The Secret of Appeal

Understanding Perception of Realistic and Stylized Faces

Eduard Zell

A thesis submitted to the
Faculty of Technology, Bielefeld University
in fulfillment of the requirements for the degree of
Doctor rerum naturalium (Dr. rer. nat.),

Bielefeld 2018

Acknowledgments

Abstract

Stylized characters are highly used in movies and games. Furthermore, stylization is mostly preferred over realism for the design of toys and social robots. However, the design process remains highly subjective because the influence of possible design choices on character perception is not well understood. Investigating the high-dimensional space of character stylization by means of perception experiments is difficult because creating and animating compelling characters of different stylization levels remains a challenging task. In this context, computer graphics algorithms enable the creation of highly controllable stimuli, simplifying examination of specific features that can strongly influence the overall perception of a character.

This thesis is separated into two parts. First, a pipeline is presented for creating virtual doubles of real people. In addition, algorithms are described suitable for the transfer of surface properties and animation between faces of different stylization levels. With ELASTIFACE, a simple and versatile method is introduced for establishing dense correspondences between textured face models. The method extends non-rigid registration techniques to allow for strongly varying input geometries. The technical part closes with an algorithm that addresses the problem of animation transfer between faces. Such facial retargeting frameworks consist of a pre-processing step, where blendshapes are transferred from one face to another. By exploring the similarities between an expressive training sequence of an actor and the blendshapes of a facial rig to be animated, the accuracy of transferring the blendshapes to actor's proportions is highly improved. Consequently, this step overall enhances the reliability and quality of facial retargeting.

The second part covers two different perception studies with stimuli created by using the previously described pipeline and algorithms. Results of both studies improve the understanding of the crucial factors for creating appealing characters across different stylization levels. The first study analyzes the most influential factors that define a character's appearance by using rating scales in four different perceptual experiments. In particular, it focuses on shape and material but considers as well shading, lighting and albedo. The study reveals that shape is the dominant factor when rating expression intensity and realism, while material is crucial for appeal. Furthermore, the results show that realism alone is a bad predictor for appeal, eeriness, or attractiveness. The second study investigates how various degrees of stylization are processed by the brain using event-related potentials (ERPs). Specifically, it focuses on the N170, early posterior negativity (EPN), and late positive potential (LPP) event-related components. The face-specific N170 shows a u-shaped modulation, with stronger reactions towards both, most abstract and most realistic compared to medium-stylized faces. In addition, LPP increases linearly with face realism, reflecting activity increase in the visual and parietal cortex for more realistic

faces. Results reveal differential effects of face stylization on distinct face processing stages and suggest a perceptual basis to the uncanny valley hypothesis.

List of Math Symbols

Symbol	Description
$\mathcal{B}$	featureless, smooth version of mesh $\mathcal{M}$
$\mathcal{M}$	triangle mesh, consisting of vertices $\mathbf{v}_i$
$\mathcal{P}$	point cloud, consisting of points $\mathbf{p}_j$
$\mathcal{S}$	shader attached to a mesh $\mathcal{M}$
$\mathcal{T}$	2D texture
$\mathbf{v}^\mathrm{s}$	variable belonging to the source mesh
$\mathbf{v}^\mathrm{t}$	variable belonging to the target mesh or point cloud
$\mathbf{v}^{bs}$	variable belonging to the smoothed source mesh
$\mathbf{v}^{bt}$	variable belonging to the smoothed target mesh
$\hat{\mathbf{v}}$	original, undeformed value of variable
$\bar{\mathbf{v}}$	average value of a variable
A	Voronoi area of a vertex
$\mathbf{a}$	marker position of captured actor
$\mathbf{b}$	vertex position of personalized blendshape
C	total amount of closest point correspondences
$\mathbf{c}$	position of the closest point correspondence
d_k	displacement of all vertices of a blendshape
$\mathbf{d}$	orthogonal displacement vectors of a PCA
$\mathbf{D}$	matrix containing all displacements $\mathbf{d}$
$f,\ F$	counting variable and total amount of frames
$\mathbf{g}$	vertex position of initial guess blendshapes
H	mean curvature
$i,\ j$	counting variables
$k,\ K$	counting variable and total amount of blendshapes
$l,\ L$	counting variable and total amount of faces
$m,\ M$	counting variable and total amount of sparse points on mesh (markers or reference points)
$n,\ N$	counting variable and total amount of vertices
$\mathbf{n}_i$	normal of a point $\mathbf{p}_i$
P	total amount of points in a point cloud
$\mathbf{p}$	point position on a 3D surface or in a point cloud
r	parameter controlling steepness of a function
$\mathbf{r}$	position of reference point
$\mathbf{R}$	rotation matrix of size 3×3
$\mathbf{s}$	vertex position withiin a sparse blendshape
t_k	trust value of a blendshape
$\mathbf{t}$	translation vector of size 3

Symbol	Description
$\mathbf{u}$	2D texture coordinates
$\mathbf{v}$	vertex position
$\mathbf{V}$	matrix containing all $\mathbf{v}$ for all blendshapes
w	weight of a blendshape or a principal component dimension
$\mathbf{w}$	vector containing all weights w
$x,\ y,\ z$	3D coordinates
$\alpha,\ \beta,\ \gamma$	angles in triangle
$\delta\mathbf{v},\ \ldots$	delta formulation of a variable, e.g., $\delta\mathbf{v} = \mathbf{v}_k - \mathbf{v}_0$
$\kappa_1,\ \kappa_2$	maximum and minimum curvature
ρ	Pearson correlation between two variables
ρ^+	positive Pearson correlation $\max(\rho, 0)$
P	correlation matrix between blendshapes
σ	singular values of data matrix of PCA
τ	trust value of specific blendshape
$\chi()$	boosting function for blendshape similarity
ψ	soft mask encoding displacement strength for blendshapes

Table 1: Overview of math symbols and description for mesh editing and animation context

Symbol	Description
c	power coefficient of the Phong BRDF
$\mathbf{h}$	normalized half-vector between $\mathbf{l}$ and $\mathbf{n}$
$\mathbf{h}_u$	unnormalized half-vector between $\mathbf{l}$ and $\mathbf{n}$
i	script for incoming light direction
$k_d,\ k_s$	color constants for diffuse and specular shading
$\mathbf{l}$	normalized light direction
m	surface roughness term
$\mathbf{n}$	normal of a point
p	path of light within a scattering medium
r	script for reflected light direction
$\mathbf{s}$	one out of two bitangent vectors of a curve
$\mathbf{t}$	tangent vector of a curve
$\mathbf{u}$	one out of two bitangent vectors of a curve
$\mathbf{v}$	normalized view direction
ϕ	azimuthal angle between two vectors
θ	longitudinal angle between two vectors
$\mathbf{F}_0$	surface's characteristic specular color
$f_R(\ldots)$	Bidirectional Reflectance Distribution Function
$f_S(\ldots)$	Bidirectional Curve Scattering Distribution Function
$D(\ldots)$	Normal Distribution Function
$F(\ldots)$	Fresnel Reflectance Function
$G(\ldots)$	Geometry Function
$L(\ldots)$	Radiance
$M(\ldots)$	Longitudinal Scattering Function
$N(\ldots)$	Azimuthal Scattering Function

Table 2: Overview of math symbols and description in rendering context

Contents

1 Introduction

Faces are one of the most powerful tools for social communication that reveal a wide range of information about a person like gender, age, ethnicity or mood (Jack & Schyns 2015, Tsao & Livingstone 2008). It is therefore not surprising that first research on faces and facial expressions is dating back to Darwin. Some studies had even a big impact beyond psychology. Ekman's Facial Action Coding System (FACS) (Ekman 1972) is a fundamental concept for facial animation in computer graphics. Conversely, computer generated faces have started to replace photographs of real people in perceptional experiments, due to better experimental stimulus control (de Borst & de Gelder 2015) and recent advances in capturing and rendering of human faces (Alexander et al. 2010, Jimenez et al. 2015). These two prominent examples demonstrate well that advances in computer graphics lead to new types of experiments in psychology and vice versa; insights in psychology induce new algorithms in computer graphics (Wallraven et al. 2007, Klehm et al. 2015). So far, most research in computer graphics as well as in facial perception is centered on real or realistic humans. However, realistic characters represent only a fraction of the characters used in computer games or animation movies, where characters of different stylization levels are common. The aim of this thesis is to improve the understanding of what makes characters more appealing. For the design of perception studies it is recommended to change only one variable over time. But creating characters of different stylization level with the same texture or animation is technologically difficult. In order to overcome these issues, new algorithms are developed for enabling new types of experiments that focus on the perception of stylized characters. The benefit of algorithms is demonstrated by two extensive studies that offer new insights on stylized face perception.

Although cartoon faces are often associated with simplicity, creation of stylized characters is not necessarily easy. In the movie *Maleficient* expensive hardware, custom software and months of work by skilled professionals and researchers were required to create the flower pixies (Seymor 2014). The challenge of creating convincing characters is the multidimensional design space in combination with various psychological factors that influence facial perception. Influencing factors that have been reported in the past are: familiarity (Dill et al. 2012, Tinwell & Grimshaw 2009), degree of realism (Mori et al. 2012), reasonable facial proportions (Seyama & Nagayama 2007), rendering style (McDonnell et al. 2012) and 2D image filters (Wallraven et al. 2007).

While it is generally accepted that stylized versions are often preferred over realistic characters (Geller 2008), the reasons for this are still unclear. In fact, the popular theory of the uncanny valley (Mori et al. 2012) has been criticized over the last years (Kätsyri et al. 2015) due to lack of empirical evidence. The challenge is

Figure 1.1: Different levels of stylization of a female character specifically created for our perception experiments.

to identify reliable and well-understood factors and translate this knowledge into feasible guidelines for generating appealing 3D virtual characters. In order to do so, three main limitations have been identified in previous work: (i) the absence of a continuous stylization for the same character, (ii) the requirement to transfer either surface properties or facial animation across characters of different stylization levels and (iii) the lack of perceptual experiments beyond rating scales. The first problem is addressed by creating a high quality dataset containing several stylization levels of two identities (Figure 1.1). For the transfer of surface properties like skin textures, dense one-to-one correspondences are required. In this way, material properties saved in textures can be transferred between faces. With ELASTIFACE we present a simple and effective framework for establishing dense correspondences between a given set of face models, both in terms of their geometries and textures. In addition, a facial retargeting algorithm is presented that is well suitable for animation transfer between realistic and stylized characters. Finally, perception of stylized characters is investigated by using rating scale perception experiments in combination with Event-Related Potentials (ERPs), which is a well-established approach in neuroscience for the investigation of processes within the brain. In the following, we list the contributions for dense correspondence matching, facial animation retargeting and the main findings from our perceptual studies.

Geometry and Texture Matching Existing methods, that establish dense correspondences between non-isometric models, require a transfer of the input meshes into a simpler domain. In contrast to most previous work, our method can handle input models of extremely different geometries without adjusting the mesh connectivity or being limited to specific domains like a plane or a sphere. Our main contributions are:

- We propose to apply aggressive simultaneous fairing, such that the simple domain evolves naturally. After obtaining two very similar featureless meshes, a simple variant of non-rigid registration is sufficient for establishing accurate correspondences.

- The algorithmic core of our method is based on solving simple bi-Laplacian linear systems and is therefore easy to implement.

- Based on the obtained dense correspondences, we show how textures can be transferred between meshes with different connectivity and parametrization.

- For the use case of facial perceptual experiments, we demonstrate how facial parts, textures and several rendering styles can be blended at interactive framerates.

Facial Retargeting with Range of Motion Alignment The transfer of facial animation between different characters, better known as facial retargeting, consists of a pre-processing step and the actual animation transfer. As a pre-processing step, a transfer of blendshapes is required that aligns the blendshapes of the face rig to be animated with the actor's proportions, which is the source of the animation. We improve existing retargeting algorithms as following:

- Based on the observation that facial motions are similar across different stylization levels, as motivated by the Facial Action Coding System (FACS) (Ekman & Friesen 1978), a metric is suggested for measuring similarity of facial expressions.

- Based on a new manifold alignment approach and the expression similarity measure, a novel algorithm creates actor-specific blendshapes in an unsupervised manner with the help of a training sequence of the actor's facial motion. Even if the facial rig and the actor differ strongly in their facial proportions, we successfully align the ranges of motion of the actor and the character face rig. This subsequently leads to accurate retargeting.

- A Cross-Expression Graph is introduced that maintains consistency between blendshapes during the expression transfer.

- A prior energy, which is based on physically-inspired deformations, addresses the few artifacts during facial retargeting that remain even in case of accurate blendshape transfer.

- Our contributions are fully compatible with most previous methods suitable for real-time applications, and produce results comparable or better than state-of-the-art offline methods (Seol et al. 2012).

Rating Scales Experiments Based on a high quality stimuli set for a male and a female character consisting of five stylization levels in addition to photographs, we investigate two of the main aspects that primarily define the stylization of a 3D character: shape and material (including texture, shading and lighting). 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, which measuring scales reveal most information, and how our initial stimuli can be improved. Results and acquired knowledge from these tests are then used to guide a second round of experiments, where we deeply explore the design space with more samples along the core factors of the previous studies. Our main contributions 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.

- Appeal, eeriness and attractiveness are highly affected by the stylization level of material; 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 was constantly perceived as less appealing and more eerie.

- Our stimuli are accessible for further investigations[1].

Event-Related Potential Experiment While previous experiments relied on rating scales measuring subjective criteria, Event-Related Potentials (ERPs), which are a systematic approach to measure electric signals of the brain, are an excellent way to analyze face perception more objectively. Based on the high temporal resolution of ERPs, it is possible to distinguish between highly automatic and more controlled stages of processing. Furthermore, the availability of rating scales for the same stimuli allows a better interpretation of specific brain responses. Due to the small amount of comparable work using ERP for stylized face perception, only the most distinctive stimuli have been chosen for this experiment. Our main findings are:

- The N170[2] amplitude peaked most for highly stylized and real faces, creating a U-shaped function across the stylization levels. However, the signal's origin seems to be caused by different regions of the brain.

- For the N170 and the Early Posterior Negativity (EPN)[3] strong effects occurred for different emotions, showing most intensive processing for angry expressions, followed by happy and finally neutral faces.

- A selectively increased N170 was found for abstract happy faces, while for more realistic faces a bias towards angry faces was observed.

- The Late Positive Potential (LPP)[4] component increased for more realistic faces in combination with larger and broader activation in occipito-parietal areas in source space.

- Our results indicate that realism and neotenic features, like big eyes or smooth skin, cause a stronger brain response for the according stimuli.

To our knowledge, both studies are the first of its kind because: (i) The rating scales experiments focus on the combination of different levels of stylization in shape and material. (ii) The ERP experiment contains several stylization levels for

---

[1]http://graphics.uni-bielefeld.de/publications/sigasia2015/

[2]The N170 is a face specific negative amplitude of an electric signal appearing about 170 ms after a stimuli is shown. Please see Section 7.1 for more details.

[3]Electric signal meassured within the time range of 250–400 ms after showing the stimuli. Section 7.1 provides more details.

[4]Electric signal meassured within the time range of 400–600 ms after showing the stimuli. Section 7.1 provides more details.

the same identity. In addition, the combination of the two experiments leads to better interpretation and understanding of the brain responses. Based on our results we propose an alternative perspective on the uncanny valley hypothesis. Rather than trying to predict appeal from perceived realism of a character, realism alone seems to be a positive trait in character perception, as information processing involves more brain regions. This is accompanied by positive reactions towards neotenic features that dominate designs of highly stylized characters. Hence both characteristics are rather exclusive features for specific stylization levels. This might explain the intuition behind the uncanny valley hypothesis, but further investigation is required to confirm this new hypothesis. Overall, the presented studies provide useful insights on how to create appealing characters, and are another example how computer graphics and perceptual experiments can benefit from each other.

Related Publications

This thesis is based on the following publications (in chronological order):

- Eduard Zell, Mario Botsch (2012), 'Developing Design Guidelines for Characters from Analyzing Empirical Studies on the Uncanny Valley', in *Proc. of ACM International Symposium on Facial Analysis and Animation (FAA)*, pp. 11:1-2.

- Eduard Zell, Mario Botsch (2013), 'ELASTIFACE: Matching and Blending Textured Faces', in *Proc. of the Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, pp. 25-33.

- Jascha Achenbach, Eduard Zell, Mario Botsch (2015), 'Accurate Face Reconstruction through Anisotropic Fitting and Eye Correction', in *Proc. of Vision, Modeling and Visualization (VMV)*, pp. 1-8.

- Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, 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), pp. 184:1–184:12.

- Sebastian Schindler*, Eduard Zell*, Mario Botsch, Johanna Kissler (2017), 'Influence of Facial Stylization on the N170 and LPP', *Scientific Reports* **7**, 45003, (* equally contributing authors).

- Roger Blanco i Ribera*, Eduard Zell*, J.P. Lewis, Junyong Noh, Mario Botsch (2017), 'Facial Retargeting with Automatic Range of Motion Alignment', *ACM Transactions on Graphics* **36**(4), (* equally contributing authors).

Additional Achievements and Publications

In addition to the work that is part of this thesis, I was involved in an industry project with Adidas with the goal to fit shoe templates to 3D scans. I have also participated in the CITmed research project and a collaboration on motion tracking, animation and visualization of stick insects with results published in the following articles:

- Eugen Dyck, Eduard Zell, Agnes Kohsik, Philip Grewe, York Winter, Martina Piefke, Mario Botsch (2012), 'OctaVis: An Easy-to-Use VR-System for Clinical Studies'*Proc. of Virtual Reality Interaction and Physical Simulation (VRIPHYS)*, pp. 127-136.

- Eduard Zell, Eugen Dyck, Agnes Kohsik, Philip Grewe, David Flentge, York Winter, Martina Piefke, Mario Botsch (2013), 'OctaVis: A Virtual Reality System for Clinical Studies and Rehabilitation', *Eurographics – Dirk Bartz Prize for Visual Computing in Medicine (3rd Place)*, pp. 9-12.

- Leslie Theunissen, Michael Hertrich, Cord Wiljes, Eduard Zell, Christian Behler, Andre Krause, Holger Bekemeier, Philipp Cimiano, Mario Botsch, Volker Dürr (2014) 'A Natural Movement Database for Management, Documentation, Visualization, Mining and Modeling of Locomotion Experiments', *Proc. of Living Machines*, pp. 308-319.

Part I

Algorithms for Realistic and Stylized Stimuli

This part of the thesis addresses technical challenges involved in creating virtual faces for perceptual experiments. The first chapter describes our pipeline for creating virtual doubles of real people. This involves various aspects, starting from 3D scanning over template fitting up to rendering and some artistic elements like the reconstruction of hairstyles. Our pipeline addresses practical issues for creating high-quality stimuli that we compare against the state-of-the-art in research.

The second chapter presents ELASTIFACE, a method for establishing dense correspondences and transferring textures between faces. This method created a basis for an independent investigation on the perception of shape and material stylization. Besides material transfer, this algorithm can be also used for creating blendshapes from different scans for both realistic and stylized characters.

The third chapter provides an algorithm for facial animation transfer, also known as facial retargeting. In this case, we specifically address the problem of animation transfer between realistic and stylized characters. While facial retargeting has a wide range of applications, our intention within the perception context is to create an animated stimuli set with consistent facial animation in the future.

