuli for the happy expression; all other stimuli can be found in the supplementary material. We analyze the interaction between shape and material for the scales most frequently used in previous work: realism, appeal, reassurance, and familiarity. Twenty-two volunteers participated in this first experiment: 14 female, 8 male, with age from 19 to 30 years (avg. 24.5).

In this section, we analyze the effects of shape and material only. Figure 5 compares the ratings of the neutral expression with averaged ratings over all expressions. Despite a smaller offset and some noise, ratings for different expressions have been very consistent, which justifies averaging over all expressions. For statistical analysis, a rm-ANOVA with three factors (shape, material, and expression) was used. We refer the reader to the supplementary material for more per-expression results.



(a) Realism – neutral (b) Appeal – neutral (c) Reassurance – neutral (d) Familiarity – neutral

(f) Realism – all (g) Appeal – all (h) Reassurance – all (i) Familiarity – all

Figure 5: *Results of Experiment 1a: Ratings for perceived realism, appeal, reassurance/eeriness, and familiarity, for different shape and material stylizations. (a)–(d) show results for the neutral expression only, (f)–(i) are averages over all expressions. Error-bars denote 95% confidence levels. Individual per-expression results are shown in the supplementary material and discussed in Section 7.*

Realism A main effect was found for shape ($F(2, 42) = 113.18$, $p < 0.0001$) and material ($F^*(1.47, 30.82) = 23.15$, $p < 0.0001$, $\epsilon = 0.734$), as well as for the interaction between shape and material ($F(4, 84) = 11.14$, $p < 0.0001$). Post-hoc tests show that the cartoon shape was perceived as least realistic, no matter which material was used. Similarly, cartoon and middle materials did not make a difference for the middle shape (Figure 5a), while the realistic material caused a more realistic perception for this shape ($p < 0.002$ for both comparisons). In contrast, all material levels differ significantly for the realistic shape ($p < 0.0002$). Interestingly, the most stylized shape does not reach the bottom of the realism scale, revealing that there is more potential for abstraction.

Appeal We found a main effect of material on the ratings of appeal ($F^*(1.41, 29.67) = 42.69$, $p < 0.0001$, $\epsilon = 0.706$), but no main effect of shape was found. An interaction between shape and material ($F(4, 84) = 13.97$, $p < 0.0001$) shows that a realistic material on a cartoon shape yields the least appealing combination, since a post-hoc analysis showed significantly lower ratings for this combination compared to all others ($p < 0.02$ in all cases). The realistic material is less favored on the middle shape as well, and the cartoon material on the realistic shape is similarly unappealing ($p < 0.02$ in all cases except the combinations mentioned above). These results (see Figure 5b) suggest that material contributes most to the perceived appeal of a CG character, and that strong mismatches in the level of stylization of shape and material can result in very unappealing characters. Furthermore, the middle shape was rated as equally appealing regardless of material, which could be due to the fact that it was never strongly mismatched with material. Our appeal ratings ranged from 2.5 to 4.2, which is similar to the appeal ratings reported by McDonnell et al. [2012] for their static images.

Reassurance Similar to the appeal ratings, we found a main effect of material on the ratings of reassurance ($F^*(1.51, 31.70) = 49.07$, $p < 0.0001$, $\epsilon = 0.755$), but no main effect was found for shape. An interaction between shape and material is present ($F(4, 84) = 12.02$, $p < 0.0001$) and post-hoc analysis showed significantly lower ratings of reassurance especially in shape-material combinations that reduce appeal as well—realistic materials on all shape levels and cartoon materials on the realistic shape ($p < 0.02$). The realistic material on the cartoon and middle shape was perceived most eerie. A Cronbach's alpha value of $\alpha = 0.88$ confirms high similarity between the appeal and the reassurance scale (see Figure 5 b,c).

Familiarity Again, a main effect has been found for material ($F(2, 42) = 12.58$, $p < 0.0001$), but not for shape. Furthermore, there is also a significant interaction between shape and material ($F(4, 84) = 17.99$, $p < 0.0001$). The results of the post-hoc test for familiarity are less similar than between the appeal and eeriness ratings. Even though the combination of realistic material and realistic shape is unappealing and eerie, it was not rated significantly less familiar than other combinations. Realistic materials on cartoon and middle shapes result in the least familiar combinations ($p < 0.02$ in all cases). See Figure 5d.

5.2 Experiment 1b: Shading and Lighting

The above experiment reveals a strong influence of material, in particular on the appeal and reassurance ratings. The realistic material was rated as the least appealing for all character shapes, while the middle material was the most appealing for the realistic shape. Materials are controlled by a large number of shader parameters, and testing each of them is infeasible. In addition, only certain parameter combinations are meaningful and would be used in a real-world scenario. We note that all shader parameters are mainly responsible

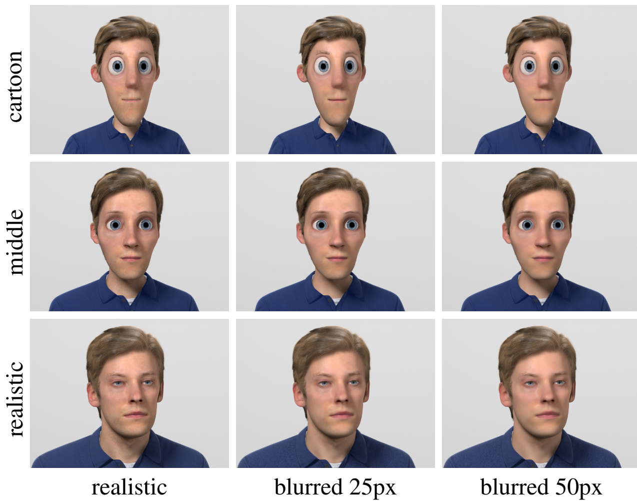


Figure 6: *Stimuli for Experiment 1c: Realistic material with realistic texture and two variants with blurred textures (Gaussian kernels of 25 and 50 pixels), for the three shape stylizations.*

for light-material interaction, while albedo textures control primarily the color. Instead of varying certain shader parameters within certain ranges, we modify the light transport more drastically by altering shading and lighting technique.

