



**HAL**  
open science

## The impact of stylization on face recognition

Nicolas Olivier, Ludovic Hoyet, Ferran Argelaguet, Fabien Danieau, Quentin Avril, Philippe Guillotel, Anatole Lécuyer, Franck Multon

► **To cite this version:**

Nicolas Olivier, Ludovic Hoyet, Ferran Argelaguet, Fabien Danieau, Quentin Avril, et al.. The impact of stylization on face recognition. SAP 2020 - ACM Symposium on Applied Perception, Sep 2020, Virtual, United States. pp.1-9, 10.1145/3385955.3407930 . hal-02951833

**HAL Id: hal-02951833**

**<https://inria.hal.science/hal-02951833v1>**

Submitted on 27 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The impact of stylization on face recognition

Nicolas Olivier  
Inria, Univ Rennes, CNRS, IRISA,  
France  
Interdigital, France

Ludovic Hoyet  
Inria, Univ Rennes, CNRS, IRISA,  
France

Ferran Argelaguet  
Inria, Univ Rennes, CNRS, IRISA,  
France

Fabien Danieau  
Interdigital, France

Quentin Avril  
Interdigital, France

Philippe Guillotel  
Interdigital, France

Anatole Lecuyer  
Inria, Univ Rennes, CNRS, IRISA,  
France

Franck Multon  
Inria, Univ Rennes, CNRS, IRISA,  
M2S, France



Figure 1: A face with various levels of two (top and bottom) non-human stylization, from low (left) to high (right). The original human face is on the left.

## ABSTRACT

While digital humans are key aspects of the rapidly evolving areas of virtual reality, gaming, and online communications, many applications would benefit from using digital personalized (stylized) representations of users, as they were shown to highly increase immersion, presence and emotional response. In particular, depending on the target application, one may want to look like a dwarf or an elf in a heroic fantasy world, or like an alien on another planet, in accordance with the style of the narrative. While creating such virtual replicas requires stylization of the user's features onto the virtual character, no formal study has however been conducted to assess the ability to recognize stylized characters. In this paper, we present a perceptual study investigating the effect of the degree of stylization on the ability to recognize an actor, and the subjective

acceptability of stylizations. Results show that recognition rates decrease when the degree of stylization increases, while acceptability of the stylization increases. These results provide recommendations to achieve good compromises between stylization and recognition, and pave the way to new stylization methods providing a tradeoff between stylization and recognition of the actor.

## CCS CONCEPTS

• **Computing methodologies** → Perception; Mesh models; • **Human-centered computing** → Virtual reality.

## KEYWORDS

Virtual Characters, Style Transfer, Facial Recognition, Perception

SAP '20, September 12–13, 2020, Virtual Event, USA

© 2020 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Symposium on Applied Perception 2020 (SAP '20)*, September 12–13, 2020, Virtual Event, USA, <https://doi.org/10.1145/3385955.3407930>.

## ACM Reference Format:

Nicolas Olivier, Ludovic Hoyet, Ferran Argelaguet, Fabien Danieau, Quentin Avril, Philippe Guillotel, Anatole Lecuyer, and Franck Multon. 2020. The impact of stylization on face recognition. In *ACM Symposium on Applied Perception 2020 (SAP '20)*, September 12–13, 2020, Virtual Event, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3385955.3407930>

## 1 INTRODUCTION

Digital humans are essential elements of the entertainment world. They are also becoming more and more commonplace in retail, sports, social media, education, health and many other fields. While generic characters are widely available, some applications in these fields would benefit from using digital personalized (stylized) representations of users, as they were shown to highly increase immersion, presence and emotional response [Fleming et al. 2016]. Such stylized characters can either be created by an artist, which involves a lengthy and costly creation process, or using a capture system, such as body or facial scanning techniques. For instance, the latter has been used by game studios to capture gamers' features and stylize in-game characters, such as capturing the player's face in NBA 2K15 (Take-Two Interactive, 2014) or the entire body as in Kinect Sports Rivals (Microsoft, 2014). While captured characters are by definition human-like, artist-made stylized characters are also very common place in animated movies, games and even web-based support services. Animation companies such as Disney, BlueSky and Pixar aim for highly stylized looks when designing characters, in accordance with the style of the narrative, e.g., cartoon movie, or fantasy story.

Depending on the target application, a real human character may therefore not be adapted to the visual style of a content or to a specific narrative, as one may want to look like a dwarf or an elf in a heroic fantasy world, or like an alien on another planet (e.g. Sigourney Weaver in Cameron's *Avatar*). However, with the development of digital personalized representations of the user, creating such virtual replicas requires stylization of the user's features onto the virtual character. In this paper, we consider stylization to be the process that from contents A and B, produces a content C similar to A, but with the style of B. In the case of a virtual character, "C similar to A" means that people could recognize A when watching C. However, despite the rising interest in stylizing virtual characters, no formal studies have been conducted to assess the ability to recognize stylized characters. Can we recognize the human face that has been stylized into a virtual character? Is there a limit in the stylization that can be applied?

In this paper, we therefore focus on the ability of recognizing a scanned actor's face (A) that has been stylized into the face of another non-human virtual character (B). In particular, we leverage the work of Danieau et al. [Danieau et al. 2019] to create stylized representations (C) of a number of actors, in order to evaluate the ability of viewers to recognize the original actor (A). In this context, we addressed the following questions: What is the effect of the degree of stylization on the ability of viewers to recognize an actor? What is the effect of the degree of stylization on the acceptability of the stylization by viewers (the judgement of the quality of the result in respect to the original style character)? E.g., do they consider the result of an orc stylization to be an orc character? Is the ability of viewers to recognize a stylized actor affected by how much non-human is the virtual character used for the stylization?