2 Creating Realistic Facial Stimuli

(a) Scan   (b) Fit   (c) Model   (d) Rendering
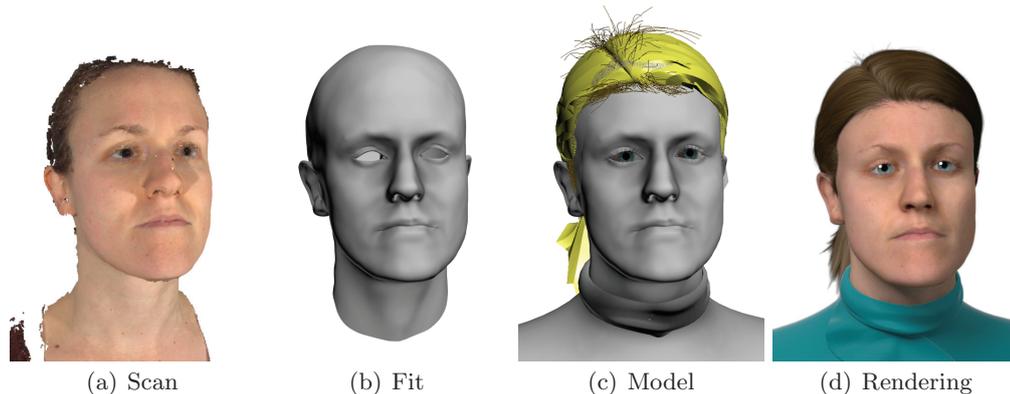
Figure 2.1: Intermediate and final results created with our pipeline for replicating realistic characters. After obtaining a high-resolution scan (a) a template is fitted towards the model (b). By adding eyes, teeth, hair and clothes the model is completed (c) and finally rendered (d).

In the introduction, we mentioned already that one shortcoming of many perceptual experiments investigating the impact of facial stylization is the absence of several stylization levels of the same identity. While there is still no automatic approach that reliably creates different stylization levels for a single character, various algorithms have been suggested to replicate different facial parts of real people. Our primary focus at this stage is to establish a pipeline to create realistic virtual doubles for perception experiments (Figure 2.1). This includes the reconstruction of skin, eyes, teeth or hair and advanced rendering algorithms for all these different facial parts. The presented character creation pipeline was state-of-the-art when the project started in 2013. In the following subsections, we provide an overview of our pipeline and techniques involved to create realistic characters together with recent developments in research and commercially available solutions. Whenever new software was released that simplifies extending or re-creating realistic stimuli, we name it as suitable alternatives. Although the presented pipeline focuses on the replication of realistic characters, the main difference to stylized characters is the acquisition process of 3D geometry and textures. Such data is typically obtained from 3D scans for realistic characters but created by artists for stylized characters.

My contribution to this work was an evaluation of published work in academic context together with existing (commercial) software solutions. Based on this analysis, a pipeline was built such that the input and output of different software packages were compatible with each other. In addition, software for fitting a template model to a scan, named as non-rigid registration, was developed and integrated within the pipeline. Analysis of the pipeline revealed that 3D model creation from a scan required a significant amount of manual corrections, especially around the eyes,

Figure 2.2: Our face scanning setup (*right*) and comparison between photographs and virtual reconstructions of our actor (*left*).

and that common design decisions, e.g., computing closest point correspondences from the template to the point cloud, do not lead to optimal results. Based on these insights, further automatization and evaluation of non-rigid registration algorithms (Achenbach et al. 2015) have been developed in a project under the lead of Jascha Achenbach.

2.1 Data Acquisition

Face Scanning A 3D scan creates the best representation in 3D of a real character. For our use case, technology is of high interest that creates complete and high-quality color scans at acquisition times lower than a second. Short acquisition times are required to scan facial expressions without motion blur artifacts, which in turn disqualifies most laser scanners. The high-quality requirement excludes most depth camera approaches that rely on low-resolution devices[1]. Instead, photogrammetry or multi-view stereo, where 3D objects are reconstructed from several high-resolution photographs, combine high quality with short acquisition times (e.g., (Ma et al. 2007, Beeler et al. 2010, Fyffe et al. 2016)). Any multi-view stereo system requires an array of high-resolution cameras capable of taking pictures simultaneously. Using wireless remote triggers proved to be most reliable approaches in practice. While some methods focus only on the reconstruction of accurate geometry (Bradley et al. 2010, Valgaerts et al. 2012), or geometry in combination with mesoscopic features (Beeler et al. 2010), other methods compute also texture maps like albedo or specular normals (Ma et al. 2007, Ghosh et al. 2011, Fyffe et al. 2016). However, the latter typically require controlled lighting and advanced hardware setups like a light stage. Very recently, methods have been suggested that reconstruct faces based on single view photographs or videos (Cao et al. 2015, Garrido et al. 2016, Thies et al. 2016) or low-resolution facial scans (Weise et al. 2011, Bouaziz et al. 2013, Li et al. 2013, Thies et al. 2015). Although the results are impressive, these methods rely highly on the quality and variety of facial databases. The models

---

[1]E.g., Kinect, RealSense, ZED or Orbbec Persee.

in the database constrain the model to valid faces, but unfortunately bias as well the result, such that the reconstruction is only close to the target face.

In order to generate the realistic models, we replicated the multi-view stereo face scanner of Beeler et al. (2010), which is a good trade-off between quality and hardware demands. It reconstructs high-resolution point clouds from photographs (Figure 2.2) without any priors. Since all photographs are taken simultaneously, the scanning process is instantaneous and therefore well suited for capturing different facial expressions.

One serious issue of state-of-the-art face reconstruction methods is the complexity of the methods paired with limited availability due to patents (e.g., (Beeler et al. 2012, Ghosh et al. 2011)) or closed source. Recently, this problem can be circumvented in practice by using custom hardware setups in combination with general purpose photogrammetry software like AgiSoft or CaptureReality[2], specialized systems (Medusa, 3dMD) or service providers (Ten24, Infinity Realities). While service providers and software packages differ in quality, any of these three options is suitable for obtaining high-quality scans. The optimal solution is therefore primarily dependent on the personal skill set, the budget and the time available.

Pores and Small Features By using high-resolution cameras, it is possible to capture geometry up to a detail level of $1\,\text{mm}$ or less. Regarding size, fine-scale features are between geometric features, represented by triangles and the surface reflectance model, simulated via a Bidirectional Scattering Distribution Function (BSDF) that is evaluated at render time (see also Section 2.4). Because even small skin structures are essential for convincing realism in close-ups, facial microstructures have been measured for static faces (Graham et al. 2013) or have been computed procedurally (von der Pahlen et al. 2014). Later, Nagano et al. (2015) demonstrated that facial microstructures change under skin deformation and influence the skin appearance. A common approximation for computing microstructures is the "dark-is-deep" heuristic (Beeler et al. 2010). While such heuristics create stunning results, visual differences persist to correctly reconstructed pores (Fyffe et al. 2016).

For our characters we followed the procedural idea proposed by von der Pahlen et al. (2014) and transferred fine-scale details to our characters from other high-resolution bump maps. Considering the visual differences between every stylization level of each character, we assumed that the perceived differences of this simplification would be rather small. Indeed, the results of our perception experiments show that participants did not even recognize more significant changes in a scene, e.g., changing from soft indirect illumination to a direct spot light with hard shadows (see also Section 6.3.2).

Lighting and Camera Besides the facial properties itself, recent publications also approximate external parameters, like camera matrix or illumination (Blanz & Vetter 1999, Thies et al. 2015, Shahlaei & Blanz 2015, Conde et al. 2015). In our experiments, we wanted to minimize possible side effects due to, e.g., a realistic environment as much as possible. All characters have been rendered using a simple three-point lighting setup that (i) does not change between the characters and (ii) is a plausible environment for each stylization level. Thus measuring accurately lighting and camera parameters was not required in our case.

---

[2]Other software packages are: 123D Catch, 3DF Zephyr, Acute3D, or Neitra 3D Pro.

2.2 Template Fitting

Scanning provides accurate spatial information of data, but only visible parts may be reconstructed, which was in our case just the frontal part of the face. To overcome these limitations, a variety of methods reconstructs the face based on a template model, that is continuously deformed towards a scan (Li et al. 2009, Weise et al. 2011) or even a picture (Blanz & Vetter 1999, Thies et al. 2016). The advantage of template fitting is that accuracy remains where data is present and missing information is plausibly approximated. Robustness of template fitting is increased by using a database of faces as a template instead of a single face (Blanz & Vetter 1999, Thies et al. 2015, Ichim et al. 2015).

In our initial setup, we had no access to a parametric model and relied on a linear non-rigid deformation method in to fit a template to the reconstructed point cloud. Appearing inaccuracies during stimuli creation were compensated through manual corrections, e.g., additional correspondences. Based on the observed limitations, the non-rigid registration has been further analyzed and improved (Achenbach et al. 2015). We report primarily the outcome of our analysis of different template fitting algorithms in this subsection and recommend readers interested in replicating our pipeline the source code of the face reconstruction tutorial provided by Bouaziz et al. (2014).

Given a facial template as a polygon mesh $\mathcal{M}$, consisting of N vertices $\mathbf{v}_i$, and let be $\mathcal{P}$ a point cloud, representing a scan of a person, with P points in total, where $\mathbf{p}_j$ represents the position of each point. For every point $\mathbf{p}_j$ a unique normal vector $\mathbf{n}_j$ is associated. Our goal is to deform the template model $\mathcal{M}$ such that it approximates the point cloud $\mathcal{P}$ as much as possible.

Rigid alignment Any template fitting starts with a rigid registration, where the optimal global rotation $\mathbf{R}$ and translation $\mathbf{t}$ are computed between the template and the scan. Unless a rough alignment can be assumed, the first optimal transformation is calculated based on a few point correspondences. Such correspondences can be either selected manually or computed automatically using computer vision algorithms (Asthana et al. 2013). Given C corresponding pairs $(\mathbf{c}_j, \mathbf{p}_j)_j$ with $C \leq P$ and $\mathbf{p}_j$ as a point of the target point cloud together with the correspondence point $\mathbf{c}_j$ on the template mesh, an optimal rigid fit is obtained by minimizing the following energy (Besl & McKay 1992):

$$E_{Rigid}(\mathbf{R}, \mathbf{t}) = \sum_{j=1}^{C} \|\mathbf{p}_j - (\mathbf{R}\mathbf{c}_j + \mathbf{t})\|^2 \ . \tag{2.1}$$

The initial alignment is then refined using the Iterative Closest Point (ICP) algorithm (Besl & McKay 1992, Chen & Medioni 1992, Rusinkiewicz & Levoy 2001). First, correspondences are established by searching for each point $\mathbf{p}_j$ of the scan the closest neighbor $\mathbf{c}_j$ on the template mesh. Second, based on these correspondences, Equation (2.1) is solved for the optimal global rotation matrix $\mathbf{R}$ and the global translation vector $\mathbf{t}$. This procedure is repeated until convergence.

As closest point is a very basic heuristic for approximating correspondences, additional heuristics (Rusinkiewicz & Levoy 2001, Chang et al. 2010) have been introduced to increase reliability (Figure 2.3). First, the distance between two points should be within a threshold. Second, the normals of the closest points should be similar, and third, correspondence points should not be located on a mesh boundary.

(a) far off correspondences | (b) different normals | (c) shape boundary
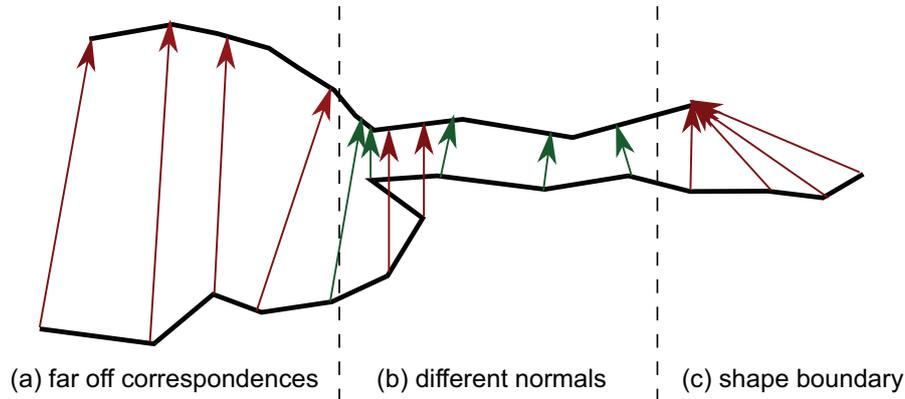
Figure 2.3: Common heuristics for pruning wrong correspondences. Green arrows encode valid closest point correspondences, while red arrows demonstrate three types of invalid correspondences: (a) far distances, (b) different normal direction, (c) matching towards a shape boundary.

Parametric model If the template model consists of $\hat{L} \gg 1$ faces with identical topology, the scanned head can be approximated by interpolating between different template faces. Fitting a morphable model to a scan or an image has been initially suggested by Blanz & Vetter (1999) and has been adopted in subsequent works, e.g., (Chen et al. 2014, Bouaziz et al. 2014, Thies et al. 2015). As faces are highly similar, a dimensionally reduced Principal Component Analysis (PCA) model is often preferred in such cases. The PCA model is constructed by first computing the average face $\bar{\mathbf{v}} = (\bar{\mathbf{v}}_1^x, \bar{\mathbf{v}}_1^y, \bar{\mathbf{v}}_1^z, \dots, \bar{\mathbf{v}}_N^x, \bar{\mathbf{v}}_N^y, \bar{\mathbf{v}}_N^z)^\mathsf{T}$ of all template faces, and saving all faces of the dataset as displacements with regard to the average face in a data matrix

$$\left[\mathbf{v}_1 - \bar{\mathbf{v}}, \, \dots \, , \mathbf{v}_{\hat{L}} - \bar{\mathbf{v}}\right]. \tag{2.2}$$

After performing the principal component analysis by singular value decomposition of the data matrix, the facial variance is encoded by the orthogonal displacement vectors $\mathbf{d}^l = (\mathbf{d}_1^l, \dots, \mathbf{d}_{3N}^l)^\mathsf{T}$. The amount L of principal components is $L \leq \hat{L}$ and can be further reduced by considering only the first principal components.

Equivalently to the rigid ICP algorithm, corresponding points $\mathbf{c}_j$ on the template mesh are computed for every point $\mathbf{p}_i$ on the scan. Because the correspondence points $\mathbf{c}_j$ are located on mesh triangles, but the PCA is formulated in terms of vertex positions, every correspondence point is encoded using barycentric coordinates.

$$\mathbf{c}_j = \alpha_j \mathbf{v}_\alpha + \beta_j \mathbf{v}_\beta + \gamma_j \mathbf{v}_\gamma. \tag{2.3}$$

Vertex positions of objects created by a parametric model are computed as a linear combination of the displacement vectors,

$$\mathbf{v}(\mathbf{w}) = \bar{\mathbf{v}} + \sum_{l=1}^{L} w_l \mathbf{d}_l = \bar{\mathbf{v}} + \mathbf{D}\mathbf{w}, \tag{2.4}$$

with $\mathbf{w} = (w_1, \dots, w_L)^\mathsf{T}$ being the weighting coefficients and $\mathbf{v} = (\mathbf{v}_1^x, \mathbf{v}_1^y, \mathbf{v}_1^z, \dots, \mathbf{v}_N^x, \mathbf{v}_N^y, \mathbf{v}_N^z)^\mathsf{T}$ containing all vertex positions. Furthermore, the sum $\sum_{l=1}^{L} w_l \mathbf{d}_l$ can be

| (a) Scan | (b) Noise | (c) Direction | (d) Fitting |
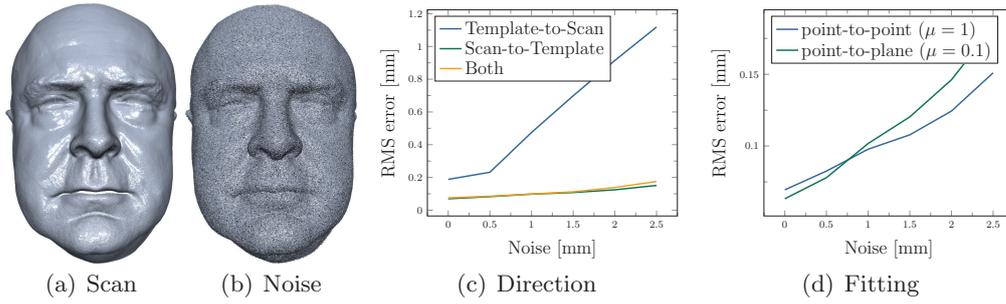
Figure 2.4: (a) A high-resolution face model from Beeler et al. (2011) and (b) a noisy point cloud by adding varying amounts of noise ($\pm 2\,\text{mm}$). (c) Comparison of correspondence directions (*templace-to-scan* vs. *scan-to-template*, and (d) correspondence distance (*point-to-point* vs. *point-to-plane*, on the high-resolution face scan with synthetically added noise (uniformly distributed, zero mean). ©Scan: Beeler et al. (2010)

written as a matrix-vector product $\mathbf{Dw}$. By combining Equations (2.3) and (2.4), we express the vector $\mathbf{c} = (\mathbf{c}_1^x, \mathbf{c}_1^y, \mathbf{c}_1^z, \ldots, \mathbf{c}_C^x, \mathbf{c}_C^y, \mathbf{c}_C^z)^\mathsf{T}$ in terms of the weighting coefficients:

$$\mathbf{c}(\mathbf{w}) = [\alpha\beta\gamma](\bar{\mathbf{v}} + \mathbf{Dw}) . \tag{2.5}$$

The matrix $[\alpha\beta\gamma]$ of size $3C \times 3N$ is constructed from the barycentric coordinates from Equation (2.3).

The best fit of the parametric model to the scan is obtained by minimizing the combination of the two energies: $E(\mathbf{w}) = E_{Model}(\mathbf{w}) + \mu E_{Prior}(\mathbf{w})$. The energy

$$E_{Model}(\mathbf{w}) = \frac{1}{C}\sum_{j=1}^{C} \|\mathbf{p}_j - \mathbf{c}_j\|^2 = \frac{1}{C} \|\mathbf{p} - [\alpha\beta\gamma](\bar{\mathbf{v}} + \mathbf{Dw})\|^2 , \tag{2.6}$$

fits the model to the scan by minimizing the distance between all C points $\mathbf{p}_j$ in the scan and according correspondence points $\mathbf{c}_j$ on the mesh. In the compact representation, all points of the scan $\mathbf{p} = (\mathbf{p}_1^x, \mathbf{p}_1^y, \mathbf{p}_1^z, \ldots, \mathbf{p}_C^x, \mathbf{p}_C^y, \mathbf{p}_C^z,)^\mathsf{T}$ are stacked in one vector. In order to prevent overfitting, a regularization energy is added based on the probability of the coefficients w_l.

$$E_{Prior}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \Lambda \mathbf{w} . \tag{2.7}$$

Λ is defined as the diagonal matrix $diag(1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma_L^2)$, and σ_l^2 are the eigenvalues of the covariance matrix

$$[\mathbf{v}_1 - \bar{\mathbf{v}}, \ldots, \mathbf{v}_L - \bar{\mathbf{v}}]^\mathsf{T}[\mathbf{v}_1 - \bar{\mathbf{v}}, \ldots, \mathbf{v}_L - \bar{\mathbf{v}}] , \tag{2.8}$$

or equivalently: σ_l are the singular values of the data matrix. The combination of Equations (2.6) and (2.7) is minimized using traditional linear least squares. This involves solving a linear system of size $L \times L$.

Non-rigid Registration Once the parametric model is fitted towards the target point cloud, the remaining step is a non-rigid mesh registration based on closest point constraints (Allen et al. 2003, Sumner & Popović 2004, Amberg et al. 2007,
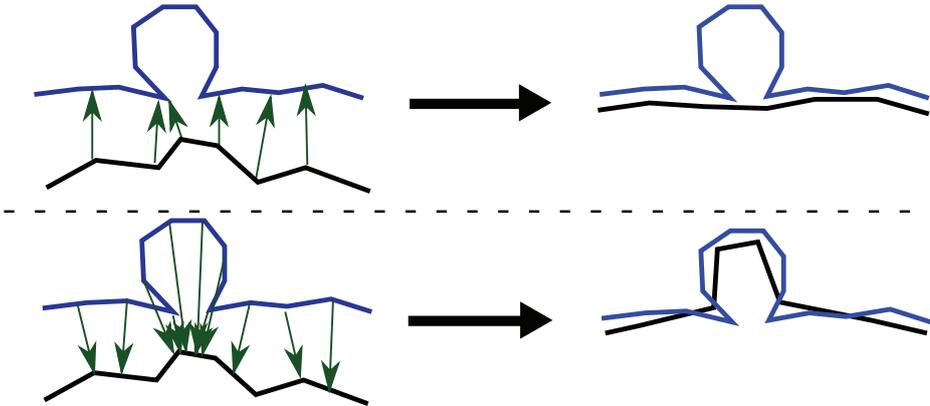
Figure 2.5: Illustrative examples are demonstrating the results for different closest point correspondences: (*top*) *template-to-scan*, (*bottom*) *scan-to-template*. Only for the *scan-to-template* example, closest point constraints drag the template (*black*) towards the bump feature of the target (*blue*).