In an experiment similar to the previous one, we tested the initial baseline characters (three matching shape/material stylization, five expressions) with two different shaders and three illumination methods. For shading we tested a simple Phong shader in addition to the sophisticated skin shader. The lighting categories were (i) global illumination and soft shadows, (ii) ambient light and soft shadows, (iii) ambient light and hard shadows. All questions and scales were the same as for the previous experiment.

Twenty new volunteers participated in this second experiment (15 female, 5 male, ages from 19 to 30 years). A rm-ANOVA with three factors (shading, lighting, expression) was used for statistical analysis. While there was a main effect of lighting on realism ($F(2, 38) = 6.66, p = 0.003$), with global illumination being rated more realistic than soft shadows ($p = 0.020$) and hard shadows ($p = 0.004$), the difference was very small (means are 3.95 ± 0.1). Besides the effect of lighting on realism, we did not find any other significant effects, neither for the other scales nor for the different shaders. These results suggest that textures have more influence than shader parameters on appearance, and therefore we explore them more in depth in the following.

5.3 Experiment 1c: Texture

One possible explanation of why the middle material was rated the most appealing for the realistic shape could be the reduced pigmentation variation as reported by Fink and Matts [2008], as discussed in Section 2. In order to analyze whether their findings on attractiveness can also explain our effects on appeal and reassurance, we designed a variation of the first experiment of Section 5.1. Our third experiment should then:

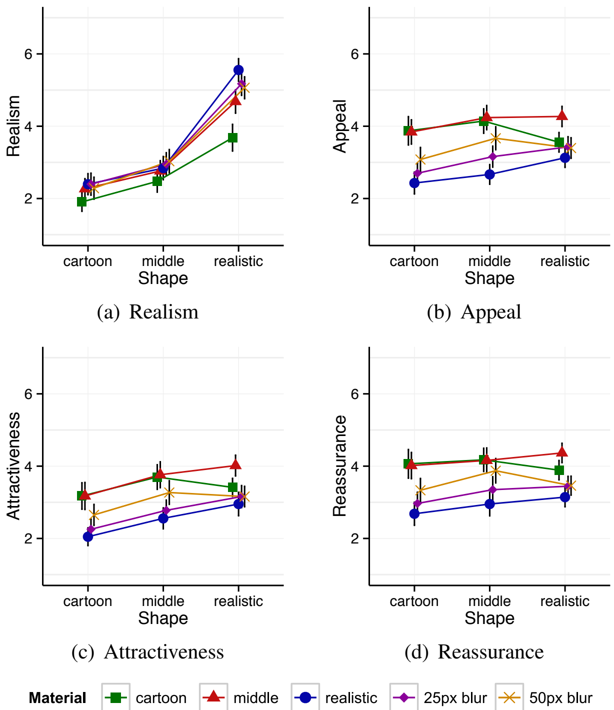
- test whether it is possible to influence appeal or realism by changing only the albedo texture,
- show a possible correlation between attractiveness and appeal/reassurance, and
- reveal whether appeal can be increased without sacrificing realism too much, simply by filtering a photo-realistic texture.

To this end, we created two additional textures with reduced skin details by applying uniform Gaussian blur of kernel sizes 25 and 50 pixels (for 4k textures), respectively. The 50px kernel covers barely 1 cm of the face, which translates into around four pixels in image-space. Lips and skin were filtered independently in order not to blur the boundary inbetween; eyebrows were not filtered. These three textures (realistic, blurred 25px, blurred 50px) were used in combination with the realistic material. To enable a comparison with Experiment 1, we also included the cartoon and middle materials (with their original textures only). This results in a set of 5 materials, which were also transferred to the middle and cartoon shapes, as shown in Figure 6.

For this experiment we tested these 5 materials on the 3 shape stylizations, but used the neutral expression only, leading to 15 stimuli in total. Note that the three realistic materials differ in their (blurred) texture only. The presentation of the stimuli was repeated three times with different random orderings. After each stimulus, participants were asked to rate it according to the previously described scales for appeal, reassurance, and realism, plus a new scale *attractiveness*. Twenty-one new volunteers (13 female, 8 male), average age 24.6 years, participated in the experiment. For statistical analysis, a rm-ANOVA with three factors (shape, material, and expression) was used. All results from Section 5.1 were confirmed, and thus we only describe the main effects related to the added material levels.

Realism Although a main effect was found for shape ($F^*(1.29, 25.78) = 124.98$, $p < 0.0001$, $\epsilon = 0.645$), material ($F(4, 80) = 17.52, p < 0.0001$) and an interaction between shape and material ($F(12, 240) = 6.42, p < 0.0001$), the post-hoc shows that this is not related to the added textures. The ratings for the two blurred textures are between the realistic and the middle texture, but are not significantly different for any shape. This confirms our initial assumption that blurring a realistic texture only slightly reduces the perceived realism of a character.

Appeal and Attractiveness Due to the high similarity between appeal and attractiveness (Cronbach's $\alpha = 0.87$) we report these results together. A main effect was found for shape for attractiveness ($F^*(1.33, 25.54) = 5.36$, $p = 0.021$, $\epsilon = 0.665$) but not for appeal. Material was significant in both cases (Appeal: $F^*(1.68, 33.60) = 27.17$, $p < 0.0001$, $\epsilon = 0.421$; Attractiveness: $F^*(1.56, 31.26) = 16.72$, $p < 0.0001$, $\epsilon = 0.391$). The interaction between shape and material is significant (Appeal: $F^*(7.05, 94.03) = 4.99$, $p < 0.0001$, $\epsilon = 0.588$; Attractiveness: $F(12, 240) = 2.88$, $p = 0.005$). As we hypothesized, the blurred textures were rated higher than the realistic texture. This effect is stronger for the cartoon and middle shapes and a significant difference between the realistic and 50px blurred version was found ($p < 0.003$ in all cases). For other comparisons between the blurred and realistic textures no significant difference was found. However, the graphs in Figure 7 show that the two blurred textures were rated equally appealing for the realistic shape. In contrast, a stronger blur is preferred for cartoon and middle shapes. We therefore conclude that blurring realistic skin textures is a reasonable approach for increasing appeal or attractiveness, without losing too much realism. Although the results of our tests are not significant in some cases, these findings are in line with research of Fink and Matts [2008]: We generalize their findings to character shapes of different stylization levels.