To answer these questions, we conducted a study (N=24) investigating the effect of the degree of stylization on the ability to recognize an actor, and the subjective acceptability of the stylization results. The face dataset used in the experiments was composed of 10 facial scans and 3 non-human virtual characters with varying

degrees of non-humanness. The stylized faces were generated using the method proposed by Danieau et al. [Danieau et al. 2019], slightly modified to enable the creation of different degrees of stylization for each actor. The results of the experiment showed that recognition rates decrease when the degree of stylization increases, while acceptability of the stylization increases. Overall, these results provide guidance to find the desired compromise between stylization levels and recognition, and lay the foundation to new stylization methods offering a tradeoff between stylization and recognition of the actor.

## 2 RELATED WORK

First, we present a number of works that investigated facial style transfer, either in the 2D image domain or in the 3D spatial domain.

Second, we review how a face is perceived identity-wise, which parts of the face hold the most information for recognition, existing biases, and high-level features relevant to recognition.

### 2.1 Facial style transfer in the 2D domain

Facial style transfer — whether for portraits or textures — is nowadays solved by leveraging deep neural networks, which has recently caused a breakthrough in the whole field of image style transfer. Using a content image and a style image, a third image can thus be computed. Several contributions have been made in that area during the last few years, with applications such as changing human head poses and expressions, turning animal faces into others [Huang et al. 2018; Liu et al. 2017, 2019], or human faces into manga faces [Kim et al. 2019]. When style differences are smaller (e.g. human to human), more classical style transfer techniques can be used, although they are not optimised for the case of head portraits, and can produce faces that prove inadequate for the human visual system. Most of facial style transfer methods are based on the MRF (Markov random fields) style transfer of Li and Wand [Li and Wand 2016]. It assumes that each pixel in a texture image is entirely characterised by its spatial neighbourhood, which allows to control the image layout at a local level. This makes the result more realistic than with Gatys's original neural style transfer method [Gatys et al. 2015], which modeled visual textures with summary statistics (Gram matrices). Selim et al. [Selim et al. 2016] proposed to use gain maps to constrain spatial configurations, which can preserve the facial structures while transferring the texture of the style image. Kaur et al. [Kaur et al. 2017] improved facial texture transfer by warping the face of the style image to the shape of the face of the content image in order to improve results facial parts wise, while preserving identity using a facial recognition network loss. Champandard's [Champandard 2016] technique can be used to semantically constrain the style transfer using the facial parts. However, these approaches are extremely specific to texture style transfer, and do not transfer when applied on 3D data.

### 2.2 Style Transfer in the 3D domain

First approaches investigating style transfer between 3D meshes were based on approximating a transformation [Ma et al. 2014] by segmenting the meshes and matching their parts, then using operations such as deformation or substitution between them [Lun

et al. 2016]. However, the low level and patch-work aspect of these approaches make them limited to simple objects (i.e. furniture).

More recently, Yin et al. [Yin et al. 2018] proposed a method able to learn general-purpose shape transforms via point displacements. While significant shape changes, e.g., skeleton-to-shape or incomplete-to-complete scans, are possible, this method is supervised and requires paired shapes from two domains. As paired data is not typically available, a number of unsupervised methods has been proposed: The VAE-CycleGAN of [Gao et al. 2018] encodes each input set into two separate latent spaces and trains a CycleGAN to translate codes between the those latent spaces. This allows animation transfer between various humans, animals, and faces. Yin et al. [Yin et al. 2019] perform cross-domain shape translation by autoencoding them to a shared autocomplete latent space. A translator network is then trained to translate the encoded shapes from one part of the latent space to another, the style being preserved thanks to a feature preservation loss. This method allows furniture style transfer, animation transfer, and low-quality scan to high-quality scan transfer.

In the case of body geometry transfer, Fleming et al. [Fleming et al. 2016] proposed a stylization technique based on interpolations between the person’s 3D scan, an artist-created style template, and an average human mesh. More precisely, they first generate a person-specific style template by adding half of the Euclidean difference between the average body shape and the person’s 3D scan to the style template. Different levels of stylization were then created by interpolating between the person’s 3D scan and the person-specific style template. Danieau et al. [Danieau et al. 2019] used a similar approach for the specific case of facial style transfer, in combination with a more advanced texture style transfer method. Although some methods exist to generate stylized faces, they have however mostly not considered the ability of viewers to recognize the identity of the actor in the results, which is however one of the main motivation for stylizing a character (i.e., that the person is still recognizable in this stylized character). To explore this limitation, we will now explore the areas of face recognition and perception of virtual characters.

## 2.3 Face perception

**2.3.1 Recognition strategies.** A human face reveals a great deal of information to a perceiver [Galton 1879]. A look at it can allow to identify an individual, and it can reveal mood and intention. Of course, it is not the only way to identify an actor. Voice [I. Pollack and Sumby 1954], gait [Hoyet et al. 2013; Johansson 1973] or even clothing can all be used to establish identity, but the face is the easiest and most common way to do it [Bruce and Young 1986].

Global (holistic) and local facial features are both crucial for recognition [Leder and Bruce 1998; V. 1988]. Galton et al. [Galton 1879] described holistic perception as “the sum of a multitude of small details, which are viewed in such rapid succession that we seem to perceive them all at a single glance”. The concept of holism is that faces are recognized better as a “whole” (global appearance), rather than by the recognition of individual parts (nose, eyes, etc.) [Tanaka and Farah 1993]. This effect is although not confirmed for faces that are inverted [Yin 1969], or scrambled [Tanaka

and Farah 1993], nor for non-face objects [Yin 1969] which suggests that holistic encoding is specific to regular faces. However, human’s attention seems to be attracted/drawn quickly by strong local features (large nose, staring eye, etc.). Thus in this case holistic features may not be used. However, their perception seems more subtle. We thus see that two main categories of facial features — holistic and local — can be leveraged for the recognition of our stylized faces. We will now look at which factors impact the use of these features.