Brown & Rusinkiewicz 2007, Li et al. 2008, Tam et al. 2013). At this stage, the template mesh is continuously deformed towards the scan. The alignment error is measured by $E_{Fit}(\mathbf{v})$ with $\mathbf{v} = (\mathbf{v}_1^x, \mathbf{v}_1^y, \mathbf{v}_1^z, \ldots, \mathbf{v}_N^x, \mathbf{v}_N^y, \mathbf{v}_N^z)^{\mathsf{T}}$ containing all unknown vertex positions after the non-rigid registration. Adding a regularization term to the total energy $(E(\mathbf{v}) = E_{Fit}(\mathbf{v}) + \lambda E_{Reg}(\mathbf{v}))$ penalizes strong deformations of the template mesh. Similar to rigid ICP, correspondences between scan and template mesh are computed using closest point search. The template mesh is deformed towards the target by continuously reducing λ. In case of high-resolution meshes, performance gains can be achieved using hierarchical approaches (Bonarrigo et al. 2014, Achenbach et al. 2015). Here, a low-resolution template is first deformed towards a sub-sampled target. After sufficient alignment, the template is subdivided and the process is repeated until the desired resolution level.

Two metrics exist for the fitting energy. The *point-to-point* energy minimizes the quadratic distance between a correspondence pair. The *point-to-plane* energy allows in addition sliding of the template along the point cloud, leading to faster convergence. Following Li et al. (2008, 2009) and Bouaziz et al. (2014), the fitting energy is defined as the linear combination of the two metrics with $\mu = 0.1$.

$$E_{Fit}(\mathbf{v}) = \frac{1}{N} \sum_{i=1}^{N} \mu \underbrace{\|\mathbf{v}_i - \mathbf{c}_i\|^2}_{\text{point-to-point}} + (1 - \mu) \underbrace{(\mathbf{n}_i^{\mathsf{T}}(\mathbf{v}_i - \mathbf{c}_i))^2}_{\text{point-to-plane}} . \tag{2.9}$$

In our experiments, the *point-to-point* distance requires indeed 30% more iterations until convergence than the combined distance measure. Nevertheless, more iterations until convergence do not necessarily lead to slower computation times. With the *point-to-point* metric alone, the problem is separable in $x/y/z$, leading to three $N \times N$ linear systems. In contrast, the *point-to-plane* distance couples the coordinates, leading to a $3N \times 3N$ linear system. Solving the larger linear system increased in our experiments computation times by factor three. Because both energy formulations converge to a comparable fit (Figure 2.4), we recommend to use the *point-to-point* metric.

template-to-scan   scan-to-template
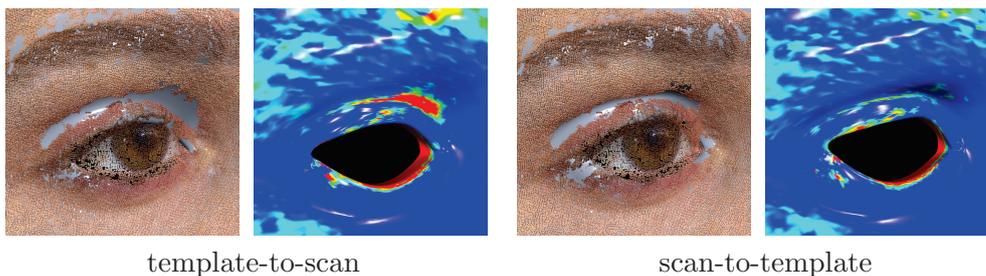
Figure 2.6: For high-resolution scanner data, *scan-to-template* correspondences (*right*) yield towards a more accurate reconstruction than the typically employed *template-to-scan* correspondences (*left*). The bottom row shows a color-coding of the two-sided Hausdorff distance of scan and template.

Besides the fitting energy itself, closest point correspondences can be computed in two directions – from *template-to-scan* and from *scan-to-template* (Figure 2.5). The former identifies for each of the N template vertices $\mathbf{v}_i$ the closest point in $\mathcal{P}$, while the latter finds for each of the P points $\mathbf{p}_j$ its closest neighbor on the template mesh $\mathcal{M}$. In practice $N \ll P$, thus the computational complexity of the *template-to-scan* correspondences is lower ($\mathcal{O}(N \log P)$ vs. $\mathcal{O}(P \log N)$). However, due to the results obtained with *scan-to-template* correspondences are more accurate (Figures 2.4, 2.6) and should be chosen if accuracy is preferred over performance. Therefore, we reformulate the *point-to-point* fitting energy from Equation (2.9) as following:

$$E_{Fit}(\mathbf{v}) = \frac{1}{C} \sum_{j=1}^{C} \|\mathbf{p}_j - \mathbf{c}_j(\mathbf{v}_{\alpha\beta\gamma})\|^2 \ . \tag{2.10}$$

In this case, the correspondences $\mathbf{c}_j(\mathbf{v}_{\alpha\beta\gamma})$ on the template mesh are expressed by barycentric coordinates of the vertices $\mathbf{v}_i$ (see Equation 2.3).

The regularization energy E_{Reg} is responsible for smooth surface deformation of the template model towards the target scan. A *linear* version, similar to Allen et al. (2003) and Sumner & Popović (2004) and a *non-linear* as-rigid-as-possible deformation model (Sorkine et al. 2004) are encountered most frequently. Both energies minimize bending and are thus discussed together. The *linear* model penalizes deviations between the original $\Delta\hat{\mathbf{v}}_i$ and the new $\Delta\mathbf{v}_i$ vertex Laplacian.

$$E_{Bend}(\mathbf{v}) = \frac{1}{\sum_{i=1}^{N} A_i} \sum_{i=1}^{N} A_i \|\Delta(\hat{\mathbf{v}}_i - \mathbf{v}_i)\|^2 \ . \tag{2.11}$$

In addition, the *non-linear* variant computes the optimal rotations $\mathbf{R}_i$ within the one ring neighborhood of each vertex and thus preserves the shape better under strong rotations.

$$E_{ARAP}(\mathbf{v}, \mathbf{R}_1, \ldots, \mathbf{R}_N) = \frac{1}{\sum_{i=1}^{N} A_i} \sum_{i=1}^{N} A_i \|\Delta\mathbf{v}_i - \mathbf{R}_i \Delta\hat{\mathbf{v}}_i\|^2 \ . \tag{2.12}$$

In both expressions, the Laplacian is discretized using the cotangent weights (Pinkall & Polthier 1993, Meyer et al. 2003), and the area A_i is the local Voronoi area of a vertex i. Both energies have the advantage that the resulting systems are separable

(a) Template   (b) Linear   (c) Non-Linear   (d) Regularization
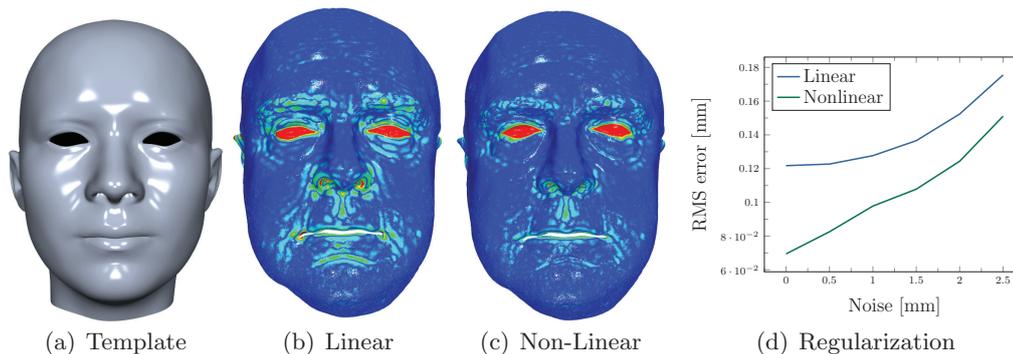
Figure 2.7: A template model (a) is fitted towards a high-resolution face model (Beeler et al. 2011) with added noise ($\pm 2\,\text{mm}$) using a linear (b) and a non-linear (c) regularization energy. Errors of more than $1\,\text{mm}$ are shown in red. (d) Comparison of regularization energies (*linear* vs. *non-linear*) on the high-resolution face scan with varying amount of noise (uniformly distributed, zero mean). ©Template: Chen et al. (2014) ©Scan: Beeler et al. (2010)

in $x/y/z$ and require only $N \times N$ systems to solve. The final solution for the *linear* bending model is obtained by solving once the resulting linear system. In contrast, the *non-linear* as-rigid-as-possible variant requires alternating optimization for the vertex positions $\mathbf{v}_i$ and local rotation matrix $\mathbf{R}_i$. In our experiments, convergence was typically reached after 2–3 iterations. While the computation time for the non-linear method is ten times higher, in our experiments the Root Mean Square (RMS) error of the linear model is twice as large as that of the non-linear model (Figure 2.7). Therefore, we recommend the *non-linear* as-rigid-as-possible regularization energy.

Another advantage of fitting a template model to a scan is that the template provides a 2D parametrization of the model, which is required for texture mapping. Textures can be generated automatically by sampling the color information of each pixel from the point cloud. If the camera matrices are available, that are computed during the point cloud reconstruction and that map the captured photographs on the scan, the color information can be extracted directly from the photographs. Both algorithms represent a variation of the texture transfer algorithm presented in Section 3.3 and are available in software frameworks like PhotoScan by AgiSoft.

2.3 Eyes, Teeth and Hair

Most face reconstruction algorithms do not address the reconstruction of critical facial parts like eyes, teeth or hair. To overcome this issue, methods exist that focus on specific facial parts. François et al. (2009) focused on the acquisition of accurate color textures and normal maps of the iris. High-quality scanning of eyes using a multi-view setup in combination with active lighting for capturing the sclera, cornea and iris was presented after the start of our project by Bérard et al. (2014). Very recently, the same authors (Bérard et al. 2016) suggested a data-based approach to create eyes from a low-quality input. Similarly, a morphable model is the core of the teeth reconstruction method of Wu et al. (2016). In contrast, Bermano et al. (2015) couple geometric deformation with image sequences to reconstruct moving eyelids. Other work focused on facial hair creation. Herrera et al. (2010) detect reliably hairy
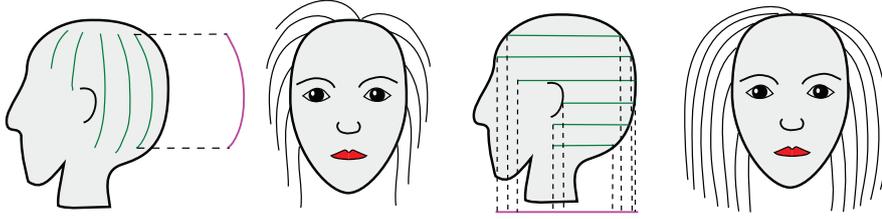
Figure 2.8: Schematic illustration on the layering of hair when cutting real hair and the importance of hair length for the overall hairstyle. Green lines represent groups of hair and the purple line the cutting line. *Left:* Horizontal layering with equal hair length and the resulting hairstyle. *Right:* Vertical layering with a cutting line above the shoulder and the resulting hairstyle.

(a) Hair Layers   (b) Hair Mesh   (c) Hair Strands   (d) Rendering
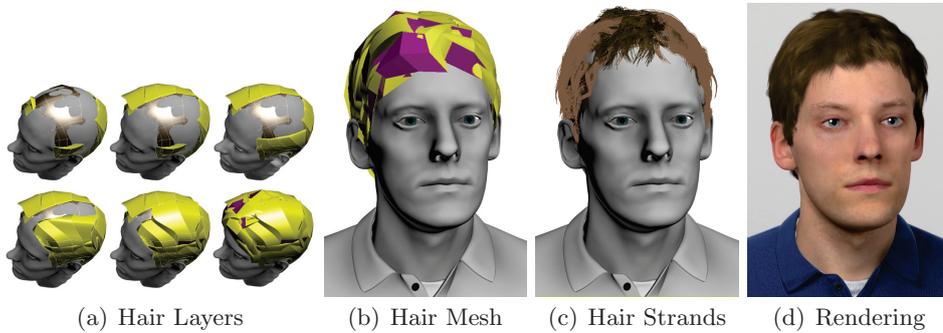
Figure 2.9: (a) Decomposition of the final hairstyle in different layers. The same character is shown using internal representations of hair as (b) hair meshes or as (c) hair strands (5% of total hair fibers) together with (d) the final rendering.

regions from images using a database. While his approach is limited to short hair, the method of Beeler et al. (2012) is even capable to reconstruct complex beards with long facial hairs in combination with the face.

Reconstruction of facial parts like teeth or eyes is in particular difficult because of small size, occlusion or transparency. While the small size and occlusion make reconstruction difficult, it also means that big parts of eyes or teeth remain hidden. Therefore, we used the teeth, gums and tongue from a template model and manually adjusted it to the facial proportions of our character using Free-Form Deformation (FFD), which is available within all major 3D packages. For the eyes, our template model consisted of two merged spheres for the sclera and cornea and a flat torus for the iris. The according textures have been adjusted by globally shifting the hue, brightness and saturation in order to match the color of the captured person's iris. Although simpler, this approach turned out to produce better results and be more effective than extracting eye color from the scan, which is very noisy around the eyes due to partial occlusion by eyelashes.

Digital hair in computer graphics is created by defining a few guide hairs and interpolating the remaining hair between the guide strands. For an overview of standard hair creation techniques, we recommend the survey of Ward et al. (2007). An interesting alternative, especially for more sophisticated hairstyles, are hair meshes

(Yuksel et al. 2009), that we used to replicate the hairstyles. The technology is available as part of Hairfarm[3], a plug-in of 3Ds Max[4]. In this case, the hull of hair strands is created similar to polygon modeling techniques and later filled with hair fibers. Interestingly, most hair reconstruction methods follow a similar approach (Paris et al. 2004, Wei et al. 2005). First, edge detection filters are applied to identify the main hair structures. Second, hair is created matching the flow of the detected edges and the silhouette. More advanced methods reconstruct structures in combination with shading using a light stage (Paris et al. 2008). Rather than just matching the overall hairstyle, Jakob et al. (2009) focus on reconstructing hair on fiber level, by using several images taken at shallow depth of field. Similar, Luo et al. (2013) reconstruct fibers, group them to ribbons and optimize for connections. Their method is even capable to reconstruct long curly hair. By using thermal imaging, Herrera et al. (2011) overcome most difficulties of multi-view hair reconstruction like hair detection, depth approximation or fiber segmentation due to anisotropic shading paired with multiple scattering effects. Recent developments focus on hair reconstruction from single images with little (Chai et al. 2012, 2015, Hu et al. 2015) to no user input (Chai et al. 2016), but rely in many cases on databases of hair styles. Due to the amount of fibers of real hair and the resulting overlapping and self-occlusions, all discussed methods approximate the visual appearance rather than fully reconstruct a hair style.

We decided to model our hairstyles in Hairfarm following cutting and stylization guidelines for real hair. The appearance and nature of hair are volumetric thanks to the thousands of different hair strands. In order to control the vast amount of different hair fibers, haircuts are divided into layers or groups (Figure 2.8, green lines). In the simple, but very frequent case all hair within a group has identical length, but for more advanced haircuts the length of hair within a layer might change gradually. Defining well the length of hair has a big impact on the shape of the haircut, especially in case of thick and curly hair (Figures 2.8 and 2.9). Once the length and basic shape of the haircut are defined, the actual hairstyling begins. This involves a finer partition of the layers and redefinition of the shapes of different hair fibers. The very last polishing step includes the creation of separate hair strands and loosely spread hair.

2.4   Rendering

The analysis and computation of realistic light-surface interaction have a long tradition in computer graphics. In general, material models in computer graphics can be separated in appearance-based models and physically-based shading. While the former models specify visual properties using intuitive parameters for artists, the latter obey different laws of physics, e.g., energy conservation. Although physically-based shading (Hill et al. 2015, McAuley & Hill 2016) is preferred over phenomenological models like the Phong shading model (Phong 1975) for realistic characters, we will show in the next paragraphs that the difference between the two models is blurry, and appearance-based material models can be converted to physically-based shading models. This means that a specific set of parameters of the Phong shading model actually represents physically correct appearance. Our analysis of different rendering

---

[3]http://www.hair-farm.com
[4]http://www.autodesk.com

solutions revealed that no single solution existed that combines all recent advancements from the research community out of the box. Lacking an optimal solution, we selected the combination of Mental Ray[5] with Hairfarm as the best compromise in terms of speed/quality. However, with the publication of the DigitalEmily 2.0 dataset[6] the combination of Vray[7] with Hairfarm should also be considered.

Skin Rendering Rendering skin is especially challenging because of its layered consistency. In general, surface reflectance is modeled by separating the upper oily layer (specular lighting) from the underlying fleshy skin (subsurface scattering). The resulting material is thus a combination of a Bidirectional Reflectance Distribution Function (BRDF) for the specular part and a Bidirectional Surface Scattering Distribution Function (BSSDF) for the diffuse/subsurface scattering part. In the following, we review the common physically-based models for skin rendering and their connection to appearance-based shading models like the Phong shading. The intention behind this detailed description is to show that the boundary between appearance-based and physically-based shading models is somewhat blurry.

For skin rendering, specular lighting is approximated using an analytic BRDF based on the microfacet distribution model (Torrance & Sparrow 1967). The general Cook-Torance BRDF is defined as:

$$f_R(\mathbf{l}, \mathbf{n}, \mathbf{v}) = \frac{D(\mathbf{h})F(\mathbf{v}, \mathbf{h})G(\mathbf{l}, \mathbf{v}, \mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})} \; . \tag{2.13}$$

Here, $\mathbf{l}$ is the light direction, $\mathbf{v}$ the view direction, $\mathbf{n}$ the surface normal, and $\mathbf{h} = \mathbf{n} + \mathbf{l}/\|\mathbf{n} + \mathbf{l}\|$ the half-vector between $\mathbf{l}$ and $\mathbf{n}$. All input vectors are used in their normalized form. The function $D(\mathbf{h})$ is the normal distribution function, $F(\mathbf{v}, \mathbf{h})$ the Fresnel reflectance and $G(\mathbf{l}, \mathbf{v}, \mathbf{h})$ the geometry function.

The Fresnel reflectance is best understood so far and describes the amount of light that is reflected by specific incident angles. Even though an exact solution exists for the Fresnel reflectance (Pharr & Humphreys 2010), computation time can be saved using Schlick's approximation (Schlick 1994) without losing too much accuracy. In addition, Schlick's expression has more intuitive input parameters and became thus a common alternative for practical applications.

$$F_{Schlick}(\mathbf{F}_0, \mathbf{l}, \mathbf{h}) = \mathbf{F}_0 + (\mathbf{1} - \mathbf{F}_0)(1 - (\mathbf{l} \cdot \mathbf{h}))^5 \; . \tag{2.14}$$

$\mathbf{F}_0$ is the surface's characteristic specular color, which is non-colored and rather low $\max(\mathbf{F}_0^r, \mathbf{F}_0^g, \mathbf{F}_0^b) < 0.2$ for dielectric materials like skin, and colored and high $0.34 < \mathbf{F}_0^r, \mathbf{F}_0^g, \mathbf{F}_0^b < 1.02$ for shiny materials like metals.