(a) Realism (b) Appeal (c) Attractiveness (d) Reassurance

Figure 7: *Results of Experiment 1c: While there is nearly no difference between the realistic and blurred textures for the realism scale, the blurred textures increase appeal and attractiveness and reduce eeriness.*

Reassurance Although the graphs of reassurance and appeal are similar (Figure 7; $\alpha = 0.89$), a main effect was found for material only ($F^*(1.44, 28.72) = 24.55$, $p < 0.0001$, $\epsilon = 0.359$), but not for shape. In addition, there is an interaction between shape and material ($F^*(7.128, 142.46) = 2.66$, $p = 0.029$, $\epsilon = 0.594$). The two blurred textures have been rated less eerie than the realistic version. Significant differences have been found between the realistic texture and the 50px blurred version for cartoon and middle shapes ($p < 0.0001$). Thus, blurring a texture does not only increase appeal, but also reduce eeriness.

5.4 Conclusion

The three tests described above allow us to draw the following main conclusions on the tested dimensions:

- Shape is the main descriptor for realism, while material is more important for perceived appeal, reassurance, and attractiveness. Strong mismatches in stylization between material and shape affect negatively the appeal and attractiveness of the characters and make them more eerie.
- Texture has stronger influence on appeal and attractiveness than shading or illumination models. Blurring a realistic texture does not significantly reduce realism but increase appeal and attractiveness.
- Ratings for appeal, reassurance, and attractiveness measure similar concepts ($\alpha > 0.87$ in all experiments), but do not correlate with the realism scale ($\alpha < 0.5$ in all experiments).

6 Experiment 2: Further Investigation of Shape and Material

The experiments in Section 5 indicate that different stylization levels of material and shape have a big impact on perceived appeal or realism. However, our set of stimuli contained only a single character, and the realism scale was not densely sampled. A more stylized character might reveal that big mismatches between material and shape cause unappealing results, or a stylization level between middle and realistic might cause uncanny reactions. To allow for a more generalized conclusion about different stylization levels, further investigation is required.

In the following experiment we analyze the effect of varying stylizations on shape and material, including matching and mismatching levels of stylization, on a significantly extended set of stimuli. In particular, we seek answers to the following questions:

- Can our findings be observed on other characters as well?
- Does a strong mismatch between material and shape create unappealing results only for realistic shapes or for all shapes?

Stimuli We extended our initial stimuli with another character of different gender, because this adds by design a clearly distinctive person. For each character, two additional stylizations were created, yielding five stylization levels from *level 0* (most stylized) to *level 4* (highly realistic). We distinguish between stylizations in material and shape by using the prefix m and s respectively. The new stylizations (level 0 and level 3) have been particularly designed by the artists to fill the gaps for perceived realism in the stylization scale. For these levels our character designs are inspired by *Pocoyo* and *Tangled*. We also changed the hairstyle of the virtual male character in order to allow a better comparison with a photograph of the actor. This provides us with baseline ratings on appeal and realism for the real person. The new set of stimuli is composed of 2 characters times 5 shape stylizations times 5 material levels times 5 expressions, leading to a total of 250 images. A representative subset of the stimuli is shown in Figure 8, for the five expressions and matching shape/material levels of the male character (left), and the 25 combinations of material and shape for the female character (right).

Procedure The largely extended stimuli require a reduction of the scales in order to keep the experiment tractable. Given that the appeal, reassurance, and attractiveness scales measure similar concepts, and that the familiarity scale did not provide much information, we decided to keep only the realism and appeal scales for this experiment. Furthermore, we increased the display time of the stimuli to 4 s, and showed the neutral male and female baseline characters before the experiment, such that participants could better estimate the range of characters from the beginning on. At the end of the actual experiment, participants rated a photograph of the real characters in neutral expression. The rest of the experiment remains similar to the previous one. With all these changes, participants finished the experiment within 50 minutes or less. Twenty-one new different volunteers (17 female, 4 male) took part, average age 23.4 years.

Our results are summarized in Figure 9 and are mostly consistent across male and female. Repeated measures ANOVA with four factors (character, shape, material, and expression) was used for statistical analysis. Differences between the two characters were significant, but since they were rather small and/or inconsistent, we exclude them from further analysis. In the following we present an in-depth discussion of the realism and appeal ratings, and report the impact of expression in Section 7.