**2.3.2 Factors impacting recognition.** Hair, face outline, eyes, mouth, and nose (not necessarily in this order) have been identified to be important for perceiving and remembering faces [G. M. Davies and Shepherd 1981; V. 1988]. The nose was found to be insignificant in frontal images [Zhao et al. 2003], but seems at least as important as the eyes or mouth when looking at profiles [V. 1988]. The perception and recognition of faces has also been studied in term of aesthetic attributes such as beauty, attractiveness, or pleasantness, where most attractive faces have been found to have the best recognition rate, followed by least attractive then mid range faces [Zhao et al. 2003]. Studies also showed that faces with distinctive features are better retained in memory and are recognized better and faster than typical faces [Bruce et al. 1994]. Average faces are perceived as less distinctive and more attractive [Deffenbacher et al. 1998; Valentine et al. 2004], although people are more sensitive to small differences in average faces [Volker Blanz 2000]. Overall, the perception of one’s face is however not completely understood, but results generally suggest that all parts can contribute to facial recognition, and that strong local features are more recognizable.

A large body of research also focuses on understanding cultural differences in face perception, which might therefore relate to our topic involving stylizing an actor’s face into non-human characters (e.g., orc, alien, elf). Previous works in this area showed that people recognize faces from their own-race better than faces from other-races, an effect called the Other Race Effect (ORE) [Meissner and Brigham 2001]. Having experience with faces from other races although does not always cause improvements in recognition for these races [Chiroro and Valentine 1995; Ng and Lindsay 1994]. In addition to perceptual experience, motivation to individuate people from the other-race increases the use of holistic strategies when recognizing faces of this race, and seems to be an important factor for ameliorating the ORE [Levin 2000]. Despite all the studies in this area, the source of the ORE is however still an open question. As a parallel, human and non-human face matching has been shown to be correlated to how human-like the non-human face was, heavily non-human features hindering matching [Danieau et al. 2019].

**2.3.3 Perception of virtual characters.** Finally, there is also a large body of work in the Computer Graphics community investigating the perception of virtual characters’ faces. Shape and texture are two crucial aspects of the appearance of a virtual character [Zell et al. 2015], and while shape is more important for perceived realism, texture is the most important for appeal, eeriness, and attractiveness. Furthermore, mismatches between shape and texture realism lead to the uncanny valley. Although movement can affect the perception of virtual characters — worsening it if the character is already perceived negatively — much of the information used to evaluate virtual characters is available in a still image [McDonnell et al.

2012]. Similarly, stylization can have an important impact on the perception of a character, as Fleming et al. [Fleming et al. 2016] showed that stylizing a character with the identity of an actor can increase its appeal.

To the best of our knowledge, no other research has focused on the recognition of stylized faces of virtual characters. The degree to which a face can be stylized and still remain recognizable is therefore unknown, as well as the degree to which it needs to be stylized to be considered “properly stylized” (looking like an orc, alien, etc. and not a human).

### 3 STYLIZATION METHOD

Before presenting the study on the recognition of stylized faces, this section briefly introduces the face stylization method used in our experiment. The method is based on the work of Danieau et al. [Danieau et al. 2019], and adapted to enable controlling of the degree of stylization. The approach separates the stylization into two parts (geometry and texture) and is illustrated with high quality human facial meshes, captured with a system similar to [Danieau et al. 2019], and non-human facial meshes obtained from the *Paragon* [Games 2018] free assets.

#### 3.1 Geometry Stylization

As previously mentioned, the shape of several facial features is an essential part of a face identity (e.g., mouth, nose). Therefore, to transfer the identity of one’s face to another, the geometrical particularities should be transferred, whether it is the size of the jaw, the angle of the nose, or the eye-to-eye distance. Following the approach of Danieau et al. [Danieau et al. 2019], all meshes are first normalized to the same topology, and then passed through a geometry relative style transfer method, which computes the variations of a given face from an average human model to capture the particularities of each specific face. As all the actors in our studies were male Caucasians, we therefore used as average model the default Caucasian male facial model from *MakeHuman* [team 2019].

In this paper, we are interested in exploring the effect of the degree of stylization, i.e., how much the features of the non-human character are transferred onto the actor’s face on a continuum from no stylization (the actor’s face) to a strong stylization (the actor’s face with strong features of the non-human character), as depicted in Figure 1. However, the original approach of Danieau et al. [Danieau et al. 2019] only enables controlling the degree of exaggeration of the actor’s features onto the non-human character, i.e., producing stylizations on a continuum from the non-human face to the non-human face with features of the corresponding actor. We therefore adapted their method using the following formula:

$$M = M_h + w(M_n - M_a) \quad (1)$$

where  $M$  is the set of vertices of the final mesh,  $M_n$  is the set of vertices of the non-human mesh,  $M_h$  the set of vertices of the human mesh and  $M_a$  the set of vertices of the average human. The weight  $w$  controls the importance of the non-human in the deformation (i.e., the degree of stylization increases with  $w$ ). Examples of geometry stylization at five different levels are presented in Figure 2.