The Normal Distribution Function (NDF) describes the amount of light that is reflected towards the direction of the half-vector. If the microfacets' normals are equally distributed in all directions, a perfectly rough surface is given. The geometry shadowing function considers the fraction of light that is either blocked by the microsurface structure or amplified due to inter-reflections. For the NDF and the respective geometry shadowing, different analytic functions were proposed, but only a fraction was applied to skin rendering. Weyrich et al. (2006) assumed that

---

the Torrance-Sparrow model (Torrance & Sparrow 1967) is most accurate.

$$G_{TS}(\mathbf{l}, \mathbf{v}) = \min\left(1, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{v})}{\mathbf{v} \cdot \mathbf{h}}, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{l})}{\mathbf{l} \cdot \mathbf{h}}\right). \tag{2.15}$$

The Torrance-Sparrow model is based on the Beckmann distribution (Beckmann & Spizzichino 1963) with $m \in [0, 1]$ being the roughness parameter and $m = 0$ defining a perfectly smooth surface.

$$D_{Beckmann}(\mathbf{h}, m) = \frac{1}{\pi m^2 (\mathbf{n} \cdot \mathbf{h})^4} e^{\frac{(\mathbf{n} \cdot \mathbf{h})^2 - 1}{m^2 (\mathbf{n} \cdot \mathbf{h})^2}}. \tag{2.16}$$

Based on physical measurements, Weyrich et al. (2006) created a reference for the roughness term for specific parts of the face. In their highly influential paper, d'Eon et al. (2007) suggested to replace the Torrance-Sparrow geometry shadowing function with the simplified version from Kelemen & Szirmay-Kalos (2001).

$$\frac{G_{Kelemen}(\mathbf{l}, \mathbf{v})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})} \approx \frac{1}{\mathbf{h}_u \cdot \mathbf{h}_u} = \frac{1}{4(\mathbf{l} \cdot \mathbf{h})^2}. \tag{2.17}$$

Here, $\mathbf{h}_u = \mathbf{n} + \mathbf{l}$ is the unnormalized half-vector computed from normalized normal and light vectors. Interestingly, Kelemen also proposed to approximate the Beckmann distribution by the Blinn-Phong model. Ngan et al. (2005) and Walter et al. (2007) fitted various analytical models to empirical data and reported that the Torrance-Sparrow model approximates the data best. Nevertheless, the Blinn-Phong model approximates well the Beckmann distribution for low roughnesses, which is the case for the skin. In addition, Ngan et al. (2005) demonstrated that combining two specular lobes improves fitting quality significantly. In fact, an energy conserving form exists for the Blinn-Phong model (Blinn 1977), which addresses the main shortcoming of the Blinn-Phong model.

$$D_{Phong}(\mathbf{h}, c) = \frac{(c + 2)}{2\pi} (\mathbf{n} \cdot \mathbf{h})^c. \tag{2.18}$$

Considering the connection between the power coefficient c and the roughness m in microfacet models, it is possible to express the Blinn-Phong distribution as a variable of the roughness m. This illustrates that the boundary between appearance-based shading models and physically-based shading models is blurry.

$$m = \sqrt{\frac{2}{c + 2}}. \tag{2.19}$$

$$D_{Phong}(\mathbf{h}, m) = \frac{1}{m^2 \pi} (\mathbf{n} \cdot \mathbf{h})^{\frac{2}{m^2} - 2}. \tag{2.20}$$

However, Walter et al. (2007) criticized even the physically-based Beckmann distribution for having a too short lobe. Burley (2012) confirmed this observation by re-evaluating different analytical models on the MERL dataset (Ngan et al. 2005). Speaking visually, this means that the Phong or Beckmann distribution can either approximate well the intensity of the specular highlight or the fading of the specular highlight, but not both at the same time. It is therefore not surprising that the combination of two lobes, one that approximates the highlight and one the falloff

(Ngan et al. 2005), leads to more accurate results. In order to overcome these limitations with a single lobe, Walter et al. (2007) suggested the GGX model.

$$D_{GGX}(\mathbf{h}, m) = \frac{m^2}{\pi((\mathbf{n} \cdot \mathbf{h})^2(m^2 - 1) + 1)^2} \cdot \qquad (2.21)$$

$$G_{GGX}(\mathbf{l}, \mathbf{v}, m) = G_1(\mathbf{l}, m)G_1(\mathbf{v}, m) \,,$$

$$G_1(\mathbf{v}, m) = \begin{cases} \frac{2(\mathbf{n} \cdot \mathbf{v})}{(\mathbf{n} \cdot \mathbf{v}) + \sqrt{m^2 + (1 - m^2)(\mathbf{n} \cdot \mathbf{v})^2}} \,, & \text{if } \frac{\mathbf{n} \cdot \mathbf{h}}{\mathbf{n} \cdot \mathbf{v}} > 0 \\ 0 \,, & \text{if } \frac{\mathbf{n} \cdot \mathbf{h}}{\mathbf{n} \cdot \mathbf{v}} < 0 \end{cases} \qquad (2.22)$$

Another advantage of the GGX model is that the limited set of parameters makes it very artist friendly. Variations of this model are used in the Unreal Engine (Karis 2013) and Disney's internal production renderer (Burley 2012, 2015).

A considerable amount of research exists on rendering translucent or semi-translucent materials using path tracing or photon mapping. Due to the number of publications, we focus only on work that is closely related to skin rendering. Subsurface scattering models the effect that incoming light is reflected at a different position. The relationship of the incoming and outgoing lights is described by a BSSDF, which is a generalization of the BRDF. While the BRDF was only dependent on the incoming and outgoing light angles, the BSSDF also considers the incoming and outgoing positions (Pharr & Humphreys 2010). One side-effect of subsurface scattering for skin and most other materials is that the light distribution becomes isotropic, due to the high amount of scattering. Nevertheless, many scattering events must be computed due to low absorption rates within the media, e.g., for red light $\sigma_a \in [0.013, 0.032]mm^{-1}$.

Jensen et al. (2001) suggested a dipole model for homogeneous materials as an efficient way to compute subsurface scattering. The dipole model has later been extended by Donner & Jensen (2005) for multi-layered translucent materials and adapted for human skin rendering (Donner & Jensen 2006). d'Eon et al. (2007) demonstrated that the dipole profile of skin can be closely approximated using a combination of Gaussians.

Due to the separability of Gaussians, d'Eon et al. (2007) approximated subsurface scattering using texture filtering in texture space. In order to increase performance, the texture filtering approach has been extended to screen space (Jimenez et al. 2009) and a single separable 2D filter (Jimenez et al. 2015). In addition, research focused on refining the reflectance model for skin (Weyrich et al. 2006, Donner et al. 2008, Jimenez et al. 2010, Iglesias-Guitian et al. 2015) or improve the diffusion theory that the dipole model is based on (d'Eon et al. 2011, Habel et al. 2013b,a). Finally, it should be noted that subsurface scattering is also used for rendering stylized characters (Burley 2012, 2015).

For skin rendering, Mental Ray has a dedicated multi-layer skin shader with subsurface scattering. Comparing this shader with different surface reflectance models is difficult, because it does not follow an algorithm published in academia, and the algorithm is not well documented. Based on the description and practical experiments, specular reflection is modeled using a two lobe approach, with every lobe modeled by a variation of the Phong specular model, which should approximate well the GGX model as we discussed previously.

We encoded different specularity of facial skin in specularity maps, following Weyrich et al. (2006). The shader simulates subsurface scattering by blurring and combining different albedo textures, which represent different skin layers. The top
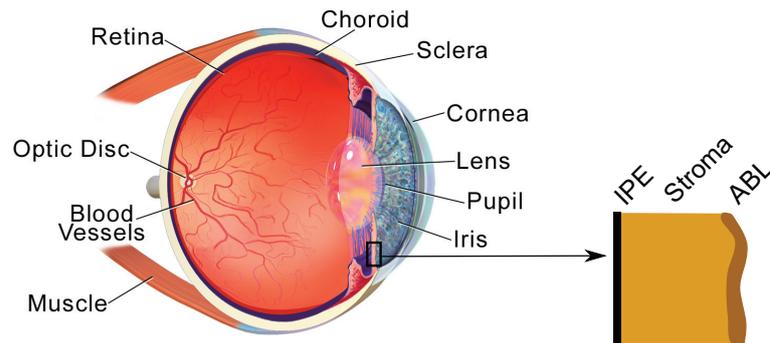
Figure 2.10: The anatomy of the human eye (Blausen 2014) with a cross-section (*right*) of the iris, showing the ABL (Anterior Border Layer), the Stromal layer and the IPE (Iris Pigment Epithelium). ©Eye: Blausen Medical

layer models the dermis and has a blueish albedo, while the lowest layer simulates the light distribution of the epidermis and has a red-tinted albedo texture. Overall, the skin shader can be considered similar to the method of d'Eon et al. (2007) without guaranteeing physical properties like energy conservation.

Eye Rendering   The human eye can be approximated by a sphere with a bulge (the cornea) in front of the iris (Figure 2.10). Most rendering work on eye rendering focuses actually on rendering the detailed structure of the iris. The iris is separated into different layers, the Anterior Border Layer (ABL), the stromal layer and the Iris Pigment Epithelium (IPE). Based on biological observations, the following assumptions are made for rendering the iris: The IPE absorbs any incoming light due to strong pigmentation. Depending on the amount of melanin in the ABL and stromal layer, the eye color ranges between brown (high melanin concentration) and grey or blue (low melanin concentration). As the ABL is much thinner than the stromal layer, scattering occurs only within the collagen fibrils of the stromal layer. The scattering occurs in Rayleigh fashion (Wilkerson et al. 1996) – short wavelengths (blue) are scattered more than long wavelengths (red). In consequence, white light will be colored grey/blue in cases of low melanin concentration, and brown for high melanin concentration.

By combining layers of different color textures, Lefohn et al. (2003) simulate the layered consistency of the iris. The idea is based on ocularist's approach to human iris synthesis. Lam & Baranoski (2006) developed a physically accurate eye rendering algorithm based on wavelength optics and the layered consistency of the iris. Due to the stochastic approach for scattering, the algorithm is based on Monte Carlo ray-tracing. By approximating the iris model of Lam & Baranoski (2006), François et al. (2009) developed a real-time version. In addition, they suggest a method that models the distortion of the iris due to refractions of the cornea. Jimenez et al. (2013) address practical aspects beyond the rendering of the iris, like eye-water around the eye, inter-reflections from eyelids, or the scattering on the sclera. Following Jimenez et al. (2013), we applied a transparent material with according index of reflection to the lens of the eye. For the iris, a subsurface scattering material was selected and the eye interior had a non-reflective black material.
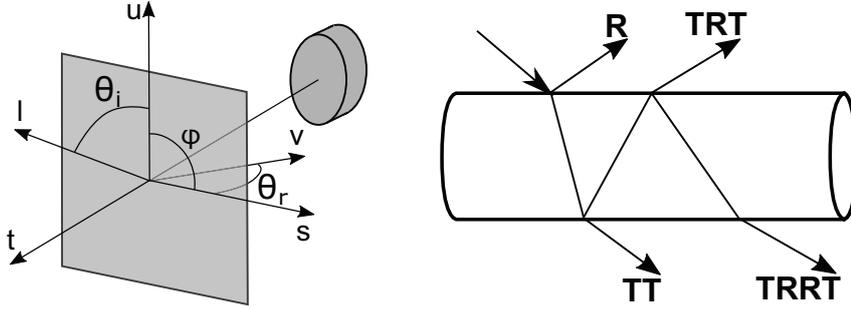
Figure 2.11: *Left:* Illustration of angles and vectors, required for the Kajiya-Kay and Marschner hair model. *Right:* Reflections (R) and transmittance (T) considered by the extended Marschner model in d'Eon et al. (2011).

Hair Rendering In contrast to skin or eyes, hair is represented as curves in computer graphics. Thus, the light interaction is not described via a BSSDF but via a Bidirectional Curve Scattering Distribution Function (BCSDF) (Zinke & Weber 2007). The most simple BCSDF is the Kajiya-Kay shading model (Kajiya & Kay 1989), that is still common in real-time applications. Similar to the Phong BRDF (Phong 1975), the Kajiya-Kay model consists of a specular and a diffuse component:

$$L(\mathbf{v}, \mathbf{t}, \mathbf{l}) = \underbrace{k_d \, \sin(\mathbf{t}, \mathbf{l})}_{\text{diffuse}} + \underbrace{k_s \, ((\mathbf{t} \cdot \mathbf{l})(\mathbf{t} \cdot \mathbf{v}) + \sin(\mathbf{t}, \mathbf{l}) \, \sin(\mathbf{t}, \mathbf{v}))^c}_{\text{specular}}, \qquad (2.23)$$

where k_d and k_s represent the diffuse and specular color tint and $\mathbf{t}$ is the tangent direction of a curve. Note that the model computes radiance directly, similarly to the Phong model. If one is only interested in the BCSDF, Equation (2.23) should be normalized by θ_i, which is the angle between the incoming light and the normal plane. The two main shortcomings of the model are that (i) translucency and (ii) a second tinted highlight cannot be modeled for hair. Variations of the Kajiya-Kay model exist (Scheuermann 2004, Goldman 1997) which address these limitations.

Based on physical principles, a more accurate BCSDF has been proposed by Marschner et al. (2003), and energy conservation has been improved by d'Eon et al. (2011). The model is based on two assumptions. First, most of the visible variation can be described by combining the first three to four modes of reflection R, TT, TRT and TRRT, where R and T stand for reflection and transmission at the hair fiber (Figure 2.11). The second assumption is that longitudinal and azimuthal scattering can be separately computed by the functions $M(\theta_i, \theta_r)$ and $N(\theta_i, \theta_r, \phi)$, with θ_i and θ_r being the longitudinal angles for the incoming and outgoing light and ϕ the azimuthal angle between the incoming and outgoing light. The separation in longitudinal and azimuthal scattering allows to summarize the BCSDF in a more generalized form (d'Eon et al. 2011), where $p \in \{R = 0, TT = 1, TRT = 2, \dots\}$ describes the light path within the fiber:

$$f_S(\theta_i, \theta_r, \phi) = \sum_{p=0}^{\infty} M_p(\theta_i, \theta_r), N_p(\theta_i, \theta_r, \phi). \qquad (2.24)$$

We omit further details on the computation of the longitudinal and azimuthal scattering functions and refer the interested reader to the original publications for

Figure 2.12: Final results obtained using our character creation pipeline and the improvements proposed by Achenbach et al. (2015).

full description and derivation. Although the evaluation of the Marschner model is computationally intensive, a real-time approximation is provided in Nguyen & Donnelly (2006) by saving precomputed results in texture look-up-tables or by simplifying some terms (Karis 2016).

On the one hand, the Marschner model simulates well most of the discriminative features of hair, on the other hand, input parameters like the index of refraction or absorption coefficients are not very artist friendly. Unintuitive input parameters in combination with the energy conservation of the Marschner model, where specular highlights are tightly coupled with translucency, make it very difficult to achieve a stylized look that typically goes beyond laws of physics. Therefore, Sadeghi et al. (2010) presented a method that combines the accuracy of the Marschner model paired with artist-friendly controls. Probably the best application case for their method is the hair in the movie *Tangled*.

Besides the complexity of the Marschner model itself, the approximation of the light scattering between different fibers is another challenge with a substantial impact on the visual results and performance. While Marschner et al. (2003) used path tracing, significant speed-up has been obtained by approximating global scattering with photon mapping (Moon & Marschner 2006) or spherical harmonics (Moon et al. 2008). Zinke et al. (2008) derive a formulation that saves forward and backscattering in spatial structures and is thus capable of rendering hair at up to interactive framerates on graphics hardware for directional light sources. In case of environmental lighting, performance can be increased by approximating the light with surface radial basis functions (Ren et al. 2010, Xu et al. 2011). For the creation of our stimuli, we used the previously mentioned plug-in Hairfarm that renders hair utilizing the method of Zinke et al. (2008). Figure 2.12 shows the final results obtained using our character creation pipeline with additional improvements for template fitting (Achenbach et al. 2015).

28

3 ElastiFace – Geometry and Texture Matching

Shape · Material · Base Character · 50% Eyes · 100% Nose

Figure 3.1: Our ELASTIFACE framework matches faces with strongly different shapes and textures. It facilitates combining shape and material of different characters (*left*) or morphing between facial parts (*right*). Examples are based on real perception studies (see Chapter 6, and Seyama & Nagayama (2007)).

Developing a method that establishes dense correspondences between two faces is driven by the idea to investigate the perceptual influence of static facial properties like shape or material independently. In computer graphics, most data varying over a polygonal surface is either saved per vertex or in a texture that is mapped to the surface. To exchange such information between various faces, dense one-to-one correspondences are required. The ideal method should be applicable across multiple face stylizations and not require topological changes in order to be suitable for building databases of face models incrementally. With ELASTIFACE we present a simple and effective framework for establishing correspondences between a given set of face models, using a combination of joint-fairing and non-rigid registration. In contrast to most previous work, our method handles input models of extremely different geometries (Figure 3.1). Nevertheless, its algorithmic core is based on solving simple bi-Laplacian linear systems and is therefore easy to implement.

Although there exists a wide variety of approaches for mapping one mesh onto another, most of them disqualify for our face morphing application due to their inherent limitations. Methods based on mapping the input models to a common simple parameter domain typically require the input models to be homeomorphic to a plane or a sphere. More general, cross-parametrization techniques have to adjust the mesh connectivity, thereby breaking the (required) one-to-one correspondence when matching to several target models. Non-rigid registration approaches avoid these topological limitations by deforming a given template model to different target shapes, but most of them are restricted to near-isometric input models. For the strongly varying face models we are interested in, their closest point correspondences fail to give valid results.

By combining the concepts of (i) deformation-based registration and (ii) transformation of models into a simpler domain, we overcome the individual limitations of these approaches. A novel simultaneous fairing technique transforms source and target meshes into a simple, featureless, and geometrically very similar state, from which accurate correspondences for a non-rigid registration of the original meshes

can be robustly determined. This extension to closest point correspondences is the key contribution of our work. Our fairing and fitting techniques are both simple to implement and computationally efficient since both are based on the minimization of a geometrically intuitive quadratic energy. In addition to matching the shape of the given input models, we further extend our framework by blending between facial parts, input textures and several rendering styles, as shown in Figure 3.1. Besides our specific use-case, blendshape based facial rig creation can highly benefit from our method as we demonstrate on the clay-figure example (Figure 3.10).

## 3.1 Overview of Dense Correspondence Estimation Methods

Establishing a mapping from one model to another has been investigated in several fields and under different names, and is therefore referred to as mesh morphing, cross-parametrization, non-rigid registration, or correspondence estimation. We only discuss the most relevant methods here and refer the reader to the book of Bronstein et al. (2008), the survey of van Kaick et al. (2011), or the courses of Chang et al. (2010) and Bouaziz et al. (2014) for more details.

An overview of early mesh morphing methods is reported by Alexa (2002). The described methods typically parametrize the input models into a simple domain, such as a disk or a sphere. In case of the former, a constrained planar parametrization (Lévy 2001, Kraevoy et al. 2003) can then be used to establish correspondences. However, in both cases, the input models are restricted to topological planes or spheres. Similar topological restrictions apply to the approaches of Blanz & Vetter (1999) or of Wang et al. (2008), who fit a model to a scan via cylindrical or disk parametrization.

In contrast, inter-surface mapping approaches (Kraevoy & Sheffer 2004, Schreiner et al. 2004) avoid the common parametrization domain by constructing a direct mapping between the two 3D models. Although these methods can match arbitrary non-isometric objects, they have to insert additional vertices to the resulting mesh, which breaks the requirements of our application of incrementally building a face database. Bronstein et al. (2006, 2008) also compute a direct mapping between two surfaces. They minimize parametric distortion through Generalized Multidimensional Scaling (GMDS). While their method does not introduce new vertices, it is designed particularly for isometric or close-to-isometric models — a prerequisite not met by our strongly varying face models.

Other approaches embed the input models into spaces where correspondences are easier to detect. For instance, Ovsjanikov et al. (2010) match objects based on a single correspondence by embedding the models using the Heat Kernel Map. Lipman & Funkhouser (2009) map models to the complex plane using Möbius transformations and find correspondences by a voting scheme that evaluates groups of three-point candidates. Again, both methods are designed for near-isometric input models. Kim et al. (2011) overcome this restriction by blending several intrinsic maps. However, their method yields to non-plausible mappings for our examples (see Section 3.5).