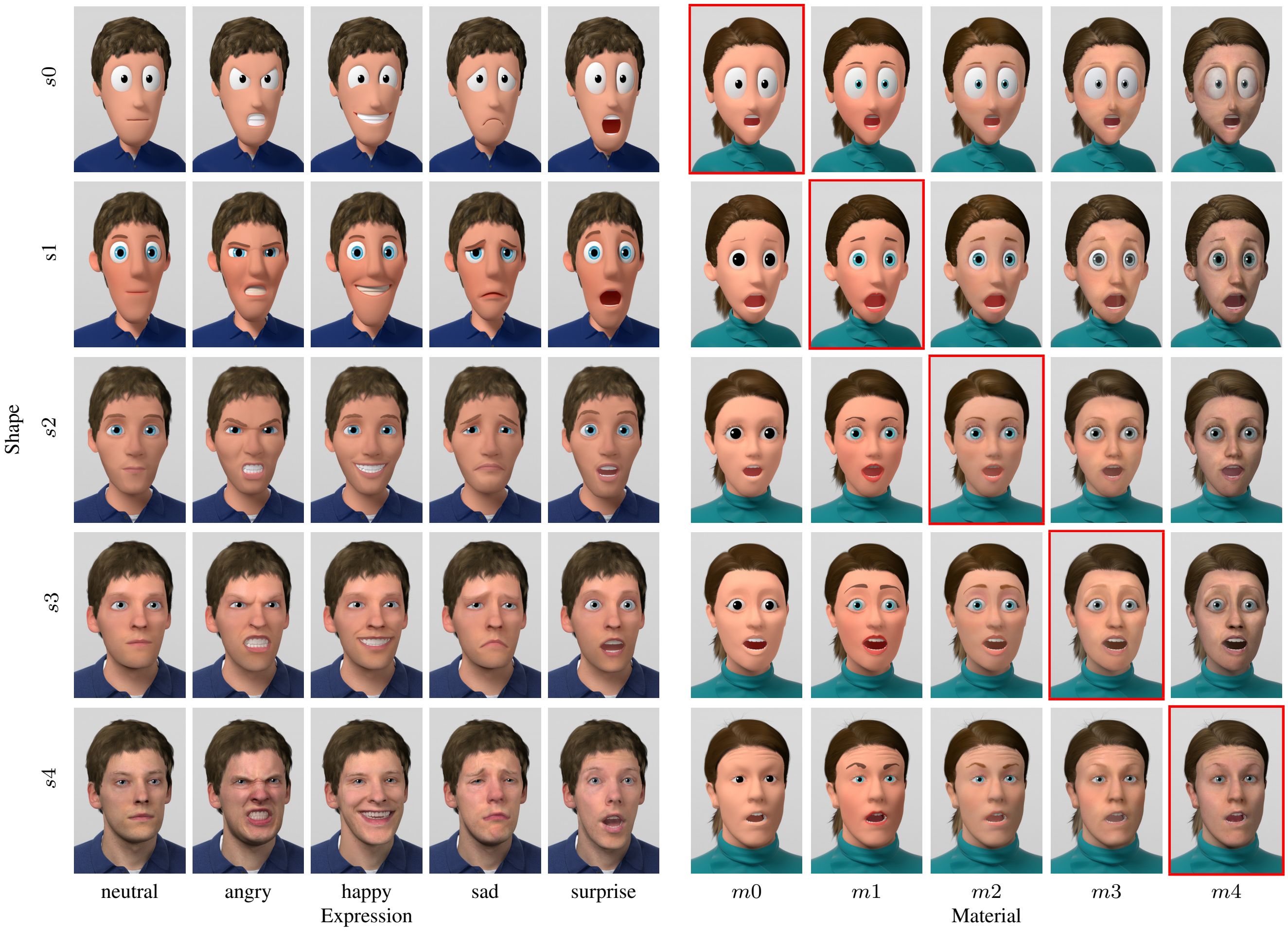


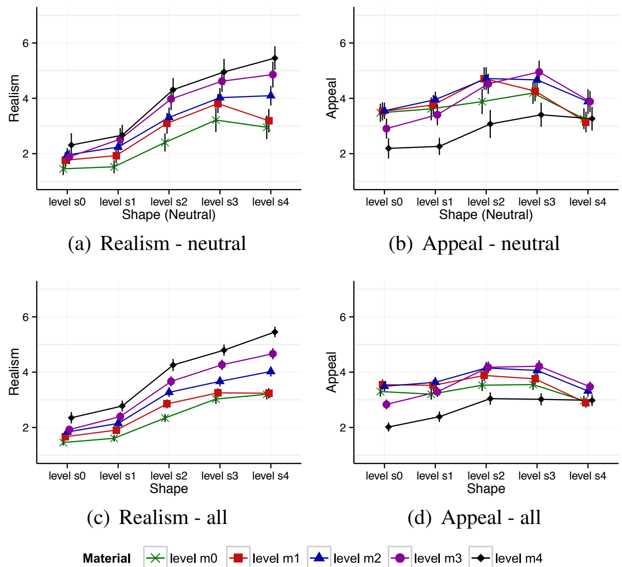
Figure 8: *Stimuli for Experiment 2:* Left: *Renderings of the male character for different stylizations (rows) and basic emotions (columns).* Right: *Combinations of shape and material stylization for the female character (surprise expression), with baseline stimuli on the diagonal.*

Realism A main effect of shape ($F^*(1.98, 39.6) = 178.67$, $p < 0.0001$, $\epsilon = 0.495$) and material ($F^*(1.33, 26.4) = 73.92$, $p < 0.0001$, $\epsilon = 0.333$) was found as well as an interaction between shape and material ($F^*(6.71, 134.1) = 11.59$, $p < 0.0001$, $\epsilon = 0.419$). Post-hoc analysis shows that all shapes ($p < 0.004$) and most of the materials ($p < 0.003$ except for level $m0$ and $m1$) differ significantly from each other. The 25 groups resulting from the combinations of shape and material differ also significantly in more than 80% of the cases. Most non-significant comparisons can be found for the shape level $s0$ (see Figure 9). For example, increasing the material from level $m1$ to $m2$ or from level $m2$ to $m3$ does not cause a significant difference. This contrasts with the case of the realistic shape levels $s3$ and $s4$ ($p < 0.002$). This is in line with the results from Section 5.1, and confirms that as the shape becomes more realistic, the material stylization becomes more dominant for perceived realism.

Appeal The main effects of shape ($F^*(2.58, 51.6) = 20.97$, $p < 0.0001$, $\epsilon = 0.645$) and material ($F^*(1.88, 37.6) = 20.39$, $p < 0.0001$, $\epsilon = 0.470$) are comparable. There is a slightly weaker interaction between shape and material ($F^*(6.06, 121.3) = 14.29$, $p < 0.0001$, $\epsilon = 0.379$). Post-hoc analysis reveals that shape levels $s2$ and $s3$ were perceived more appealing than the other shape levels ($p < 0.0002$ in all cases between the two groups).