#### 3.2 Texture Stylization

Additionally to geometric stylization, texture stylization enables the transfer of one individual’s face textural characteristics to a non-human texture. The method computes texture differences from an average (reference) human facial texture. As all the actors in our studies were male Caucasians, we used the default Caucasian male texture from *MakeHuman* [team 2019] as the average facial texture. Similarly to [Danieau et al. 2019], the average human texture displayed an artificially flawless skin, to enable transferring facial features such as hair, scars or wrinkles. Finally, as in Danieau et al. [Danieau et al. 2019] the stylization method is based on a neural network optimization [Gatys et al. 2015], where the following loss function is minimized:

$$loss = ((T - T_h \cdot w_{tex}) - (T_n - T_a \cdot w_{tex}) \cdot w_{ctex})^2 \quad (2)$$

where  $T$  is the result of the style transfer process,  $T_h$  is the human texture,  $T_n$  is the non-human texture,  $T_a$  is the default human texture,  $w_{tex}$  is the weight controlling how pronounced the identity features are in the output and  $w_{ctex}$  is the weight controlling the balance between content (non-human) and style (identity).

We also introduced an additional texture normalization step to the original method in order to avoid any influence in our experiments due to differences in colorimetry resulting from the scanning process. In particular, we found that textures generated with the same level of stylization (especially when not using a full stylization) displayed visually varying degrees of perceived stylization when the normalization step was not used. For instance, the three top examples of Figure 4 present differences in intensity, as well as differences in the appearance of some of the key visual features of the non-human model (diamond shape on the forehead), which were more consistent after normalization (right-most three examples). We therefore normalized the original textures (before stylization) by aligning the mean color distribution of the textures used with:

$$T_h = \hat{T}_h - (mean(\hat{T}_h) - targetcolor) \quad (3)$$

where  $\hat{T}_h$  is the original scanned texture,  $mean()$  is the RGB-wise average of an image, and  $targetcolor$  is the RGB value that  $T_h$ ’s mean will be aligned on (we used the mean color of a reference texture with good colorimetry). Examples of texture stylization (including normalization) at 3 different levels are presented in Figure 3.

### 4 STUDY: STYLE AND IDENTITY TRADE-OFF

Style transfer raises the question of the trade-off to reach between the original content and the style to ensure that both are identifiable in the stylized content. In the particular context of face stylization, we are exploring the following questions: To what degree a human face can be stylized and still remain recognizable? To what degree a stylized face can be considered as stylized enough? To answer these questions, we conducted a study exploring the relationship between the degree of stylization and the recognition performance for a set of non-human styles. In addition to recognition accuracy, we also explored what constitutes an acceptable level of stylization. In summary, the main hypotheses of the experiment were:

H1 Lower degrees of stylization will result in higher recognition rates.



Figure 2: Left: actor scanned face; Right: non-human character face; Middle: 5 different levels of stylizations, with style levels 0.40, 0.55, 0.70, 0.85, 1.0.



Figure 3: Left: Original identity; Right: style texture; Middle: Stylized textures at style levels 0.40, 0.75, 1.0.



Figure 4: Three stylized textures, un-normalized (top), then normalized (bottom). Normalisation helps with colometry and textural features (e.g. forehead spots, hair)

H2 Higher degrees of stylization will increase subjective recognition of the original model species. E.g., someone stylized as an orc might not be considered to look like a proper orc at low stylization levels, but will at high stylization levels.

#### 4.1 Population

Twenty-four participants took part in the experiment (6 females). They were between 23 and 61 years old (mean and SD age:  $41.8 \pm 10.9$ ), Caucasian, and were recruited from our laboratory among students and staff. They were all naive to the purpose of the experiment, had normal or correct-to-normal vision, and gave written and informed consent. The study conformed to the declaration of Helsinki. They were not compensated for their participation. None of the participants knew the human faces used in the study.

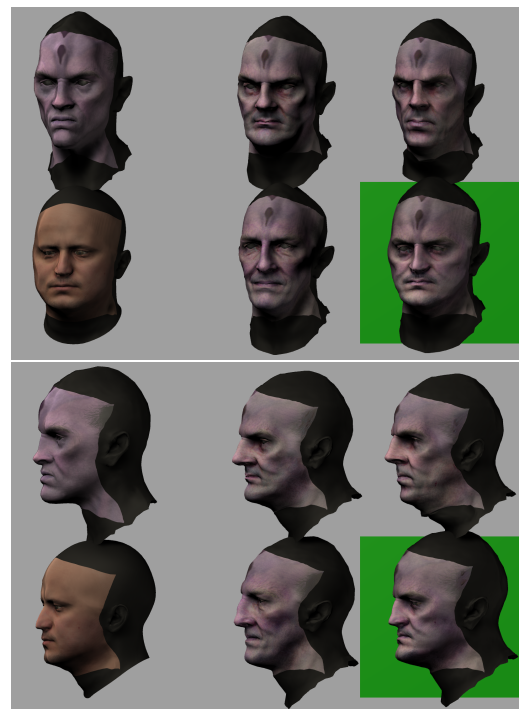


Figure 5: Front and profile views for a trial example of the experiment. Participants could switch between the views by pressing spacebar. The face selected by the participant as being the stylized face of the actor is highlighted in green.

#### 4.2 Stimuli

Ten human face scans were used in the study. In order to reduce recognition due to outliers (bearded face, under-represented gender, etc.), the 10 faces were all unbearded Caucasian males, with

neutral facial expressions (age  $M = 44.86$ ;  $SD = 7.20$ ). As eye color could have also been an easily distinguishable feature, it was not transferred during the stylization process.

To explore the effects of the degree of stylization, we used 5 levels of stylization (from low to high, see Figure 1). As texture and geometry stylization weights in Equations 1 and 2 are not necessarily equivalent, we experimentally selected the following weights for the geometry (0.40, 0.55, 0.70, 0.85, 1.00) and texture (0.60, 0.70, 0.80, 0.90, 1.00). We also used 3 non-human faces, obtained from the Paragon [Games 2018] free assets: an *Alien*, an *Orc* and a *Monkey*. They were picked amongst 5 non-human faces used in a pilot study, respectively as the most recognized, the least recognized, and the one which recognition was the least correlated with the others. They were selected for their relative proximity to a normal human face, where the Alien face was the closest to the average human face (vertex to vertex difference), followed by the Monkey, then the Orc.