Non-rigid registration approaches overcome the limitation of certain topological types by deforming a template model until it matches the given target model. The approach of Lee & Magnenat-Thalmann (2000) employs free-form deformation with
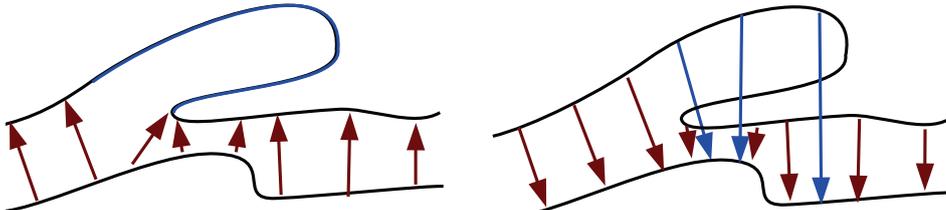
Figure 3.2: Difficult case for ICP-based non-rigid registration. *Left:* The lower surface mainly aligns to the black part of the surface and discards the details marked blue. *Right:* The blue correspondences cause self-intersections when aligning the upper surface to the lower one.

---

manually specified curve constraints to fit a generic head model to photographs. Similarly, Bui et al. (2003) use a Radial Basis Function (RBF) space warp with feature points determined by a genetic algorithm. Both approaches have the drawback that their low degree of freedom space warps are not capable of mapping strongly differing geometries, e.g., to unfold a "source ear" to a "target non-ear" (see Figure 3.3). Similar restrictions apply to the correspondence estimation of Noh & Neumann (2001), which employs an RBF warp followed by a cylindrical projection.

More recent non-rigid registration approaches fit several scans of deforming objects (Li et al. 2008, Huang et al. 2008, Tevs et al. 2009) or fit a template model to 3D scans (Allen et al. 2003, Amberg et al. 2007, Weise et al. 2009, 2011). Rather than registering a surface to another, Sahillioğlu & Kavan (2016) suggest unfolding 3D shapes into poses that are invariant to nonrigid transformations. But most of these methods assume near-isometric deformations. For models of very different shape, e.g., one model has ears while the other one does not, these methods fail to find valid correspondences (Figure 3.2), which in turn leads to fold-overs and inter-penetrations in the mapped model, as we demonstrate in Section 3.5.

Our method can be considered as a combination of the ideas presented above: Similar to non-rigid registration, it deforms a source/template model into given target shapes. But it avoids the problem of invalid closest point correspondences by first mapping source and target models into a simpler space and computing correspondences there. The simpler space, however, is not a planar or spherical parametrization, but a smoothed, feature-less version of the input models, computed by the joint-fairing technique proposed in the next section.

Generalized barycentric coordinates on a surface (Rustamov 2010) are as well suitable for establishing dense correspondences between meshes. However, relying on geodesics, the computation time quickly increases with the number of vertices. Panozzo et al. (2013) overcome this limitation by replacing geodesics with Phong projections in a higher dimensional space. The method is also suitable for establishing dense correspondences between entire meshes. While a single correspondence for a vertex is computed in a few microseconds, pre-computation times of several minutes are required in advance. Similar to our approach, their method involves least-square meshes and is thus also suffering of thinning artifacts.

Conceptually similar to our joint-fairing is the joint-flattening in lifted bijections (Aigerman et al. 2014), published after our work. First, points are connected using minimal geodesic distances. Afterwards the meshes are cut along the graph and

jointly parametrized in 2D. Possible ambiguities at overlapping areas are resolved by constructing paths to the bijective mappings at the cut. However, inconsistencies may appear at the cuts which is addressed in follow-up work (Aigerman et al. 2015). In contrast to Panozzo et al. (2013) and our approach, Aigerman et al. (2015) overcome the thinning artifacts, at the cost of high computation times (20–30 minutes) and numerical instability for meshes of over 80K vertices. In order to maintain mesh connectivity Aigerman et al. (2015) cut meshes only along existing edges, implying that reference points must be placed at vertices. It remains open whether the method allows only a limited amount of references points as otherwise cuts along existing vertices cannot be constructed. Both limitations might be a serious issue in practice for low resolution, textured meshes. For the stylized characters from Chapter 5 more than 40 reference points were required for establishing dense correspondences that were also semantically correct. As some facial parts grow (eyes) while other parts shrink (nose) such properties cannot be considered by optimizing only conformal distortions.
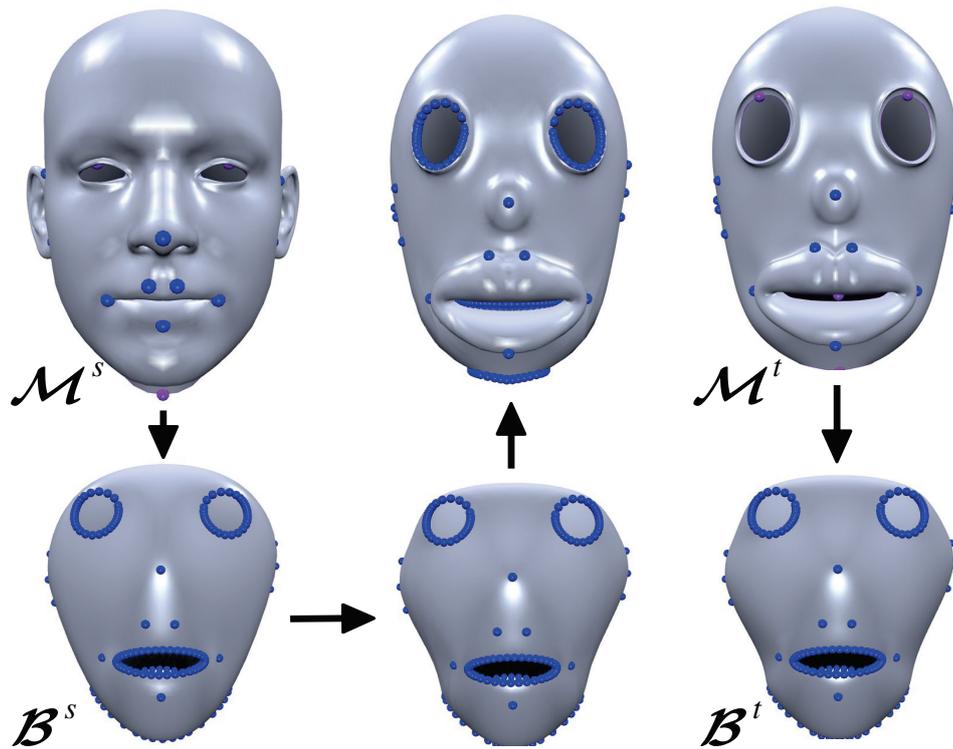
3.2 Geometry Matching

The first step of our face matching is to geometrically map the source face mesh $\mathcal{M}^{\text{s}}$ onto the target face model $\mathcal{M}^{\text{t}}$. To this end we adjust the source model's vertex positions only, and keep its connectivity fixed. After the user has manually marked a few correspondences (Section 3.2.1), we first transform both the source and the target model $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{t}}$ into smoothed base meshes $\mathcal{B}^{\text{s}}$ and $\mathcal{B}^{\text{t}}$ (Section 3.2.2), which afterwards are brought into correspondence by a non-rigid registration approach (Section 3.2.3). The correspondences derived from the smoothed models $\mathcal{B}^{\text{s}}$ and $\mathcal{B}^{\text{t}}$ are then used as initial guess for the registration of the original models $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{t}}$. This process is depicted in Figure 3.3 and explained below. In the following, we denote the vertices of the source and target mesh by $\mathbf{v}_i^{\text{s}}$, $i = 1, \ldots, N^{\text{s}}$, and $\mathbf{v}_j^{\text{t}}$, $j = 1, \ldots, N^{\text{t}}$, respectively. If we particularly emphasize properties of the original undeformed meshes we denote this by a hat (ˆ). Vertex positions of the smoothed base meshes $\mathcal{B}^{\text{s}}$ and $\mathcal{B}^{\text{t}}$ are denoted as $\mathbf{v}_i^{bs}$ and $\mathbf{v}_j^{bt}$ respectively.

3.2.1 Manual Correspondence Specification

As in most morphing or cross-parametrization approaches, the user initially specifies a few correspondences by selecting the respective feature points on both the source model $\mathcal{M}^{\text{s}}$ and the target model $\mathcal{M}^{\text{t}}$. In our application, the user typically marks about 15–20 correspondences for nose, eyes, mouth, and ears, as shown in Figure 3.3, left column. In contrast to most other approaches, however, a simple vertex-to-vertex correspondence ($\mathbf{v}_i^{\text{s}} \mapsto \mathbf{v}_j^{\text{t}}$) is often not accurate enough in our context, since our face models are equipped with high-resolution textures. For instance, accurately selecting correspondences for the eyebrows in Figure 3.4 requires to specify reference points within triangles. Each of these reference points is represented by a barycentric combination of its triangle vertices. Our set of correspondence constraints therefore consists of tuples of reference points ($\mathbf{r}_m^{\text{s}}, \mathbf{r}_m^{\text{t}}$), $m = 1, \ldots, M$, represented as:

$$\alpha_m^{\text{s}}\mathbf{v}_\alpha^{\text{s}} + \beta_m^{\text{s}}\mathbf{v}_\beta^{\text{s}} + \gamma_m^{\text{s}}\mathbf{v}_\gamma^{\text{s}} \quad \text{and} \quad \alpha_m^{\text{t}}\mathbf{v}_\alpha^{\text{t}} + \beta_m^{\text{t}}\mathbf{v}_\beta^{\text{t}} + \gamma_m^{\text{t}}\mathbf{v}_\gamma^{\text{t}} . \qquad (3.1)$$

In the above equation, $\alpha_m, \beta_m, \gamma_m$ denote the barycentric coordinates of $\mathbf{r}_m$ with respect to the containing triangles on $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{t}}$.

Figure 3.3: Overview of the geometric matching: After selecting correspondence constraints (*blue and pink dots*), the source mesh $\mathcal{M}^s$ and target mesh $\mathcal{M}^t$ are smoothed into base meshes $\mathcal{B}^s$ and $\mathcal{B}^t$. Then $\mathcal{B}^s$ is fitted to $\mathcal{B}^t$ (*bottom center*), and – based on the resulting correspondences – $\mathcal{M}^s$ is fitted to $\mathcal{M}^t$ (*top center*).
©Loki: Mark Pauly
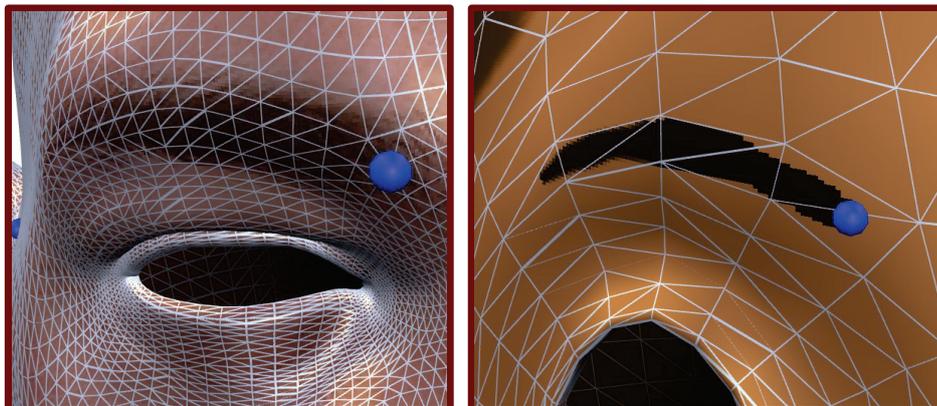
Figure 3.4: In order to accurately specify corresponding points at the eyebrow, we have to use interior triangle points instead of simple vertex-to-vertex correspondences. ©Viktor: Faceware Technologies
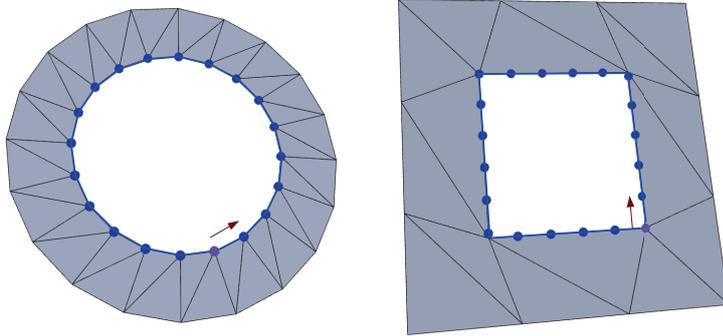
Figure 3.5: Starting from one or more correspondence constraints on a boundary loop (*purple*) we automatically assign correspondences for all other boundary loop vertices (*blue*) based on relative distance along the boundary loop.

In order to accurately map boundaries (e.g., eyes, mouth, neck), the user can manually specify one or more points on corresponding boundary loops. The system then automatically determines target positions for all other vertices on the source boundary loop by preserving the relative edge lengths on the target boundary loop (see Figure 3.5). Note that the target positions will lie on boundary edges, and can therefore also be represented as a barycentric combination of boundary vertices in the form of Equation (3.1).

Based on these correspondences we initialize the registration by aligning the two models using the best-matching similarity transform between source and target. This amounts to minimizing the squared distances of transformed source points to their corresponding target points:

$$\min_{\mathbf{R},\,\mathbf{t},\,s} \sum_{m=1}^{M} \left\| s\,\mathbf{R}\,\mathbf{r}_m^{\mathrm{s}} + \mathbf{t} \, - \, \mathbf{r}_m^{\mathrm{t}} \right\|^2 .$$

The optimal rotation $\mathbf{R}$, translation $\mathbf{t}$, and uniform scaling s can be computed in closed form (Horn 1987, Umeyama 1991).

3.2.2 Joint-Fairing

By obtaining the reference points $(\mathbf{r}_m^{\mathrm{s}}, \mathbf{r}_m^{\mathrm{t}})$ for models with similar proportions and a similar level of geometric detail, it would be possible to establish a mapping between $\mathcal{M}^{\mathrm{s}}$ and $\mathcal{M}^{\mathrm{t}}$ using a deformation-based registration with ICP-like closest point constraints, such as the method of Weise et al. (2009). However, if the models differ significantly, e.g., one model has ears while the other one does not, the closest point constraints fail to give reasonable results (see Figure 3.2). This typically leads to fold-overs and self-intersections in the deformed source mesh. To overcome the limitations of closest point correspondences in 3D, other approaches first transform both models into a simpler space and find correspondences there. For instance, Lipman & Funkhouser (2009) map models into the complex plane using a Möbius transform, which relaxes isometry to conformal equivalence. However, these methods typically require both models to differ only by a near-isometry, which is not the case for our application.

In contrast, we propose to transform both models into a simple, similar, and feature-less shape, on which we then compute robust correspondences. This is conceptually similar to Kraevoy & Sheffer (2004) and Schreiner et al. (2004), but avoids intermediate steps like path constructions between feature points. The final shapes will be non-isometric such that the initial source and target models can differ significantly. The main idea is to (i) apply aggressive fairing to remove geometric details, (ii) force corresponding points to coincide to achieve a sufficient geometric similarity, and (iii) allow correspondences to move to a certain degree in order to unfold geometrically complex regions like mouth and nose (see Figure 3.3, bottom).

A fair surface is a smooth surface of minimal area spanned across a fix boundary. For our application, we are interested in minimizing the mean curvature $H = (\kappa_1 + \kappa_2)/2$ over the entire mesh, with κ_1 being the maximum curvature and κ_2 the minimum curvature. For a continuous surface a fair surface is obtained by minimizing the following energy:

$$E_{\text{Fair}} = \int H^2 dA \,. \tag{3.2}$$

In case of discrete triangle meshes, the mean curvature vector is directly related to the Laplace-Beltrami operator, which in turn is defined through the well known cotangent formula (Pinkall & Polthier 1993, Desbrun et al. 1999, Meyer et al. 2003).

$$H^2(\mathbf{v}_i) = \frac{1}{2} \|\Delta \mathbf{v}_i\|^2 \,, \tag{3.3}$$

$$\Delta \mathbf{v}_i = \frac{1}{2A_i} \sum_{\mathbf{v}_j \in \mathcal{N}_i} (\cot \alpha_{i,j} + \cot \beta_{i,j})(\mathbf{v}_j - \mathbf{v}_i) \,. \tag{3.4}$$

$\mathbf{n}_i$ and A_i indicate the normal and the Voronoi area of vertex $\mathbf{v}_i$. α and β are the angles opposite to the edge $\overline{\mathbf{v}_i \mathbf{v}_j}$ and $\mathcal{N}_i$ denotes the one-ring neighborhood of vertex $\mathbf{v}_i$. Thus instead of minimizing the squared mean curvature over the entire surface, we can also minimize the squared norm of the Laplacian.

$$\min_{\mathbf{v}_i} \sum_{i=1}^{N} \frac{1}{2} A_i \|\Delta \mathbf{v}_i\|^2 \,. \tag{3.5}$$

The solution is found by solving the linear system.

$$\mathbf{L}^2 \begin{bmatrix} \mathbf{v}_1^\mathsf{T} \\ \vdots \\ \mathbf{v}_N^\mathsf{T} \end{bmatrix} = \mathbf{0} \,. \tag{3.6}$$

The linear system is separable in the $x/y/z$ coordinates of $\mathbf{v}_i$ leading to a $N \times N$ Laplace matrix $\mathbf{L}$ and a $N \times 3$ zero matrix $\mathbf{0}$. However, solving this linear system without boundary constraints leads to the trivial solution $[\mathbf{v}_1, \ldots, \mathbf{v}_N]^\mathsf{T} = \mathbf{0}$. We thus extend our fairing energy by adding reference constraints. In addition, we fair both input meshes $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{t}}$ simultaneously, obtaining our final energy function:

$$E_{\text{Fair}}\left(\mathbf{v}_1^{\text{s}}, \ldots, \mathbf{v}_{N^{\text{s}}}^{\text{s}}, \mathbf{v}_1^{\text{t}}, \ldots, \mathbf{v}_{N^{\text{t}}}^{\text{t}}\right) \;= \tag{3.7}$$

$$\frac{\lambda_1}{\sum_i A_i^{\text{s}} + \sum_j A_j^{\text{t}}} \left(\sum_{i=1}^{N^{\text{s}}} A_i^{\text{s}} \left\|\Delta\mathbf{v}_i^{\text{s}}\right\|^2 \;+\; \sum_{j=1}^{N^{\text{t}}} A_j^{\text{t}} \left\|\Delta\mathbf{v}_j^{\text{t}}\right\|^2 \right) \tag{3.8}$$

$$+\; \frac{\lambda_2}{M} \sum_{m=1}^{M} \left\|\mathbf{r}_m^{\text{s}} - \mathbf{r}_m^{\text{t}}\right\|^2 \tag{3.9}$$

$$+\; \frac{\lambda_3}{M} \sum_{m=1}^{M} \left\|\frac{1}{2}(\mathbf{r}_m^{\text{s}} + \mathbf{r}_m^{\text{t}}) - \frac{1}{2}(\hat{\mathbf{r}}_m^{\text{s}} + \hat{\mathbf{r}}_m^{\text{t}})\right\|^2 . \tag{3.10}$$

The first term (3.8) penalizes the squared norms of per-vertex Laplacians $\Delta\mathbf{v}_i$ and leads to as smooth as possible surfaces. The second term (3.9) penalizes the deviation of corresponding reference points, hence is responsible for making them coincident. Note that $\mathbf{r}_m^{\text{s}}$ and $\mathbf{r}_m^{\text{t}}$ are barycentric combinations of vertices $\mathbf{v}_i^{\text{s}}$ and $\mathbf{v}_j^{\text{t}}$ (3.1), such that this objective can be formulated in terms of the latter. The last term (3.10) is required to avoid the trivial solution. It basically prescribes a target position for the two corresponding points $\mathbf{r}_m^{\text{s}}$ and $\mathbf{r}_m^{\text{t}}$, which can be considered to be coincident due to the second term. The target position is chosen as the point where $\mathbf{r}_m^{\text{s}}$ and $\mathbf{r}_m^{\text{t}}$ can meet with least movement, which is the average $\frac{1}{2}(\hat{\mathbf{r}}_m^{\text{s}} + \hat{\mathbf{r}}_m^{\text{t}})$ of their original positions (before the optimization).