For the materials, only the most realistic version (level $m4$) was significantly less appealing than all other materials ($p < 0.0002$). This supports our assumptions from Section 5.3 that smooth(ed) skin pigmentations are perceived more appealing. For the abstract shape $s0$, material levels $m0$, $m1$, and $m2$ form a cluster without any significant difference; this cluster is found significantly more appealing than material levels $m3$ and $m4$ ($p < 0.03$). On the other hand, shape level $s3$ is rated significantly higher with matching material levels ($m2$ and $m3$), with both more stylized ($m0$ and $m1$) and more realistic ($m4$) materials being rated significantly lower. These results support that in all cases a strong mismatch between shape and material is perceived as unappealing.

Photograph At the end of the experiment, participants rated a photograph of the real actor in neutral pose. As expected, the average realism rating is very high (6.98, $SD = 0.15$). The average appeal rating was 4.5 ($SD = 1.40$), which is higher than the average ratings for the realistic *s4/m4* characters (3.26, $SD = 1.33$). This dip in appeal rating for the *s4/m4* character is in agreement with the uncanny valley theory [Mori et al. 2012]. However, appeal for stylizations *s2/m2 and s3/m3* (4.71, $SD = 1.25$ and 4.95, $SD = 1.25$) were rated highest. In addition, Figure 12 depicts that realism *alone* is a bad predictor for appeal; instead, our results show that the compatibility of shape and material stylizations, i.e., their matching degrees of realism, has a stronger (and predictable) influence on appeal.



(a) Realism - neutral (b) Appeal - neutral

(c) Realism - all (d) Appeal - all

Figure 9: *Results of Experiment 2: Ratings for perceived realism and appeal for different shape and material stylizations. Upper row: neutral expression averaged over male and female characters. Bottom row: averaged over all expressions and characters.*

7 Experiment 3: Effect of Expressions

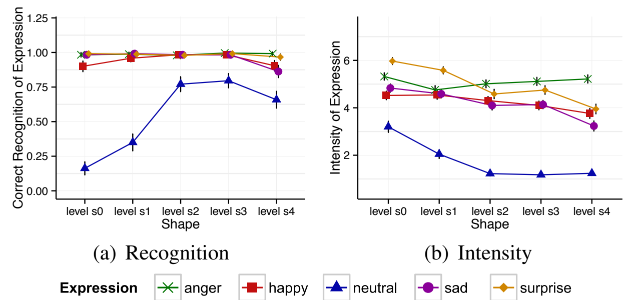
In previous experiments, we have analyzed the overall effect that shape and material have on the perception of faces. Here, we first analyze whether different levels of stylization in shape and material, including mismatches between them, affect the recognition and intensity of expressions (*anger*, *happy*, *neutral*, *sad* and *surprise*). We then discuss how ratings are affected by particular expressions (Figure 8). This is interesting since previous findings suggest that valence of emotion affects character perception [Calder 1996; Tinwell et al. 2011a], making negative expressions to be rated less appealing than positive expressions. In particular, we seek answers to the following questions:

- Does the level of stylization affect the intensity of expressions? Are they easier or more difficult to recognize?
- Do negative expressions affect the perceived appeal of characters? Is this influenced by stylization of shape or material?

7.1 Intensity and Recognition of Expressions

As discussed previously, stylization is a well-known tool for artists to enhance the expressivity of 3D characters, removing unnecessary details and enhancing specific features. In this experiment we explore how the different stylizations of shape and material affect *recognition* and the perceived *intensity* of the expressions, and which of the two dimensions is dominant for expression recognition. The extended 250-stimuli set from Experiment 2 is used again.

Each stimulus was presented for 4 seconds in random order; participants were first asked to classify the expression according to the following options: anger, happy, neutral, sad, surprised. After each answer (except for neutral), a follow-up question asked to rate the expression intensity with respect to a seven-point Likert scale bounded by *extremely low* and *extremely high intensity*. When participants rated an expression as neutral, its intensity was set to the lowest value. Twenty-four new volunteers (16 female, 8 male,



(a) Recognition (b) Intensity

Figure 10: *Results of Experiment 3: Effect of shape on the recognition and intensity of the expression. All expressions, except neutral, have been recognized well or outstandingly well independent of the shape. However the intensity reduced continuously with higher shape stylization levels.*

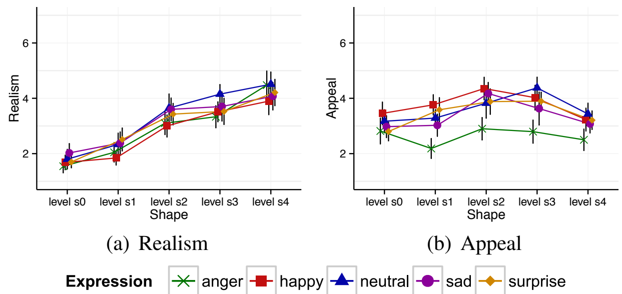
23.6 years old on average) took part in this experiment. Results are shown in Figure 10 and again a rm-ANOVA with four scales (character, shape, material, and expression) was used for statistical analysis.

Recognition We found a main effect of expression ($F^*(1.22, 28.04) = 74.00$, $p < 0.0001$, $\epsilon = 0.305$), as well as several interaction effects between expression and shape ($F^*(4.56, 104.9) = 41.3$, $p < 0.0001$, $\epsilon = 0.285$), texture ($F(16, 368) = 4.97$, $p < 0.0001$) and model ($F^*(2.3, 51.26) = 4.23$, $p = 0.016$, $\epsilon = 0.557$). The neutral expression is mainly responsible for all these effects; its recognition rate was lower ($p < 0.002$) than the other expressions, varying strongly across different shape levels. This neutral expression was in general poorly recognized for the more stylized shapes ($s0$ and $s1$): For instance, some participants reported that the big round eyes made them look surprised. This might be explained by the fact that cartoons are usually designed to enhance expressivity, not to be posed displaying a neutral emotion.