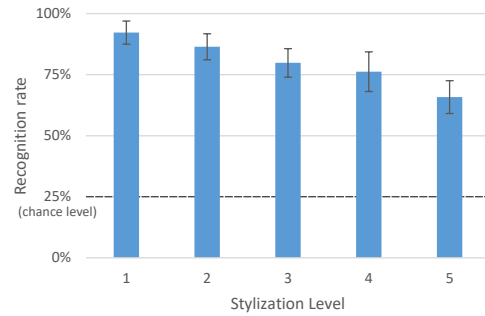
Regarding the presentation of the faces, we controlled the lighting parameters and the viewing angle. For example, Jonhston et al. [Johnston et al. 1992] showed that bottom lighting makes harder to identify familiar faces. Similarly, Hill and Bruce [Hill and Bruce 1996] showed the importance of top lighting for face recognition by matching surface images of faces to determine whether they were identical. Therefore, three directional lights coming from above were used, with intensities 0.9 (front), 0.3 (right), and 0.25 (left). During the experiment, participants could switch between two point-of-views, where meshes were viewed either 30° or 90° from the right (see Figure 5).

### 4.3 Protocol

The task required participants to recognize one face (human scan, displayed on the bottom left corner of the screen) among four stylized faces (displayed on the right side of the screen). The non-human mesh was also displayed for reference (top left corner of the screen). Figure 5 displays a trial example. Participants were instructed to select the stylized face which they thought matched the human face using keys identified on the keyboard. Character faces were presented at a 30° angle at the beginning of the trial, and could be switched back and forth to the 90° angle by pressing the space key. Each trial lasted a maximum of 20 seconds, and was displayed on a 24-inch screen located 50cm away from participants.

We used a between-subject design, as each participant was shown stylization results for only one non-human out of three (Alien, Monkey or Orc) during the study. In total participants performed 100 trials, displayed in random order: 10 Actor Faces  $\times$  5 Stylization Levels (low to high stylization)  $\times$  2 Repetitions. In each trial, the four stylized faces presented to participants all had the same degree of stylization: one of them was always the correct stylized actor, while the three others were selected randomly from the nine remaining actor faces. Before the experiment, one training trial was included to familiarize participants with the user interface, using a human face never used again in the experiment.

Finally, participants were asked to fill in a post-study questionnaire to evaluate their subjective perception of the stylized faces, both in terms of recognizing the actor (Identity subjective rating) and the species of the reference model (Species subjective rating).



**Figure 6: Main effect of Stylization Level on recognition rates. Error bars represent standard deviation.**

For each Actor Face, participants were showed simultaneously all 5 stylization levels, as well as both the reference actor and non-human faces, and asked two questions: “From the least stylized face to the most, up to which stylization level can you recognize the individual?” and “From the most stylized to the least, down to which stylization level would you consider the face to be of the same species than the reference non-human?” Participants required approximately 35 minutes to complete the entire experiment.

## 4.4 Results

**4.4.1 Recognition rates.** To analyze recognition rates, we first performed a 3-way mixed-design Repeated Measure Analysis of Variance (ANOVA) with within-subject factors Stylization Level and Actor Face, and between-subject factor Non-Human Model. Effects are then explored further using Neuman-Keuls post-hoc test for pair-wise comparisons of means. We first found a main effect of Stylization Level ( $F_{4,88} = 34.99$ ,  $p < 0.0001$ ), where post-hoc analysis showed that recognition rates were significantly decreasing with increasing stylization levels ( $p < 0.05$ ), except between Levels 3 and 4 (Figure 6). More specifically, we found that recognition rates ranged on average from 92% (lowest level of stylization) to 66% (highest level of stylization), where chance level was 25%. We also found a main effect of the Non-Human Model presented to participants ( $F_{2,22} = 3.52$ ,  $p = 0.047$ ), which was not confirmed by the post-hoc analysis (Alien-Monkey:  $p = 0.656$ ; Alien-Orc:  $p = 0.057$ ; Monkey-Orc:  $p = 0.059$ ), even though there seems to be a tendency for actors to be on average less recognized when stylized onto the Orc model than onto the other two. We finally found a main effect of Actor Face ( $F_{9,198} = 3.955$ ,  $p < 0.001$ ), where post-hoc analysis showed that some actors were on average significantly more recognized than others. Moreover, recognition rates were high on average, even for the least recognized actors, ranging from 73% to 89%. We however did not find any interaction effect between factors.

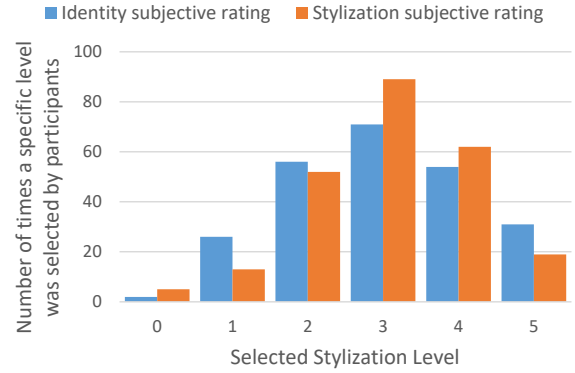
In order to further understand the relation between recognition rates and the degree of stylization, we then computed person correlations between these two variables (averaged over the non-human models for each participant, as we did not find a significant effect of this factor in the first analysis). Results showed that recognition rates are negatively correlated with the level of stylization ( $r = -0.59$ ,  $p < 0.001$ ). All these results therefore support **H1**, showing that *lower degrees of stylization result in higher recognition rates*.