Note that we do not combine the terms (3.9) and (3.10) since we want to enforce strongly that corresponding points become coincident, while we only enforce weakly a specific target position. This allows the reference points to move to a certain degree in order to further decrease the curvature term, which effectively leads to an unfolding of geometrically difficult parts, such as nose, mouth, and ears (Figure 3.3). If we strictly enforce $\mathbf{r}_m^{\text{s}} = \mathbf{r}_m^{\text{t}} = \frac{1}{2}(\hat{\mathbf{r}}_m^{\text{s}} + \hat{\mathbf{r}}_m^{\text{t}})$ the surface would not be able to unfold to a state without self-intersections. This behavior was achieved by $\lambda_1 = 0.1$, $\lambda_2 = 100$, $\lambda_3 = 1$ in almost all examples; only the Slimer model (Figure 3.12) and the clay faces (Figure 3.10) require a higher smoothing weight ($\lambda_1 = 10$) to fully unfold.

If we keep the cotangent weights and Voronoi areas fixed, then minimizing the quadratic objective function (3.7) amounts to solving three $(N^{\text{s}} + N^{\text{t}}) \times (N^{\text{s}} + N^{\text{t}})$ linear systems of normal equations for the x, y, and z coordinates of the vertex positions of $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{t}}$. This system is highly sparse, symmetric, and positive definite and we solve it using a sparse Cholesky factorization (Chen et al. 2008).

The results of this energy minimization noticeably depend on the underlying triangulation, since the cotangent weights and Voronoi areas, in fact, depend non-linearly on the vertex positions. Hence, the simple linear solve might not be a sufficiently accurate approximation to the true nonlinear solution. To take this mesh-dependence into account, we iteratively update the weights and re-solve the linear system. To avoid numerical problems due to degenerated cotangent weights, we follow Kazhdan et al. (2012) and update the Voronoi areas only, while keeping the cotangent weights fixed. Although this process might not converge in a theoretical sense, in practice, there are no noticeable changes after five to ten iterations. The joint-fairing approach turned out to be numerically very robust, and it yields two highly smooth and geometrically very similar base meshes $\mathcal{B}^{\text{s}}$ and $\mathcal{B}^{\text{t}}$. This is even the case for highly different geometries and tessellations, as shown in the second column of Figure 3.3.

3.2.3 Non-Rigid Registration

The two base meshes resulting from the joint-fairing process are void of any geometric details due to smoothing and are geometrically very similar since they correspond to discrete curvature-minimizing thin plate surfaces with identical Dirichlet boundary constraints. As such, they are an easy task for a deformation-based non-rigid registration approach. We therefore first deform the smooth source mesh $\mathcal{B}^{\mathrm{s}}$ onto the smoothed target model $\mathcal{B}^{\mathrm{t}}$, and then use their resulting vertex correspondences as an initial guess for the registration of the original models $\mathcal{M}^{\mathrm{s}}$ and $\mathcal{M}^{\mathrm{t}}$.

For matching $\mathcal{B}^{\mathrm{s}}$ to $\mathcal{B}^{\mathrm{t}}$, we adjust the vertex positions of the former. To this end, we iteratively minimize an energy consisting of a fitting term and a smoothness term, as it is done by most non-rigid registration approaches. The smoothness term (3.11) penalizes bending (i.e., change of curvature) of the source model, measured by the Laplacians of vertex displacements $\Delta(\mathbf{v}_i^{bs} - \hat{\mathbf{v}}_i^{bs})$, where $\hat{\mathbf{v}}_i^{bs}$ and $\mathbf{v}_i^{bs}$ denotes the vertex position on mesh $\mathcal{B}^{\mathrm{s}}$ before and after the deformation. The fitting term (3.12) tries to minimize the distance of each source vertex $\mathbf{v}_j^{bs}$ to its closest point $\mathbf{c}_j^{bt}$ on the smoothed target model $\mathcal{B}^{\mathrm{t}}$, unless pruning heuristics apply (Section 2.2). The third term (3.13) ensures that the coincident reference points $\mathbf{r}_m^b$ remain at their position:

$$E_{\mathrm{Regist}}\left(\mathbf{v}_1^{bs}, \ldots, \mathbf{v}_{N^{\mathrm{s}}}^{bs}\right) = \frac{\mu_1}{\sum_i A_i^{bs}} \sum_{i=1}^{N^{\mathrm{s}}} A_i^{bs} \left\| \Delta\left(\mathbf{v}_i^{bs} - \hat{\mathbf{v}}_i^{bs}\right) \right\|^2 \qquad (3.11)$$

$$+ \frac{\mu_2}{C} \sum_{j=1}^{C} \left\| \mathbf{v}_j^{bs} - \mathbf{c}_j^{bt} \right\|^2 \qquad (3.12)$$

$$+ \frac{\mu_3}{M} \sum_{m=1}^{M} \left\| \mathbf{r}_m^{bs} - \hat{\mathbf{r}}_m^{bt} \right\|^2 . \qquad (3.13)$$

While we use the Voronoi area A_i^{bs} of the smoothed base mesh $\mathcal{B}^{\mathrm{s}}$, we keep the cotangent weights that have been computed on the initial mesh $\mathcal{M}^{\mathrm{s}}$. Note that in contrast to most other non-rigid registration approaches we can use a simple quadratic energy, since both meshes are already very similar, such that the closest point constraints $(\mathbf{v}_j^{bs}, \mathbf{c}_j^{bt})$ are meaningful, equal by amount ($C \approx N$) and a linear deformation model is sufficient. The energy is again minimized by solving a sparse linear system, similar to the one of Section 3.2.2.

In an ICP-like manner (Besl & McKay 1992) we iteratively update closest point correspondences $(\mathbf{v}_j^{bs}, \mathbf{c}_j^{bt})$ and re-solve the linear system, which typically converges after three to four iterations. Similar to other methods, we start with a rather stiff surface ($\mu_1 = 10$), which is then made softer ($\mu_1 = 1$ and $\mu_1 = 0.1$) in order to allow for a more precise fit. The other weights are set to $\mu_2 = 1$ and $\mu_3 = 10$. In total, the registration takes about ten iterations and accurately maps the smoothed source model $\mathcal{B}^{\mathrm{s}}$ onto the smoothed target model $\mathcal{B}^{\mathrm{t}}$ (see Figure 3.3, bottom center).

As the last step of our registration pipeline, we take the final closest point correspondences $(\mathbf{v}_j^{bs}, \mathbf{c}_j^{bt})$ computed on the smooth meshes, and use them for matching the original source model $\mathcal{M}^{\mathrm{s}}$ to the original target model $\mathcal{M}^{\mathrm{t}}$. To this end, we simply replace $\mathbf{c}_j^{bt}$, which is represented by triangle index and barycentric coordinates on $\mathcal{B}^{\mathrm{t}}$, by the equivalent point $\mathbf{c}_j^{\mathrm{t}}$ on the original target model $\mathcal{M}^{\mathrm{t}}$ in the registration energy (3.12). Similarly, we also replace all reference points and weights of the
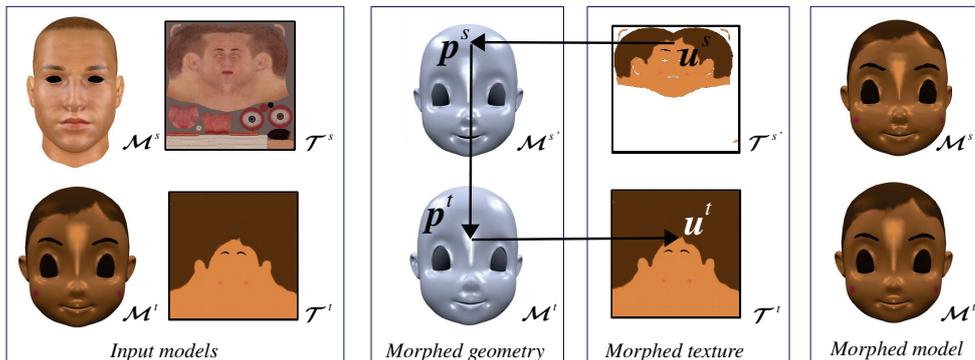
Figure 3.6: Overview of the texture transfer algorithm. *From left to right:* Original, textured meshes as input models, morphed and target geometry, morphed and target texture, and the final morphed model. ©Dave: TurboSquid

smoothed meshes $(A_i^{bs}, \mathbf{v}_i^{bs}, \mathbf{r}_m^{bt})$ by their counterparts of the original meshes and set the fitting weight very high (typically 300), while keeping the other weights at 0.1. Since the correspondences computed on the smoothed meshes are very good, we only need one or two iterations on the original models to achieve all shown results (Section 3.5).

3.3 Texture Matching

Once the source mesh $\mathcal{M}^s$ has been deformed to match the target mesh $\mathcal{M}^t$ geometrically, the source texture has to be adjusted, such that the textured versions of both meshes look identical as well. In the following, we denote by $\mathcal{T}^s$ and $\mathcal{T}^t$ the texture images of the source and target mesh, and by $\mathbf{u}_i^s$ and $\mathbf{u}_j^t$ their texture coordinates or uv-coordinates. The planar triangle mesh with uv-coordinates assigned as vertex positions is referred to as the uv-layout. Moreover, we now denote by source mesh $\mathcal{M}^s$ the deformed source mesh after the geometric matching unless stated otherwise.

Note that there are two options for mapping the target texture onto the source mesh. We can either adjust the texture coordinates of the source mesh to properly access the target texture, or we can transform the target texture to match the uv-layout of the source mesh. However, since we later want to morph the geometry and appearance of several models by blending either their vertex positions or texture images, all meshes must have the same mesh connectivity and uv-layout. Consequently, we cannot adjust the texture coordinates $\mathbf{u}_i^s$, but instead have to replace the source texture $\mathcal{T}^s$ by a transformed version of the target texture $\mathcal{T}^t$.

Another issue is that the uv-layouts of source and target might have incompatible seams or even consist of a different number of connected components. As a consequence, a smooth (or even continuous) 2D warp $\mathbf{f} \colon \mathcal{T}^t \to \mathcal{T}^s$ between the source and the target uv-layouts does not exist in general. Therefore, we perform the inverse transformation $\mathbf{f}^{-1}$ in a pixel-by-pixel manner: For each pixel $\mathbf{u}^s \in \mathcal{T}^s$ we find its pre-image $\mathbf{u}^t = \mathbf{f}^{-1}(\mathbf{u}^s) \in \mathcal{T}^t$ and copy its color value to the source texture.

As illustrated in Figure 3.6, this per-pixel mapping is computed through 3D closest point correspondence of $\mathcal{M}^s$ and $\mathcal{M}^t$. For each pixel $\mathbf{u}^s \in \mathcal{T}^s$ we first find

the 2D triangle covering it in the texture layout. Using barycentric coordinates with respect to this triangle, the pixel $\mathbf{u}^{\mathrm{s}}$ can be mapped to its corresponding 3D point $\mathbf{p}^{\mathrm{s}}$ on the source mesh. A closest point query using binary space partitioning trees reveals the corresponding point $\mathbf{p}^{\mathrm{t}}$ on the target mesh, which is finally mapped to $\mathbf{u}^{\mathrm{t}}$ using its (interpolated) texture coordinates. Because $\mathbf{u}^{\mathrm{t}}$ is a non-integer texture coordinate in general, its color value is obtained using bilinear texture interpolation and assigned to the pixel $\mathbf{u}^{\mathrm{s}}$. Although this interpolation inevitably leads to a slight texture blurring, it did not create noticeable artifacts in our examples.

This algorithm successfully transfers colors of the target texture to all pixels covered by a uv-triangle in the source texture, and it leaves blank all pixels not covered by a uv-triangle. However, round-off errors during the mapping might cause artifacts at texture boundaries, where a few pixels might be missing. We address this issue by performing a simple one-pixel dilation, i.e., by filling all transparent pixels with the average of their opaque neighbors' colors, which effectively eliminates this problem. Chapter 6 and Section A.2 show over 200 stimuli with textures transferred between characters of different stylization level.
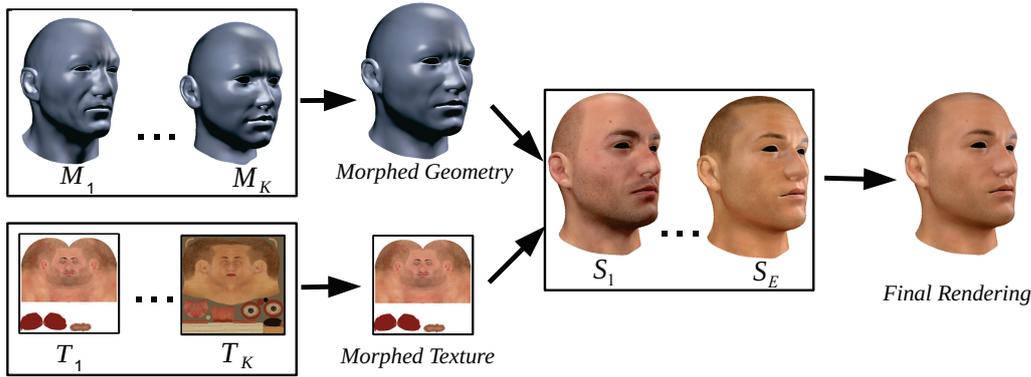
3.4   Face Morphing

The geometry and texture matching introduced in the previous sections allows us to map a source model to one or more target models. This precomputation results in a set of meshes with identical mesh connectivity and uv-layout (that of the source model), but with different shapes and texture images (that of the target models). Due to their identical connectivity and texture layout, these models are all in one-to-one vertex and pixel correspondence, which enables us to easily morph their geometries and appearances by blending their vertex positions and texture images (Figure 3.7, left). In the following, we describe a prototype for morphing faces in real-time. Such a system can be the foundation for interactive facial perception experiments that adopt stimuli during a trial.

For the blending between entire faces, we employ simple linear interpolation of vertex positions, which we prefer over more sophisticated techniques due to its simplicity, efficiency, and the fact that for faces there are almost no visible differences between linear and nonlinear geometry interpolation. In order to blend only parts of the input models (as shown in Figure 3.1) we closely follow Alexa (2003): Instead of vertex positions, we interpolate per-vertex Laplacians using Spherical linear interpolations (Slerp)(Shoemake 1985) and solve a Poisson system for the desired vertex positions. The only difference is that instead of the uniform graph Laplacian used by Alexa, we employ the cotangent discretization (Pinkall & Polthier 1993, Meyer et al. 2003), which avoids distortion in the case of irregular meshes.

For full texture blending we use simple linear interpolation, but for local blending, a gradient-based technique should be employed (Pérez et al. 2003). Additionally, we incorporate several real-time rendering styles into our face morphing application, which model a wide range of effects, ranging from realistic skin (d'Eon et al. 2007), over illustrative game characters (Mitchell et al. 2007), to non-photorealistic shading (Barla et al. 2006), as shown in Figure 3.8.

In our face morphing application, the user can interactively adjust the blending weights, and the face rendering is adjusted in real time. Our pipeline for blending geometry, texture, and rendering style is illustrated in Figure 3.7. After interpolating vertex positions, normal vectors, and texture images, the resulting morphed model

Figure 3.7: Overview of the face morphing algorithm: In a first step, the face meshes $(\mathcal{M}_1, \ldots, \mathcal{M}_L)$ and textures $(\mathcal{T}_1, \ldots, \mathcal{T}_L)$ are interpolated. The resulting interpolated mesh is rendered with the interpolated texture into several off-screen buffers using different rendering styles $(\mathcal{S}_1, \ldots, \mathcal{S}_E)$. At the end, the final rendering is obtained by blending between the off-screen buffers.
©Dave: TurboSquid, ©Viktor: Faceware Technologies

Figure 3.8: Different models and rendering styles (*odd columns*) and 50%-blends between them (*even columns*). ©Dave: TurboSquid, ©Loki: Mark Pauly, ©Viktor: Faceware Technologies

is rendered into a set of off-screen buffers using all active rendering styles, and the final image is a simple linear interpolation of these off-screen buffers. Since most computations are performed on the GPU using a combination of OpenCL and GLSL, the morphing can be performed in real-time even for complex models.

3.5 Evaluation of ElastiFace

For evaluating the capabilities of our face matching framework, we first present qualitative registration and morphing results for several characters, then provide a quantitative analysis of fitting accuracy and computational time, and finally discuss and compare to related work.

Figure 3.9 demonstrates that our system is capable of robustly mapping a source face model to a set of target meshes ranging from realistic to highly abstract characters. These results have been obtained by five iterations of joint-fairing and five iterations of the geometric registration. Besides accurately mapping the face geometries, it also successfully transfers the textures from target to source. After mapping the source model to these four target models, we can blend between the shapes and
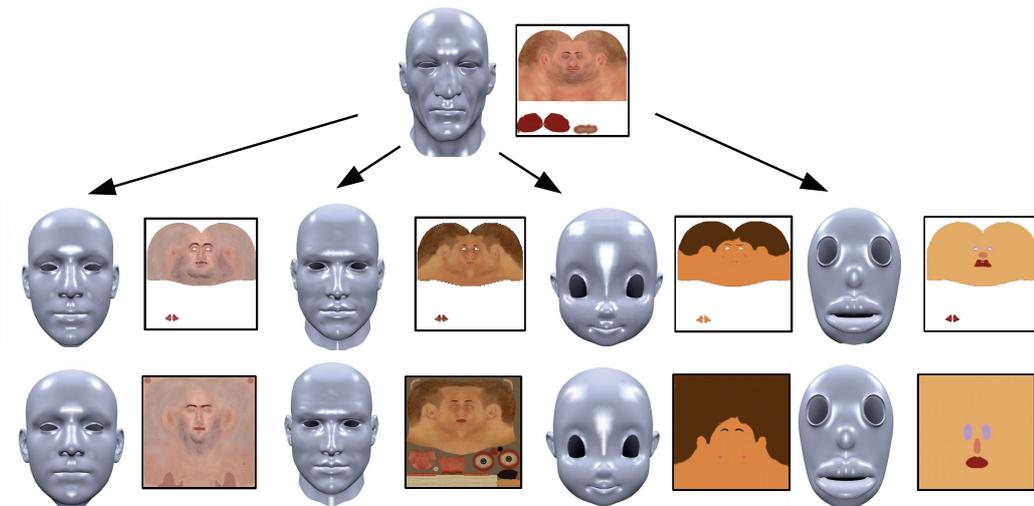
Figure 3.9: Mapping geometry and texture of the source model Viktor (*top*) to four other faces, ranging from realistic to highly abstract (*from left to right:* Loki, Dave, Girl, Kissmouth). The target models are shown in the bottom row, the morphed source models in the middle row.

©Dave: TurboSquid, ©Loki: Mark Pauly

Fitting	N^{s}	N^{t}	Time	Error
Viktor to Dave	19k	10k	7s	0.5%
Viktor to Loki	19k	6.7k	5s	0.4%
Viktor to Girl	19k	2.8k	5s	0.3%
Viktor to Kissmouth	19k	7.8k	7s	0.3%

Table 3.1: Statistics for different mapping examples, listing number of source and target vertices N^{s} and N^{t}, total time for fitting, and relative Hausdorff distance.

appearances of the deformed source models, which now have identical mesh connectivity and texture layouts. Figure 3.8 shows the results of 50%-morphs between these four characters.

In order to push our system to its limits, we sculpted several faces in clay, with strong deformation as they might appear in cartoon animation. These clay-figures have been 3D-scanned and converted to a blend-shape model by matching the neutral model (source) to the other expressions (targets). Photographs of the clay-figures and rendering of the blend shapes are shown in Figure 3.10. As mentioned in Section 3.2.2, for these extreme examples the joint-fairing did not unfold sufficiently with the default weights. However, increasing the smoothing weight allowed for the successful matching shown in Figure 3.10. Although our method has been developed primarily for faces, which typically are genus 0 objects with holes and boundaries, Figure 3.11 demonstrates on a teapot-to-cup morph that our method also works for higher genus models and for objects of disk-topology, such as the face scans shown in Figure 3.11.