We also found a main effect for shape ($F(4, 92) = 44.23$, $p < 0.0001$), which is mainly determined by the neutral expression, as discussed above, and a main effect for material ($F(4, 92) = 10.09$, $p < 0.0001$). The material level $m4$ reduced the recognition rate significantly ($p < 0.015$) but only by 2%.

Intensity Main effects of shape ($F^*(2.11, 48.61) = 91.40$, $p < 0.0001$, $\epsilon = 0.528$) and material ($F^*(2.47, 56.90) = 30.46$, $p < 0.0001$, $\epsilon = 0.618$) were found. Apart from the angry expression, the perceived intensity of expressions is continuously reduced with increasing realism of shape ($p < 0.0002$). Only in the case of shape levels $s2$ and $s3$ does the intensity remain constant. In the case of material, the absolute difference was very small (0.5 between the lowest and highest mean) and only the material level $m4$ had a higher intensity ($p < 0.0002$). This matches previous research [Wallraven et al. 2007; Wallraven et al. 2008], which found that details such as wrinkles increase the expressivity of realistic characters, although in our case the effect is weaker.

In addition, a main effect of expression ($F^*(2.57, 59.10) = 204.6$, $p < 0.0001$, $\epsilon = 0.642$) and interactions between shape and expression ($F^*(5.78, 132.94) = 19.00$, $p < 0.0001$, $\epsilon = 0.361$), material and expression ($F(16, 368) = 5.04$, $p < 0.0001$), and expression and model ($F(4, 92) = 19.55$, $p < 0.0001$) were found. In particular, the happy, sad, and surprise expressions are perceived with lower intensity as the realism of shape increases. This difference is significant in the majority of cases for shape levels $s3$ and $s4$



(a) Realism (b) Appeal

Figure 11: *Results of Experiment 2: While emotions do not differ in realism, the anger expression was perceived as more eerie and unappealing for all stylization levels.*

($p < 0.01$), but is less frequent for lower shape levels. The perception of the angry expression, on the other hand, remains constant along shape abstractions.

Overall we found that expressions of cartoon shapes are perceived as more intense, which confirms that properly stylizing features helps increase expressivity. The neutral expression is hard to read for very stylized character shapes, which suggests that low-intensity subtle expressions are harder to convey in abstract characters designed to enhance expressivity. Additionally, we found no or small impact of material on the intensity or expression recognition, which indicates that shape is the dominant dimension when designing expressive characters.

7.2 Effect of Expression on Realism and Appeal

In our previous experiments on material and shape with the five basic expressions (Section 5), we found that appeal and eeriness measure similar concepts, while effects for familiarity were generally small. We focus here on the effect of expressions on realism and appeal with the extended stimuli set. The rest of the analysis can be found in the supplemental material. Figure 11 shows the results, which we analyze below.

Realism A main effect of expression ($F(4, 80) = 10.38$, $p < 0.0001$) was found, which could be mainly attributed to the neutral and sad expressions, which have been perceived as more realistic ($p < 0.006$). Because the means are located within a small range (± 0.16), we classify this effect as noise and omit similar examples for the rest of this section. Nevertheless, equal realism ratings confirm that expressions were well designed by the artists.

Appeal A main effect was found for appeal ($F^*(1.56, 31.36) = 19.34$, $p < 0.0001$, $\epsilon = 0.392$), which is primarily caused by the anger expression ($p < 0.001$). These results reveal that anger is rated much lower with respect to appeal. Previous studies reported that negative emotions trigger unpleasant responses from the observers [Calder 1996]; our results confirm these studies. Moreover, this effect is maintained even in the presence of highly stylized and appealing characters, suggesting that negative expressions are perceived as unappealing independent of stylization level.

Additionally, we observe that ratings are unsteady across different stylization levels for the rest of the expressions. In many cases, interaction effects between expression and shape or material are found with $p < 0.0001$. Zhu et al. [2014] showed with photographs that different instances of the same expression do indeed vary in perceived appeal. We believe that this might also be the primary

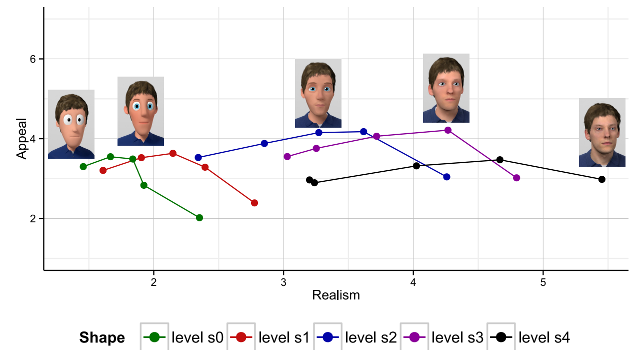


Figure 12: *Participant ratings for our male stimuli plotted on a realism-vs-appeal diagram, similar to [Tinwell et al. 2011b; MacDorman et al. 2009]. Each graph corresponds to one shape stylization, while graph nodes correspond to material levels. The icons are placed above the nodes of matching shape/material levels. The diagram reveals that perceived realism is a bad predictor for appeal. Instead, it is the compatible degree of realism of both shape and material that matters.*

reason for the variations in our ratings. We rule out recognition as an error source, since all the expressions were recognized outstandingly well (see previous subsection).

8 Discussion

Shape and material are two of the main aspects that define the appearance of virtual characters, which in turn are crucial when defining the visual look of animated feature films. We have analyzed the perceptual effects of different stylizations along these dimensions on computer-generated faces. In particular, we have studied five different stylizations of two virtual characters—male and female—ranging from very realistic to highly stylized, varying both the shape and the material.