**4.4.2 Subjective scores.** At the end of the experiment, participants were asked to provide subjective ratings for each Actor Face regarding the stylization level from which they considered that they could not recognize each individual (Identity subjective rating), as well as the level from which they considered that the stylized face was not anymore from the same species as the reference non-human (Species subjective rating). The distributions of both Identity and Species subjective ratings are displayed in Figure 7.

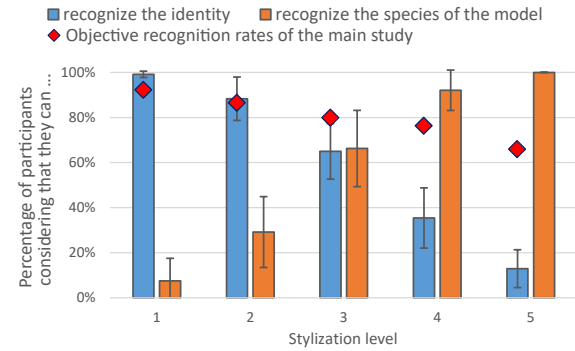
We first analyzed the Stylization Level selected by participants for each Actor Face for both questions by performing two mixed RM ANOVA with within-subject factor Actor Face and between-subject factor Non-Human model. We only found a main effect of Actor Face for the Identity subjective ratings ( $F_{9,189} = 2.777, p < 0.01$ ), where post-hoc analysis showed that overall there were mostly no significant differences between actors, except from slight differences between a few extremes. These results suggest that the displayed model or actor did not influence the level at which participants considered that they could not recognize an individual or species.

To further explore this data, we transformed the subjective ratings to represent the probability of recognizing the identity or species at a given level of stylization. For the identity, answers were converted to a binary format where a stylization level for an individual was assigned a value of 1 if it was below or equal to the level selected by the participant (i.e., participants answered that they were able to recognize the identity at this given level of stylization), and 0 otherwise. The opposite was performed for the species subjective ratings, meaning that the species of the model at a given level of stylization would be considered to be recognized by participants if it was above or equal to the subjective rating. We then performed two mixed RM ANOVA with within-subject factor Stylization Level and between-subject factor Non-Human model. We found a main effect of Stylization Level for both identity ( $F_{4,84} = 150.04, p < 0.0001$ ) and species ( $F_{4,84} = 100.43, p < 0.0001$ ), where post-hoc analyses showed that all the stylization levels were significantly different and decreasing for identity (all  $p < 0.05$ ), and significantly different and increasing for style (all  $p < 0.001$ ) except for the two highest levels, which supports **H2**. These results are displayed in Figure 8, concurrently with the objective recognition rates analyzed in Section 4.4.1. It is also interesting to notice that participants' subjective perception of their ability to recognize an identity was drastically lower than their objective performance, especially for higher stylization levels, which is further discussed in Section 5. Finally, identity subjective ratings were found to be negatively correlated with the species subjective rating ( $r = -0.74, p < 0$ ), suggesting that identity preservation and stylization might be based on a compromise.

**4.4.3 Non-uniformity of the choices repartition.** To explore whether the distinctiveness of the actors' faces (i.e., in term of how different they are from the other faces) influenced user performance, we computed for each actor scan the geometrical difference to the average of the ten faces used in the study (sum of vertex-to-vertex distances). We then explored potential correlations with the number of times each actor face was selected in the experiment, with the average recognition rate per actor, as well as with the average precision per actor (number of correct recognition divided by number of times the given face was selected). We only found a slightly significant positive correlation of the geometrical difference with precision



**Figure 7: Distribution of the Identity and Stylization subjective ratings.**



**Figure 8: Percentage of participants considering that they can recognize the identity or the style at a given stylization level. It is interesting to notice that subjective ratings are usually lower than objective performance, especially for higher stylization levels.**

( $r = 0.63, p = 0.0497$ ), suggesting that more distinctive faces (in terms of geometric distances) might tend to be slightly more accurately recognized than less distinctive ones.

However, as geometrical differences might not capture the subtle idiosyncrasies of a person's face, we also decided to explore facial differences in terms of features extracted from a view of each human face using a facial recognition neural network [King 2009]. With this metric, we also found a similar positive correlation between how different the face was and precision ( $r = 0.61, p = 0.0603$ ). Interestingly, we also found a significant negative correlation with the number of times a face was chosen ( $r = -0.79, p < 0.01$ ), suggesting that less distinctive faces were more often chosen by participants.

These results suggest that novel metrics (e.g., neural network based methods) can provide face similarity metrics similar to common geometric methods, despite working in other domains (e.g., 2D domains for images used with neural networks). This therefore opens a possibility for novel manners of analyzing such effects, which is further discussed in Section 5.



**4.4.4 Time spent.** Participants spent an average of 10.8 seconds on each trial (min=5.0, max=16.5), with an average total time of the study of 18 minutes. Correlations between the total mean time spent and performance were measured for each style weight level and are all negative, but only significant for the lower level of stylization ( $r = -0.535$ ,  $p < 0.01$ ). This result suggests that participants needing time to recognize an identity perform worse on average, especially when faces are only partially stylized.

## 5 DISCUSSION

In this paper, we have investigated the recognizability of stylized faces for different identities and stylization levels, when both facial geometrical and textural styles are transferred. More precisely, we conducted a study involving three facial models and five stylization levels. Recognition was found to decrease linearly with the level of stylization, ranging from 93% recognition accuracy at the lowest level of stylization to 66% for the highest level. A similar tendency was observed from subjective ratings collected at the end of the experiment, where participants were directly asked to select the level of stylization up to which they considered that they could recognize the individual. Interestingly, while the average recognition rate reached 66% for the highest level of stylization in our experiment (where chance level was 25%), participants' subjective perception of their ability to recognize an identity was drastically lower, and below 50% for the two highest levels of stylization (respectively 35% and 13%). This result suggests that while participants might consider that they are not able to recognize a stylized individual when it is presented alone, the ability to recognize the person might actually be higher when presented amongst other stylized individuals. However, the limit to which this is possible, and facial features potentially responsible for such differences, cannot be determined from the experiment presented here, and would therefore require to be explored in further studies.