In order to quantitatively evaluate our method, we give performance numbers and fitting accuracies in Table 3.1. The timings of our non-optimized single-threaded
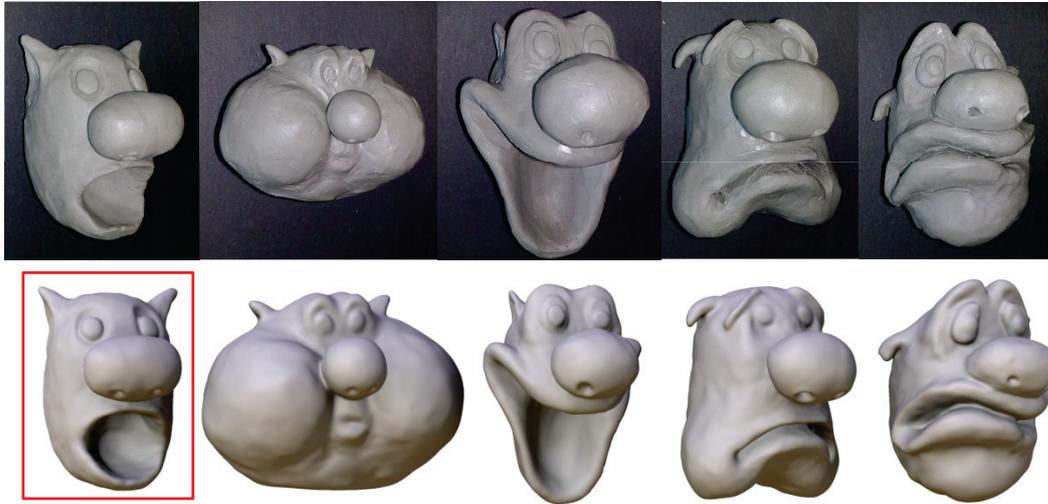
Figure 3.10: An application example to demonstrate the possibilities of our face matching technique. Clay-figures with extreme deformations have been sculpted, scanned, and transformed into a blend-shape model by matching the neutral model (*red frame*) to the other expressions. The top row shows photographs of the clay-figures, the bottom row shows the resulting blend shapes.
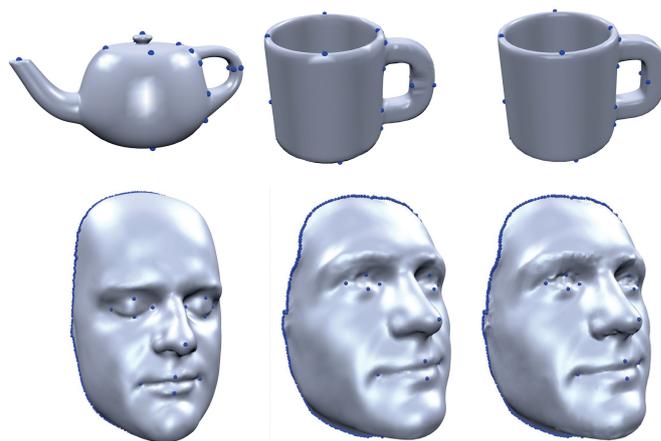
Figure 3.11: Fitting results for an object of genus one, teapot towards a cappuccino mug (*top*), and for face scans with disk topology (*bottom*). From left to right: source model, fitting result, target model. ©Left Scan: Infinite Realities

implementation were measured on a MacPro with Intel Xeon 2.67GHz. For all synthetic face models the fitting took just a few seconds, and even for the high resolution scanned models (Figures 3.10, 3.11) it only took about 30 seconds. The computational cost is dominated by the computation of closest point correspondences, which is $O(N^s \log N^t)$ due to the hierarchical kD-tree. Solving the linear systems using sparse Cholesky factorization is close to $O(N^s)$. Note that our simple and efficient method allows optimizing the position of each source vertex, which leads to highly accurate fits with errors (in terms of Hausdorff distance) well below 1% of the bounding box diagonal.

As mentioned in Section 3.1, there are many approaches for mapping one model to another. In Figure 3.12 we compare our method to the non-rigid registration of Weise et al. (2011) and blended intrinsic maps (Kim et al. 2011). Blended intrinsic maps compute correspondences in a fully automatic manner, but fail for all our examples, presumably because face models are highly symmetric and the geometric features are less prominent than the protruding arms and legs shown in the original paper (Kim et al. 2011). To allow for a fair comparison to Weise et al. (2011), we manually specified more feature correspondences at the boundaries for Weise's method in order to compensate for our automatic boundary correspondences (cf. Figure 3.5). While their results are similar to ours for near-isometric models (Figure 3.12, bottom row), self-intersections in the nose or mouth area can be observed for the first and second row. For the Slimer example (third row) their method yields wrong alignments near the self-intersecting regions of the chin and neck of the target model.

We would have also liked to compare to Kraevoy & Sheffer (2004), but even with the strong involvement of the authors we could not restore a completely working version of their software from original code fragments. Our final implementation is able to produce the initial cross-parametrization, but cannot perform the subsequent smoothing of the parametrization. As a consequence, we cannot produce full cross-parametrizations similar to the examples shown in Figure 3.12. However, we can evaluate how many additional vertices have to be inserted during their initial path creation step. When testing the top row example of Figure 3.12, the numbers of vertices for the source and target meshes increase from 7224 and 2911 to 8611 and 4071, respectively. This massive change in connectivity strongly contradicts our goals described at the beginning.

Hence, compared to other approaches for establishing correspondences, our method has a few important advantages. First, in contrast to cross-parametrization techniques (Kraevoy & Sheffer 2004, Schreiner et al. 2004), it does not require inserting new vertices or edges, which would prevent the compatible fitting to multiple target models (Figure 3.9). Second, in contrast to parametrization-based methods (Blanz & Vetter 1999, Lipman & Funkhouser 2009), it can handle models of higher genus (Figure 3.11). Third, in contrast to most non-rigid registration techniques (Huang et al. 2008, Weise et al. 2011) it can handle highly non-isometric input models, where existing methods often produce self-intersections (Figure 3.12). Finally, our method is easier to implement, very robust, and more efficient than most other approaches.

A limitation of our technique is that the aggressive fairing might collapse protruding extremities, such as arms, or legs. Because of this, the subsequent registration is not capable of determining correct correspondences, which results in strongly distorted triangles (see Figure 3.13). This collapsing can be avoided by manually specifying additional reference points but requires considerably more work.
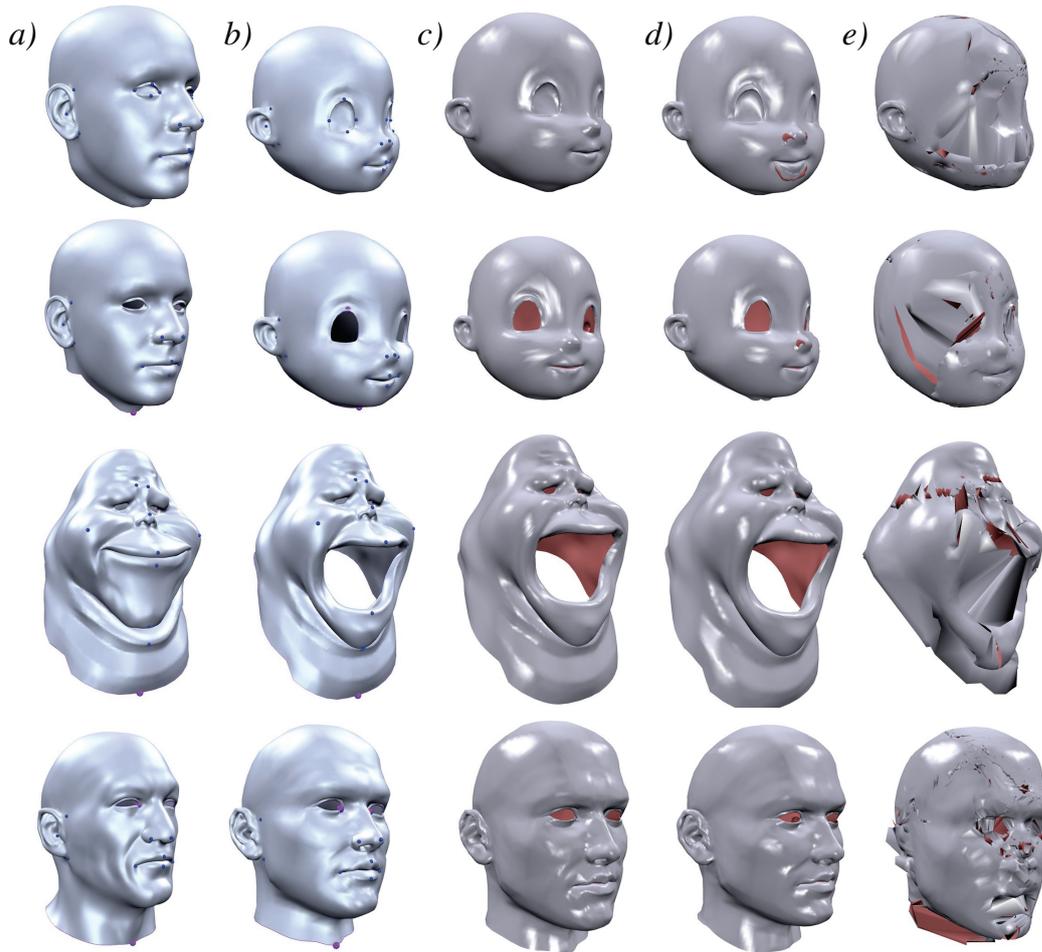
Figure 3.12: Comparison of fitting source models (*column a*) to target models (*column b*). While our method (*column c*) handles all examples without problems, the method of Weise et al. [2011] (*column d*) causes self-intersections around the nose for the upper two examples and around the mouth for the third example. The blended intrinsic maps [Kim et al. 2011] (*column e*) are not suitable for face matching. ©Dave: TurboSquid, ©Loki: Mark Pauly, ©Slimer: Thibaut Weise, ©Viktor: Faceware Technologies
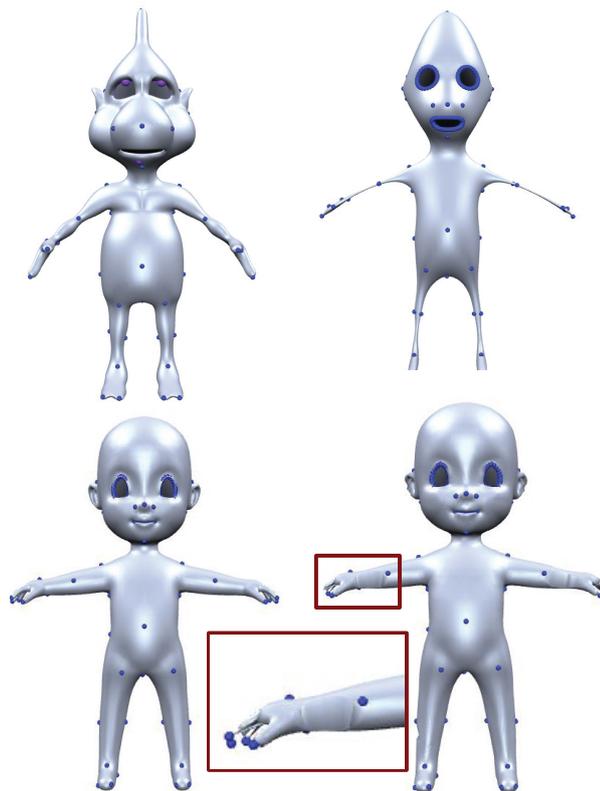
Figure 3.13: When fitting the source character (*top left*) to the target character (*bottom left*), our method fails at extremities. These tend to collapse during the joint-fairing (*top right*), thus causing wrong correspondences for the subsequent fitting and leading to distorted triangles in the fitting result (*bottom right*).

4 Facial Retargeting with Range of Motion Alignment

Figure 4.1: Captured facial expression of two actors (*left* and *right*) with retargeting results for realistic and stylized characters. Our method automatically aligns the ranges of motion of the captured actor and the target blendshape rig, such that expressions are restored faithfully even for stylized characters.
©Motion capture: Feel Ghood Music, Face rigs: Mark Pauly, meryproject.com, Jana Bergevin

The previous chapter addressed the problem of transferring properties between static faces. Although static faces provide significant information for judging the appearance of a person (McDonnell et al. 2012) and are for practical reasons the primary type of stimuli in perceptual studies, e.g., Chapters 6 and 7, it is more common to encounter moving or animated faces in real-life and in applications like movies or games. When testing the perception of animated faces, having a semantically equivalent sequence of different characters is desirable, because otherwise, variance in perception might originate from different animations. Transferring facial animations across faces of different proportions is a well-known problem in computer graphics.

Facial animation retargeting (Figure 4.1) addresses the general problem of animation transfer between virtual characters, with the transfer of performance capture to virtual characters being the primary application within the industry. Recent developments in vision- and depth-sensor-based facial motion capture (Weise et al. 2011, Li et al. 2013, Ichim et al. 2015, Cao et al. 2014, Thies et al. 2016) made accurate captures of an actor, traditionally limited to big film or game studios, affordable to a much broader audience. Current real-time capture systems typically adopt a realistic generic blendshape model to the actor. Since the modified and the original character have semantically equivalent blendshapes, the captured actor performance is then transferred between the characters by directly mapping the blendshape weights. The special case of equivalent blendshapes between two characters is often named as parallel parametrization in retargeting context.

In practice, it is uncommon to encounter facial rigs with a complete set of semantically equivalent blendshapes. Creating facial rigs for animation is time-consuming and requires highly skilled artists. Therefore, a rig is carefully designed to fit the animation needs, only modeling the necessary expressions. In addition, expressive

digital characters are often stylized and exaggerate facial proportions of humans. An effective retargeting method must either transfer animation from facial markers to a blendshape rig or between faces with different blendshape sets. Several retargeting approaches generate their own parallel parametrization, by transferring the blendshapes of the character face rig to align with the actor's proportions. However, especially for stylized characters this step often fails, due to differences in ranges of motion or shortcomings of current methods. The subsequent blendshape estimation becomes erroneous, which has been addressed so far by incorporating additional priors.

In this chapter, we propose a novel algorithm for creating actor-specific blendshapes with the help of a training sequence consisting of an actor's facial motions that semantically correspond to the blendshapes of the character face rig. We show that given a training sequence that sufficiently covers the actor's range of motion, it is possible to create, in an unsupervised manner, an accurate parallel parametrization – even if the facial rig and the actor differ strongly in their facial proportions. The key observation is that facial motions are similar across different stylization levels, as motivated by the Facial Action Coding System (FACS) (Ekman & Friesen 1978). The FACS describes facial expressions on the basis of muscle activations and is a frequent reference for blendshape creation for stylized and realistic human characters. Based on a new manifold alignment approach and a novel energy measuring similarity of facial expressions, we successfully align the ranges of motion of the actor and the character face rig. This subsequently leads to accurate retargeting.

Our second contribution is a prior energy based on physically-inspired deformations, which can be computed efficiently even in real-time applications. Our geometric prior addresses the few artifacts that remain even in case of accurate parallel parametrizations. Both contributions are fully compatible with most previous methods, suitable for real-time applications, and produce results comparable or better than state-of-the-art offline methods (Seol et al. 2012) (Figure 4.15).

Contribution   This project is the result of a collaboration with the KAIST Visual Media Lab (Daejeon, South Korea). On the basis of an earlier developed blendshape visualization software (comparable to Figure 4.6), I developed first prototypes for blendshape transfer based on spherical harmonics or k-means clustering. The last prototype served as a proof-of-concept with reasonable retargeting results for many characters. Roger Ribera i Blanco implemented the final version in close collaboration with me, including the geometric prior. Out of all ideas that are described in the paper, Roger Ribera i Blanco contributed most to the manifold alignment (Section 4.3.3), key expression extraction (Section 4.3.2) and geometric constraint sections (Section 4.3.4), while I proposed most of the ideas related to the original prototype, measuring expression similarity and contrast enhancement (Section 4.3.1). Ideas on the Cross-Expression Graph (Section 4.3.5) and geometric prior (Section 4.4) can be attributed equally to both first authors.

4.1   Overview of Retargeting Methods

As a key element of human-centered applications, research on virtual faces and face animation has been an active field of research for decades, resulting in a wide range of publications on this topic. For a general overview, we recommend the book of Parke & Waters (2008) and the more recent surveys focusing on rigging (Orvalho

et al. 2012) and blendshape animation (Lewis et al. 2014). In the following, we focus mainly on facial retargeting and assume a certain familiarity with blendshape-based facial animation (Section 4.2).

Cross-Mapping   To overcome problems at the transfer stage, cross-mapping methods learn directly from semantically corresponding facial expressions of the captured actor and a target face rig and synthesize new poses based on these training examples. Different learning techniques were proposed, starting with piece-wise linear mapping (Buck et al. 2000) and locally linear embedding (Wang et al. 2004), followed by more advanced machine learning algorithms like RBFs (Deng et al. 2006), kCCA (Song et al. 2011), simplicial basis (Kholgade et al. 2011), or shared Gaussian Process Latent Variable Models (sGPLVM) (Bouaziz & Pauly 2014). A key advantage of all cross-mapping approaches is that they are applicable to any type of character (e.g., even having a different number of eyes) or any facial rig (blendshapes, muscles, etc.). Unfortunately, the performance of these methods is strongly tied to the quality and number of given training examples. Often, at least 15–20 corresponding example pairs are required for sufficient results. For a moderate facial rig with 40 blendshapes, this leads to 600–800 parameters which must be defined consistently by hand. However, even in case of consistent training examples, the resulting expressions still remain (sophisticated) interpolations of the training examples. This often leads to inaccurate results for expressions that are too different from the training examples.

Parallel Parametrization   The simplest way to transfer an animation from one character to another is by creating two semantically equivalent facial rigs. In this case, the animation can simply be transferred by copying the control parameters from one rig to another. For blendshape-based facial rigs, manually creating semantically corresponding sets of blendshapes is a labor-intensive task, requiring not only excellent modeling skills and anatomical knowledge of the face but also a considerable amount of time. To automate this process, several approaches for transferring blendshapes from a generic face model to a neutral target have been suggested.

Given a source blendshape rig and the neutral face of a target character, Noh & Neumann (2001) suggested first to establish dense correspondences and then to transfer per-vertex displacements for each expression. This was later improved using deformation gradients (Sumner & Popović 2004) or Radial Basis Functions (Orvalho et al. 2008, Seol et al. 2011, 2012). Several improvements have been suggested since then, ranging from incorporating examples (Li et al. 2010), adding contact constraints (Saito 2013), interactive editing (Xu et al. 2014), to iterative refinement schemes for real humans (Bouaziz et al. 2013, Ichim et al. 2015, Seol et al. 2016). However, if the assumption that source and target models are of similar shape is violated, deformation transfer and similar methods often fail to preserve the semantics of the facial expressions (see Figure 4.2). The resulting proportional mismatch then leads to artifacts at the retargeting stage: exaggerated actor blendshapes cause dampened animations because smaller weights are sufficient to reach a target pose; conversely, dampened blendshapes cause larger weights and exaggerated animations – up to the point of unnatural face deformations (Figure 4.3).