Our results show that the main contributor for the perceived *realism* is shape, and the effect of material stylizations grows when shape realism is increased. This implies that mismatches in material and shape are less prominent on abstract characters. The resulting asymmetry is shown in Figure 9 (a,c), where the curves spread out as the level of realism increases.

On the other hand, we have found that material is the main factor for perceived *appeal*, specifically the albedo texture. In general, *appeal*, *attractiveness*, and *eeriness* are highly dependent on the material stylization. Matching levels of stylization of geometry and material cause the highest ratings of appeal, while strong mismatches (e.g., very realistic material on a stylized shape) result in unappealing and eerie characters.

Interestingly, as shown in Experiment 1c and later backed-up in Experiment 2, subtle stylization of a realistic material (edge-preserving blur on the albedo texture) increases appeal without sacrificing realism. These stylizations de-emphasize unwanted skin impurities, pores, and wrinkles, and our results are in accordance to empirical knowledge regarding the effect of makeup. Moreover, our results relate with previous findings on face perception showing that smooth homogeneous skin is generally rated more attractive, since it is a good estimate of a young and healthy subject [Fink et al. 2012]. However, this trend is only observed for mild stylizations, and stronger ones quickly reduce realism.

Our results are consistent across all tested *expressions*, except for anger, which was consistently rated less appealing and more eerie. This can be explained by negative or aggressive expressions triggering a defense response and a negative reaction of the viewer [Calder 1996]. Our results are also consistent between different *characters*. Although small differences between the characters exist, all reported trends are consistent and well visible.

Realism alone was shown to be a bad predictor for appeal (Figure 12), which is not well aligned with the theory of the uncanny valley, although a similar finding was reported for rendering style [McDonnell et al. 2012]. One possible explanation is that some of our characters were difficult to categorize by the participants, due to their mismatched appearance parameters [Saygin et al. 2012; McDonnell et al. 2012].

Finally, our experiments show how stylization affects the *intensity of expressions*, and that *shape* is the main factor in this case, whereas material has no significant influence for stylized shapes. This confirms previous knowledge on modeling or drawing expressive stylized characters, where expressivity is mainly determined by the global shape of the character. However, for realistic shapes, we have observed that material stylization slightly, but significantly, reduces the perceived intensity of expressions. Another possible explanation, which also merits further investigation, is that realistic characters make *suspension of disbelief*[1] harder to maintain, and therefore observers find it more difficult to emotionally connect with the virtual character. These results are consistent with previous work [Wallraven et al. 2007; Wallraven et al. 2008] and may explain the conscious disturbing effect of stylizing hyper-realistic characters in some movies (e.g., *A Scanner Darkly* or *Renaissance*).

8.1 Limitations and Future Work

As in all user studies, our results are only strictly valid for our particular set of stimuli. We have focused on a specific set of stylizations for two realistic characters, varying shape and material following typical designs used in feature animation. This of course limits the universality of the conclusions, which may not generalize if the character styles differ greatly from ours. However, since our design space was densely sampled and the observed trends are consistent between the different characters, we believe that our observations can be used as valid guidelines for creating digital characters within a reasonable range of styles.

In our statistical analysis, we employed a common significance threshold of $p < 0.05$. With the amount of results we report, it might be that some significant results are false positives. Because we only focus on clear, reoccurring trends, and since many significances have $p < 0.001$, it is unlikely that one of our main conclusions is a false positive.

Investigating the effect of realistic and stylized animation was outside the scope of this paper, due to the number of stimuli combinations that would have to be tested, and the technical complexity of creating a scale of stylized animations. Previous work has shown no difference in ratings for realism, and only small differences in appeal ratings for static or motion-captured characters [McDonnell et al. 2012]. However, we would expect a more complex interaction between motion and appeal when combining characters and animations of different levels of stylization. Therefore, a full investigation of motion stylization is an important future direction.

Finally, note that we analyzed clear *peak* expressions, avoiding the less attractive transitions between expressions common in the real world [Zhu et al. 2014]. Evaluating the impact of these transitions for different stylizations could be also an interesting avenue of future work.

In summary, we believe that our work provides many interesting insights for the creation of virtual characters, and offers a set of guidelines that we hope will help practitioners and inspire future research on the perception of virtual characters.

[1]In fiction, the suspension of disbelief is a semi-conscious decision by the viewer to accept as real what clearly is not. This allows her to connect with the story.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable suggestions, Kaan Kucuk, Fiona Healy, Sergio Garcia and Jorge Francos for character creation, and the actors for the scans. Eduard Zell and Mario Botsch were supported by the Cluster of Excellence Cognitive Interaction Technology “CITEC” (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG). Katja Zibrek and Rachel McDonnell were sponsored by Science Foundation Ireland (project Cartoon Motion). Carlos Aliaga, Adrian Jarabo and Diego Gutierrez acknowledge support from the Spanish Ministry of Science and Technology (project LIGHTSLICE) and the BBVA Foundation. Diego Gutierrez is additionally supported by a Google Faculty Research Award, and Carlos Aliaga by a grant from the Gobierno de Aragón.

References

ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, J.-Y., MA, W.-C., WANG, C.-C., AND DEBEVEC, P. 2010. The Digital Emily project: Achieving a photorealistic digital actor. *IEEE Computer Graphics and Applications 30*, 4.

ALLEY, T. R., AND CUNNINGHAM, M. R. 1991. Averaged faces are attractive, but very attractive faces are not average. *Psychological Science 2*, 2.

BEELER, T., BICKEL, B., BEARDSLEY, P., SUMNER, B., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics 29*, 4.

BENSON, P., AND PERRETT, D. 1992. Face to face with the perfect image. *New Scientist 133*, 1809.

BURLEIGH, T. J., SCHOENHERR, J. R., AND LACROIX, G. L. 2013. Does the Uncanny Valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior 29*, 3.