As for the perception of human faces, we also found slight differences in the recognition of stylized individuals. These differences were however small, and interestingly recognition rates were always relatively high (min: 73%; max: 89%) and all above chance level. It is also important to point that the experiment was conducted using only 10 human faces, which were selected to avoid introducing any recognition bias (e.g., ethnicity, age, gender). However, future experiments involving a broader number and variety of faces would allow stronger generalisation of the findings to other faces.

Similarly, in order to explore potential effects of the target character face, we also used three different non-human faces in our experiment, which were selected for their relative proximity to a normal human face (the Alien face was the closest to the average human face, followed by the Monkey, then the Orc). Unexpectedly, we did not find any effect of the character model on participants' objective or subjective ability to recognize individuals. However, we observed a slight tendency for individuals to be on average less recognized when stylized onto the Orc model, most different from the average human face. This result is in accordance with the results of Danieau and al. [Danieau et al. 2019], that showed that the less human-like a face was, the less it was recognized. Despite being non-significant, this tendency raises the question of how different

the character face can be while stylizing an individual, and the limits of the stylization.

Concurrently to the ability of observers to recognize the stylized actor, we also collected subjective ratings regarding whether observers considered the stylized faces to be of the same species than the reference non-human character. This subjective rating was positively influenced by the degree of stylization, where only 8% of the participants answered that they considered the lowest level of stylization to be of the same species, while 92.5% considered it to be the case for the highest level of stylization. The degree to which participants were convinced by the style of the result was also found to be negatively correlated to the degree to which the actor face was recognized, therefore suggesting that a balance might be required to enable observers to recognize the actor while providing sufficient cues about the type of character he/she is stylized into. It is also possible that achieving both high recognition and stylization might be currently limited by the stylization method available. Other stylization approaches could therefore be explored in the future to reduce this limitation.

As studies in Psychology demonstrated that distinctive faces affect recognition [Bruce et al. 1994], we also explored a possible influence of the distinctiveness of the actors' faces on participants' ability to recognize stylized faces. We looked into this by assessing the level of distinctiveness using two methods: the geometrical distance of each face to the average of the faces, and the distance of their projection in a facial recognition neural network to the average of the faces. We found some small correlations between distinctiveness and recognition precision for both methods, suggesting that more distinctive faces might tend to be slightly more accurately recognized when stylized than less distinctive ones. While these tendencies are similar to results observed for facial recognition, it also suggests that novel metrics, such as those based on neural network methods, might provide novel manners of analyzing such effects. Also, beyond considering the limits of a metric for measuring objective face distance, face perception remains a subjective matter not only based on the position of faces in a shape or texture space, but also on aspects such as attractiveness and familiarity [Melissa Peskin 2002], which could also be evaluated using such metrics in the future.

Despite providing the first insights about the influence of stylization level on the perception of faces, the controlled settings of our study might have introduced some limitations which should be explored in further studies. For instance, we presented participants with only 2 possible points of view during the experiment, as well as with a static lighting and no facial animation. As recognition of an individual involves more features than a static face, future studies could explore the influence of other factors on the ability of viewers to recognize an individual, including facial expressions, body shapes to motion, or the tone of the voice. Also, for this first study on the topic, the stylization process applied a global stylization to the face, as to our knowledge there is no information about how stylizing different parts would affect recognition, while it was simultaneously restricted for some areas of the face (i.e., no ear or eye stylization), which might have reduced recognition abilities. However it would be interesting in the future to identify which are the features most important for stylization, which would open the door to more localized stylizations. Finally, our experiment involved

the use of a single facial style transfer method, as there are currently no other method available in the literature. With the growing interest for this field of study, new methods might appear in the future and would benefit from comparison using our experimental protocol as a baseline.

## 6 CONCLUSION

In this paper, we have investigated the recognizability of stylized faces for different identities and stylization levels. We have presented a study, based on 3 character faces and 5 stylization levels, which main goal was to measure the balance between style and recognition. Recognition was found to linearly decrease with stylization, while participants' ability to match the stylized content with the original character species increased linearly with the level of stylization. Recognition is thus inversely proportional to being convinced by the style. These results provide new insights about necessary compromises between stylization and recognition, and pave the way for the new field of study of 3D facial style transfer, combining both knowledge in facial perception and computer graphics. We believe that this field of study will have many applications in line with the increasing interest in digital humans.