CALDER, A. J. 1996. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology 13*, 5.

CHEN, H., RUSSELL, R., AND NAKAYAMA, K. 2010. Crossing the “Uncanny Valley”: Adaptation to cartoon faces can influence perception of human faces. *Perception*, 39.

CUNNINGHAM, D., AND WALLRAVEN, C. 2011. *Experimental Design: From User Studies to Psychophysics*. AK Peters/CRC Press.

DILL, V., FLACH, L. M., HOCEVAR, R., LYKAWKA, C., MUSSE, S. R., AND PINHO, M. S. 2012. Evaluation of the Uncanny Valley in CG characters. In *Intelligent Virtual Agents*.

EKMAN, P. 1972. Universal and cultural differences in facial expression of emotion. *Proceedings of Nebraska Symposium on Motivation 19*.

FAN, S., NG, T.-T., HERBERG, J. S., KOENIG, B. L., AND XIN, S. 2012. Real or fake? Human judgments about photographs and computer-generated images of faces. In *SIGGRAPH Asia 2012 Technical Briefs*.

FERREY, A. E., BURLEIGH, T. J., AND FENSKE, M. J. 2015. Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Frontiers in psychology 6*.

FINK, B., AND MATTS, P. J. 2008. The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology 22*, 4.

FINK, B., GRAMMER, K., AND MATTS, P. J. 2006. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior 27*, 6.

FINK, B., BUNSE, L., MATTS, P. J., AND D'EMILIANO, D. 2012. Visible skin colouration predicts perception of male facial age, health and attractiveness. *International Journal of Cosmetic Science 34*, 4.

GELLER, T. 2008. Overcoming the Uncanny Valley. *IEEE Computer Graphics and Applications 28*, 4.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics 23*, 1.

GREEN, R. D., MACDORMAN, K. F., HO, C.-C., AND VASUDEVAN, S. 2008. Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior 24*, 5.

HANSON, D. 2005. Expanding the Aesthetics Possibilities for Humanlike Robots. In *Proceedings of IEEE Humanoid Robotics Conference, special session on the Uncanny Valley*.

HO, C.-C., AND MACDORMAN, K. F. 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior 26*, 6.

JACK, D. R., GARROD, M. O., AND SCHYNS, P. 2014. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology 24*, 2.

JIMENEZ, J., ZSOLNAI, K., JARABO, A., FREUDE, C., AUZINGER, T., WU, X.-C., VON DER PAHLEN, J., WIMMER, M., AND GUTIERREZ, D. 2015. Separable subsurface scattering. *Computer Graphics Forum*.

KLEHM, O., ROUSSELLE, F., PAPAS, M., BRADLEY, D., HERY, C., BICKEL, B., JAROSZ, W., AND BEELER, T. 2015. Recent advances in facial appearance capture. *Computer Graphics Forum (Proceedings of Eurographics) 34*, 2.

LITTLE, A. C., AND HANCOCK, P. J. 2002. The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology 93*, 4.

MACDORMAN, K. F., GREEN, R. D., HO, C.-C., AND KOCH, C. T. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior 25*, 3.

MACDORMAN, K. F. 2006. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the Uncanny Valley. In *Proceedings of ICCS/CogSci Long Symposium: Toward Social Mechanisms of Android Science*.

MCCLOUD, S. 1993. *Understanding Comics*. William Morrow Paperbacks.

MCDONNELL, R., BREIDT, M., AND BÜLTHOFF, H. H. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics 31*, 4.

MORI, M., MACDORMAN, K. F., AND KAGEKI, N. 2012. The Uncanny Valley [From the field]. *IEEE Robotics and Automation Magazine 19*, 2.

NITZSCHE, B., AND ROSE, K. 2011. *Bodyshop: The Photoshop Retouching Guide for the Face and Body*. Wiley.

SAYGIN, A., CHAMINADE, T., ISHIGURO, H., DRIVER, J., AND FRITH, C. 2012. The thing that should not be: Predictive coding and the Uncanny Valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience 7*, 4.

SCHNEIDER, E., WANG, Y., AND YANG, S. 2007. Exploring the Uncanny Valley with japanese video game characters. In *Situated Play: Proceedings of the Digital Games Research Association Conference*.

SEYAMA, J., AND NAGAYAMA, R. S. 2007. The Uncanny Valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments 16*, 4.

TINWELL, A., AND GRIMSHAW, M. 2009. Bridging the uncanny: An impossible traverse? In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*.

TINWELL, A., GRIMSHAW, M., NABI, D. A., AND WILLIAMS, A. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior 27*, 2.

TINWELL, A., GRIMSHAW, M., AND WILLIAMS, A. 2011. The Uncanny Wall. *International Journal of Arts and Technology 4*, 3.

TINWELL, A. 2014. *The Uncanny Valley in Games and Animation*. AK Peters / CRC Press.

WALLRAVEN, C., BÜLTHOFF, H. H., CUNNINGHAM, D. W., FISCHER, J., AND BARTZ, D. 2007. Evaluation of real-world and computer-generated stylized facial expressions. *ACM Transactions on Applied Perception 4*, 3.

WALLRAVEN, C., BREIDT, M., CUNNINGHAM, D. W., AND BÜLTHOFF, H. H. 2008. Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perceptiona 4*, 4.

WEISE, T., BOUAZIZ, S., LI, H., AND PAULY, M. 2011. Realtime performance-based facial animation. *ACM Transactions on Graphics 30*, 4.

ZELL, E., AND BOTSCH, M. 2013. ElastiFace: Matching and blending textured faces. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*.

ZHU, J.-Y., AGARWALA, A., EFROS, A. A., SHECHTMAN, E., AND WANG, J. 2014. Mirror mirror: Crowdsourcing better portraits. *ACM Transactions on Graphics 33*, 6.

ZIBREK, K., AND MCDONNELL, R. 2014. Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM Symposium on Applied Perception*.