## REFERENCES

- Vicki Bruce, Mike A. Burton, and Neal Dench. 1994. What's Distinctive about a Distinctive Face? *The Quarterly Journal of Experimental Psychology Section A* 47, 1 (Feb. 1994), 119–141.
- Vicki Bruce and Andy Young. 1986. Understanding face recognition. *British Journal of Psychology* 77, 3 (1986), 305–327.
- Alex J. Champandard. 2016. Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks. *ArXiv* (2016).
- Patrick Chioro and Tim Valentine. 1995. An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A* 48, 4 (Nov. 1995), 879–894.
- Fabien Danieau, Ilja Gubins, Nicolas Olivier, Olivier Dumas, Bernard Denis, Thomas Lopez, Nicolas Mollet, Brian Frager, and Quentin Avril. 2019. Automatic Generation and Stylization of 3D Facial Rigs. *IEEE VR* (March 2019), 784–792.
- Kenneth A Deffenbacher, Thomas Vetter, John Johanson, and Alice J O'Toole. 1998. Facial Aging, Attractiveness, and Distinctiveness. *Perception* 27, 10 (Oct. 1998), 1233–1243.
- Reuben Fleming, Betty J. Mohler, Javier Romero, Michael J. Black, and Martin Breidt. 2016. Appealing Female Avatars from 3D Body Scans: Perceptual Effects of Stylization. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, Rome, Italy, 333–343.
- H. D. Ellis G. M. Davies and J. W. Shepherd. 1981. Studies of cue saliency. *Perceiving and remembering faces, vol 96, University of Illinois Press* (1981).
- Sir Francis Galton. 1879. Composite portraits, made by combining those of many different persons into a single, resultant figure. *Journal of the Anthropological Institute* (1879).
- Epic Games. 2018. Paragon Assets. <https://www.unrealengine.com/en-US/paragon>
- Lin Gao, Jie Yang, Yi-Ling Qiao, Yu-Kun Lai, Paul L. Rosin, Weiwei Xu, and Shihong Xia. 2018. Automatic unpaired shape deformation transfer. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 1–15.
- Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. *arXiv:1508.06576 [cs, q-bio]* (Sept. 2015).
- Harold Hill and Vicki Bruce. 1996. The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance* 22, 4 (1996), 986–1004.
- Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehhee Lee, Jessica Hodgins, and Carol O'Sullivan. 2013. Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies. *ACM Transactions on Graphics* 32, 6 (Nov. 2013), 1–11.
- Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal Unsupervised Image-to-Image Translation. *arXiv:1804.04732 [cs, stat]* (April 2018).
- J. M. Pickett I. Pollack and W. H. Sumbly. 1954. On the Identification of Speakers by Voice. *The Journal of the Acoustical Society of America* 26, 403 (1954).
- G. Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* (1973).
- Alan Johnston, Harold Hill, and Nicole Carman. 1992. Recognising Faces: Effects of Lighting Direction, Inversion, and Brightness Reversal. *Perception* 21, 3 (June 1992), 365–375.
- Parneet Kaur, Hang Zhang, and Kristin J. Dana. 2017. Photo-realistic Facial Texture Transfer. *arXiv:1706.04306 [cs]* (June 2017).
- Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. 2019. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. *arXiv:1907.10830 [cs, eess]* (July 2019).
- Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.
- Helmut Leder and Vicki Bruce. 1998. Local and Relational Aspects of Face Distinctiveness. *The Quarterly Journal of Experimental Psychology Section A* 51, 3 (Aug. 1998), 449–473.
- Daniel T. Levin. 2000. Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General* 129, 4 (2000), 559–574.
- Chuan Li and Michael Wand. 2016. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. *arXiv:1601.04589 [cs]* (Jan. 2016).
- Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Long Beach, California, USA, 700–708.
- Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. 2019. Few-Shot Unsupervised Image-to-Image Translation. *arXiv:1905.01723 [cs, stat]* (May 2019).
- Zhaoliang Lun, Evangelos Kalogerakis, Rui Wang, and Alla Sheffer. 2016. Functionality preserving shape style transfer. *ACM Transactions on Graphics* 35, 6 (Nov. 2016), 1–14.
- Chongyang Ma, Haibin Huang, Alla Sheffer, Evangelos Kalogerakis, and Rui Wang. 2014. Analogy-driven 3D style transfer: Analogy-driven 3D style transfer. *Computer Graphics Forum* 33, 2 (May 2014), 175–184.
- Rachel McDonnell, Martin Breidt, and Heinrich H. Bühlhoff. 2012. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics* 31, 4 (July 2012), 1–11.
- Christian A. Meissner and John C. Brigham. 2001. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law* 7, 1 (2001), 3–35.
- Fiona N. Newell Melissa Peskin. 2002. Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces. *Perception* (2002).
- Wei-Jen Ng and R. C.L. Lindsay. 1994. Cross-Race Facial Recognition: Failure of the Contact Hypothesis. *Journal of Cross-Cultural Psychology* 25, 2 (June 1994), 217–232.
- Ahmed Selim, Mohamed Elgharib, and Linda Doyle. 2016. Painting style transfer for head portraits using convolutional neural networks. *ACM Transactions on Graphics* 35, 4 (July 2016), 1–18.
- James W. Tanaka and Martha J. Farah. 1993. Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A* 46, 2 (May 1993), 225–245.
- The MakeHuman team. 2000–2019. *MakeHuman*. [www.makehumancommunity.org](http://www.makehumancommunity.org)
- Bruce V. 1988. *Recognizing faces*. Lawrence Erlbaum Associates.
- Tim Valentine, Stephen Darling, and Mary Donnelly. 2004. Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review* 11, 3 (June 2004), 482–487.
- Thomas Vetter Heather A. Wild Volker Blanz, Alice J. O'Toole. 2000. On The Other Side of the Mean: The Perception of Dissimilarity in Human Faces. *Perception* (2000).
- Kangxue Yin, Zhiqin Chen, Hui Huang, Daniel Cohen-Or, and Hao Zhang. 2019. LO-GAN: Unpaired Shape Transform in Latent Overcomplete Space. *arXiv:1903.10170 [cs]* (March 2019).
- Kangxue Yin, Hui Huang, Daniel Cohen-Or, and Hao Zhang. 2018. P2P-NET: bidirectional point displacement net for shape transform. *ACM Transactions on Graphics (TOG)* 37, 4 (July 2018), 152:1–152:13.
- Robert K. Yin. 1969. Looking at upside-down faces. *Journal of Experimental Psychology* 81, 1 (1969), 141–145.
- Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize?: the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics* 34, 6 (Oct. 2015), 1–12.
- W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. 2003. Face recognition: A literature survey. *Comput. Surveys* 35, 4 (Dec. 2003), 399–458.