Seol et al. (2012) address artifacts resulting from erroneous expression transfer by integrating velocities over a sequence of captured frames. In contrast, we improve the transfer process, such that the transferred blendshapes automatically adapt

49
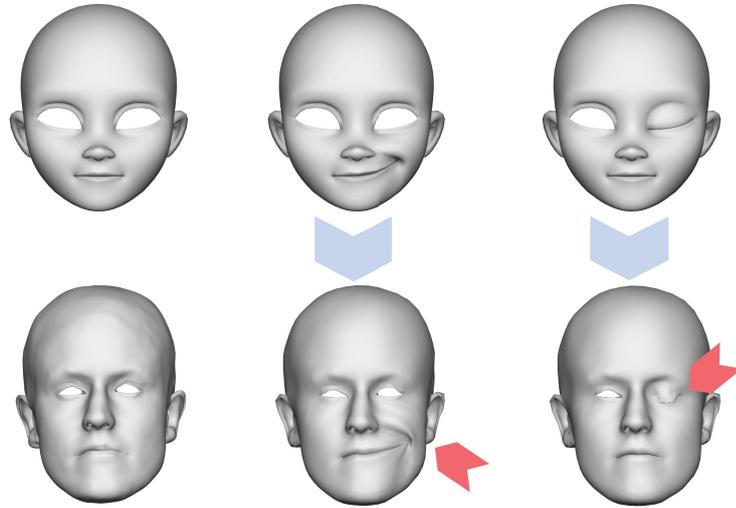
Figure 4.2: Transferring blendshapes using deformation transfer leads to unnatural deformations (*center*) or broken expressions (*right*) if facial proportions are too different. Neutral expressions are shown on the left.
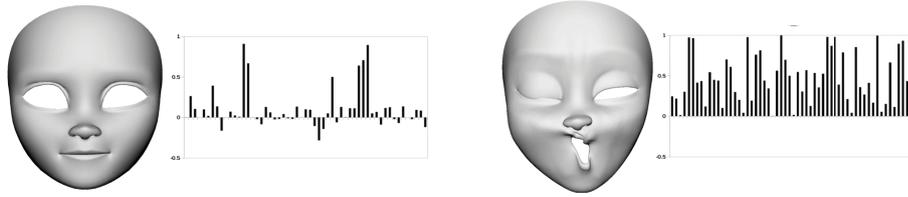
©Face rig: meryproject.com
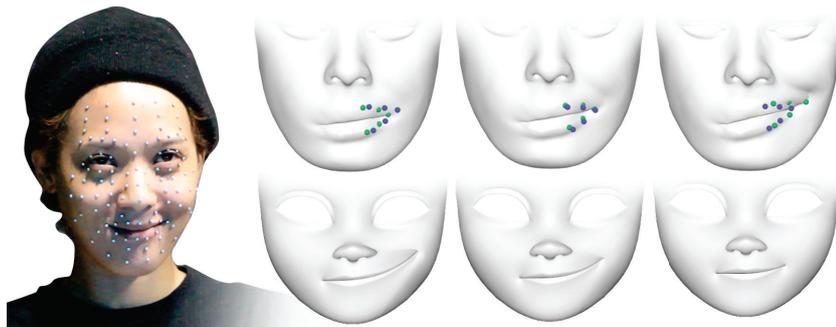
Figure 4.3: Retargeting a smile (*left*) to differently personalized blendshapes (*top*). Close-ups show the positions of the captured markers (*blue*) and the corresponding vertices (*green*). Only well-matching blendshapes result in accurate retargeting (*center*). Dampened personalized blendshapes (*top left*) cause over-exaggerated retargeting results (*bottom left*), since large weights are necessary to fit the captured markers. Inversely, exaggerated blendshapes (*top right*) cause damped retargeting results (*bottom right*).

©Motion capture: Feel Ghood Music, ©Face rig: meryproject.com

Figure 4.4: Complex interaction of blendshape weights. *Left:* Weight cancellation effects lead to a valid neutral face. *Right:* Constraining all weights to the interval $[0, 1]$ does still not guarantee valid face expressions.
©Face rig: meryproject.com

to the actor's range of motion. Given a motion sequence of an actor and sparse correspondences between the actor and the character model (e.g., in form of optical markers), our method automatically transfers the blendshapes of the face rig to the actor space.

Manifold-based Techniques Aligning the ranges of motion between the actor's motion sequence and the character's blendshape rig significantly improves our expression transfer and is inspired by the success of manifold alignment methods (Pan & Yang 2010). These approaches aim at registering two different high-dimensional data sets in a lower-dimensional embedding space. The mapping into the lower-dimensional space has to minimize the distance between the individual manifolds as well as to keep the original relationship between the data elements by preserving the geometric structure of the manifolds. Several unsupervised methods (Wang & Mahadevan 2009, Fan et al. 2016) have been proposed for various applications, including transfer learning (Pan & Yang 2010), data mining, automatic translation or image set matching (Pei et al. 2012, Cui et al. 2012). An important aspect is the dimensionality reduction, where additional constraints ensure optimal embedding spaces. Often, transformations between embedding spaces are then solved by eigen-decomposition of the graph Laplacian (Fan et al. 2016, Wang & Mahadevan 2011, 2013). In some sense, manifold alignment techniques aim to find a low-dimensional space where Euclidean distances better represent the similarity between the different data instances. In contrast, we want to identify character blendshapes that match the proportions and ranges of motion of the actor in the high-dimensional space. This requires the transfer of the original blendshapes to the actor space, instead of projecting into a low-dimensional space.

Expression Regularization Common approaches reduce artifacts in blendshape animation by restricting the blendshape weights to a fixed interval (Bregler et al. 2002, Chuang & Bregler 2002) or by penalizing large weights (Seo et al. 2011). However, such heuristics do not always succeed because combinations of blendshape weights outside the specified range can still produce valid faces (Seol et al. 2011), and restricting blendshape weights to $[0, 1]$ will not necessarily result in plausible expressions (Figure 4.4). This phenomenon is commonly known as *blendshape interference* (Lewis et al. 2005). Alternatively, PCA-based priors have been proposed for direct blendshape manipulation (Lau et al. 2009, Anjyo et al. 2012) and retargeting (Seol et al. 2012). But these approaches strongly depend on the quality and
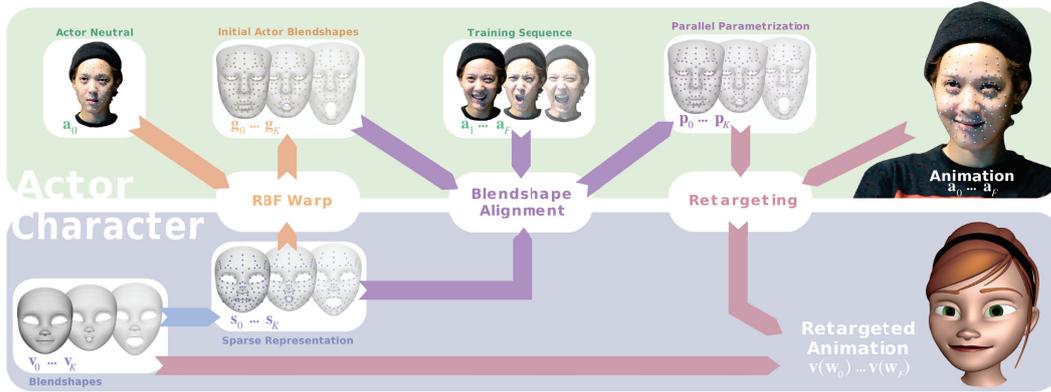
Figure 4.5: Overview of facial animation retargeting pipeline, described for illustration purpose on the use case of retargeting marker-based animation. Our method addresses the problem of obtaining personalized blendshapes by aligning the range of motion of the blendshape model with the range of motion of the actor. In addition, we present a new prior for the retargeting step. ©Motion capture: Feel Ghood Music ©Face rig: meryproject.com

amount of training examples, where an insufficient set of example poses biases the solution towards the closest training data (Anjyo et al. 2012). In contrast, we consider smooth skin deformation as a key factor and propose a prior that penalizes surface deformation similar to many physically-inspired facial animation methods (Bickel et al. 2007, Ichim et al. 2016, Barrielle et al. 2016). While all these methods outperform linear blendshapes in physical accuracy, the visual improvements often do not justify the additional computation costs for many applications. Therefore, we reformulate the large-scale deformation energy of Bickel et al. (2007) to a suitable geometric prior for blendshape weights. Furthermore, in order to additionally achieve sparse weight activation, one can either regularize using the L_1 norm of blendshape weights (Bouaziz et al. 2013) or transfer common practices in manual key-framing (Seol et al. 2011).

4.2 Blendshape Animation and Retargeting

In this section, we briefly review blendshape facial animation and blendshape-based facial retargeting and set up our notation. Section 4.3 then introduces our improved blendshape transfer, which is a pre-processing step before the actual retargeting. The retargeting itself can be further regularized using our geometric prior (Section 4.4). Finally, we compare our proposed approach to state-of-the-art methods in Section 4.5.

Let the facial rig be given as a polygon mesh $\mathcal{M}$, consisting of N vertices, posed in a neutral expression, and being equipped with K expression blendshapes that all share the connectivity of $\mathcal{M}$. We denote the vector of stacked vertex positions of the neutral face by $\mathbf{v}_0 = (\mathbf{v}_0^1, \ldots, \mathbf{v}_0^N)^\mathsf{T}$, and of the k-th blendshape by $\mathbf{v}_k = (\mathbf{v}_k^1, \ldots, \mathbf{v}_k^N)^\mathsf{T}$. Due to the coupling of the x/y/z-coordinates, the blendshapes $\mathbf{v}_k$ denote $3N$-dimensional vectors.

For blendshape face animation, we employ the delta-blendshape formulation (Lewis et al. 2014), where the neutral expression $\mathbf{v}_0$ is subtracted from the blendshape ex-

pressions $\mathbf{v}_k$ to yield a displacement field for activating a particular expression: $\delta\mathbf{v}_k = \mathbf{v}_k - \mathbf{v}_0$. New facial poses $\mathbf{v}(\mathbf{w})$ are computed by displacing the neutral face by a weighted sum of delta-blendshapes, with weights $\mathbf{w} = (w_1, \ldots, w_K)^{\mathsf{T}}$, which can also be written in matrix form using the delta-blendshape matrix $\delta\mathbf{V} = [\delta\mathbf{v}_1, \ldots, \delta\mathbf{v}_K]$:

$$\mathbf{v}(\mathbf{w}) \;=\; \mathbf{v}_0 + \sum_{k=1}^{K} w_k \delta\mathbf{v}_k = \mathbf{v}_0 + \delta\mathbf{V}\,\mathbf{w}\,. \tag{4.1}$$

The main application for facial retargeting is the transfer of an actor's performance capture to a virtual character, mostly using marker-based optical motion capture. We will therefore formulate our approach for this problem setting, but note that our method is not limited to marker-based retargeting, since any given facial animation or marker-less performance capture can be easily converted to a marker-based performance capture by tracing a subset of "marker vertices" through time.

The actor's performance is given as a $3M$-dimensional vector of M stacked marker positions $(\mathbf{a}^1, \ldots, \mathbf{a}^M)^{\mathsf{T}}$ that vary over time. For a particular motion capture frame f, this data is denoted as $\mathbf{a}_f = (\mathbf{a}_f^1, \ldots, \mathbf{a}_f^M)^{\mathsf{T}}$, and $\mathbf{a}_0$ represents a calibration frame of the actor in neutral expression. Like all retargeting methods based on parallel parametrization, we require sparse correspondences between the actor's face animation and the character's face rig. These correspondences are specified as pairs of points $\{\mathbf{a}_0^m, \mathbf{s}_0^m\}$, $m = 1, \ldots, M$ on the neutral expressions of the actor and the character rig. The same set of vertices on the expression blendshape $\mathbf{v}_k$ is denoted by $\mathbf{s}_k = (\mathbf{s}_k^1, \ldots, \mathbf{s}_k^M)^{\mathsf{T}}$. Since the number M of markers and corresponding vertices is much lower than the number of mesh vertices ($M \ll N$), the $\mathbf{s}_k$ are called the sparse representation of the blendshape $\mathbf{v}_k$. We employ the same delta-formulation as above for sparse blendshapes ($\delta\mathbf{s}_k = \mathbf{s}_k - \mathbf{s}_0$) and animation data ($\delta\mathbf{a}_f = \mathbf{a}_f - \mathbf{a}_0$).

The goal of any blendshape retargeting system is to compute the time-varying weights $\mathbf{w}$ that reproduce the facial expressions on the target face rig for a given actor's performance capture. This requires a set of personalized sparse actor blendshapes $\mathbf{b}_k = (\mathbf{b}_k^1, \ldots, \mathbf{b}_k^M)^{\mathsf{T}}$ that are semantically equivalent to the sparse blendshapes $\mathbf{s}_k$ of the character rig. For each captured frame f, the blendshape weights $\mathbf{w}_f$ can be computed by minimizing the squared distance between the marker displacements $\delta\mathbf{a}_f$ and a weighted combination of the actor's sparse delta-blendshapes $\delta\mathbf{b}_k = \mathbf{b}_k - \mathbf{b}_0$:

$$E_{\mathrm{Fit}}(\mathbf{w}) \;=\; \frac{1}{M} \left\| \delta\mathbf{a}_f - \sum_{k=1}^{K} w_k\, \delta\mathbf{b}_k \right\|^2 . \tag{4.2}$$

The required personalized actor blendshapes $\mathbf{b}_k$ are either manually created or transferred from the face rig to actor space (Sumner & Popović 2004, Orvalho et al. 2008, Seol et al. 2012). But as discussed above and shown in Figures 4.2 and 4.3, this blendshape transfer often fails for highly different facial proportions, such as stylized characters. Therefore, we propose an improved approach for automatic blendshape transfer with a range of motion adjustment in Section 4.3.

In order to resolve ambiguities, prevent over-fitting, or penalize artifacts, the above blendshape fitting process is typically regularized through additional energy terms:

$$E_{\mathrm{Retarget}}(\mathbf{w}) \;=\; E_{\mathrm{Fit}}(\mathbf{w}) + E_{\mathrm{Reg}}(\mathbf{w})\,. \tag{4.3}$$

Typical choices for the energy $E_{\text{Reg}}(\mathbf{w})$ are (weighted combinations of) L_2 regularization $\|\mathbf{w}\|^2$ to penalize large weights (Lewis et al. 2014), L_1 regularization $\|\mathbf{w}\|_1$ for inducing sparsity (Bouaziz et al. 2013), and penalization of temporal changes $\|\mathbf{w}_{f-1} - \mathbf{w}_f\|^2$ to remove jitter (Lewis et al. 2014). However, as we will show in Section 4.4, our novel geometric prior, which operates on differential mesh properties instead of on blendshape weights, prevents geometric artifacts while at the same time allowing for more expressive animation.

4.3 Automatic Blendshape Transfer

A correct set of actor-specific blendshapes $\mathbf{b}_1, \ldots, \mathbf{b}_K$, used in the fitting term (4.2), is a crucial component of any retargeting method. Existing approaches for transferring the character's blendshapes to the actor space compensate for proportional mismatches between character rig and actor to a certain degree. Unfortunately, they often fail to properly align the respective ranges of motion, in particular for stylized characters (see Figures 4.2, 4.3). Our approach addresses these shortcomings and is motivated by two main observations:

- Blendshapes typically define the strongest deformation caused by activating isolated facial muscles.

- Semantically equivalent expressions are highly similar for different characters because facial muscles are consistent across humans and remain consistent even for highly stylized characters for the sake of easy expression recognition. Although semantically equivalent expressions are similar across identities, they are not equal and differ with respect to direction and magnitude.

We improve the blendshape transfer from the facial rig to the actor's proportions by aligning their facial expression manifolds. We learn the actor's expression manifold from a short, captured training sequence, consisting of F animation frames $\mathbf{a}_1, \ldots, \mathbf{a}_F$. The actor's training sequence should contain semantically equivalent expressions to the blendshapes of the facial rig; it should cover the actor's range of motion since blendshapes often correspond to extreme expressions. However, the training expressions might combine several blendshapes in arbitrary order. As introduced above, the animation data $\mathbf{a}_f$ consist of M markers with point correspondences on the facial rig. The character's sparse blendshapes $\mathbf{s}_1, \ldots, \mathbf{s}_K$ define well the expression manifold of the character rig.

Given the blendshape rig and a set of training expressions, our method computes personalized actor blendshapes based on concepts from manifold alignment (E_{Match}, Section 4.3.3). To this end, we measure the similarity between the character's blendshapes and actor's captured performance (Section 4.3.1) and extract the most essential frames from the training data (Section 4.3.2). Two conditions regularize the alignment process: The actor-specific blendshapes should not deviate too much from an initial guess derived by RBF deformation (E_{Mesh}, Section 4.3.4), and relations between individual blendshapes should be preserved by constructing a Cross-Expression Graph (E_{CEG}, Section 4.3.5). Figure 4.5 illustrates our blendshape transfer schematically in the context of facial retargeting.

Our blendshape transfer process is formulated as the minimization of the following energy (with $\alpha = 0.01$ and $\beta = 0.1$ in all shown examples) and optimizes
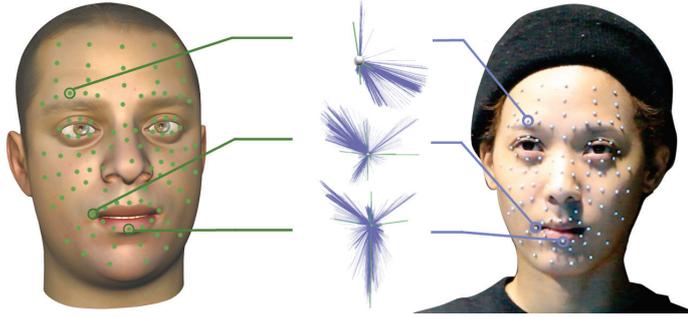
Figure 4.6: Motion space comparison between an actor (*right*) and a blendshape model (*left*). *Center*: Overlaid motion space in delta representation for selected markers of the actor (*blue*) and corresponding vertices in the blendshape model (*green*). Each line corresponds to a different frame/blendshape.

Face rig courtesy of *Jason Osipa*. ©Motion capture: Feel Ghood Music

all personalized delta-blendshapes at the same time, where the $3MK$-dimensional vector $\delta\mathbf{b} = (\delta\mathbf{b}_1, \ldots, \delta\mathbf{b}_K)^T$ contains the stacked blendshapes:

$$E_{\text{Align}}(\delta\mathbf{b}) \;=\; E_{\text{Match}}(\delta\mathbf{b}) + \alpha E_{\text{Mesh}}(\delta\mathbf{b}) + \beta E_{\text{CEG}}(\delta\mathbf{b}) \,. \tag{4.4}$$

4.3.1 Facial Motion Similarity

The similarity of the ranges of motion is best visualized by considering the displacements of a single marker $(\delta\mathbf{a}_f^m = \mathbf{a}_f^m - \mathbf{a}_0^m)$ over the entire training sequence and the delta-blendshapes of the corresponding vertex on the facial rig $(\delta\mathbf{s}_k^m = \mathbf{s}_k^m - \mathbf{s}_0^m)$, as shown in Figure 4.6. In some cases clear clusters of marker displacements can be identified, indicating that different motions on that particular area occur mainly independently. For example, on the right brow (Figure 4.6, eyebrow) three clearly distinctive motions can be identified, corresponding to raising and lowering the brow and frowning. Considering that the captured training sequence consists of over 2000 frames, the clear separation into these motion clusters is surprising. This observation fuels our motivation to consider an expressive training sequence to generate an actor-specific parallel parametrization. Still, a clear solution may not always exist (Figure 4.6, lips). Even though there might be a lot of data available, it is possible that no clear one-to-one correspondence between the performance and specific blendshapes can be established. When a clear correspondence can be identified, it is best to exactly match the specific expression. However, aligning blendshapes with unrelated expressions will result in a complete loss of the semantic equivalence between the actor's performance and the retargeted animation.

In order to quantify the similarities between a blendshape k and the actor's expression at frame f, we compute the Pearson Correlation Coefficient between $\mathbf{a}_f$ and $\mathbf{s}_k$. In our case the mean of a sampling set is replaced by the more meaningful neutral facial expression. The computation of the correlation coefficient $\rho_{k,f}$ between an actor's expression $\mathbf{a}_f$ and a sparse blendshape $\mathbf{s}_k$ then simplifies to the following equation in delta-representation:

$$\rho_{k,f} = \frac{\delta\mathbf{a}_f \cdot \delta\mathbf{s}_k}{\|\delta\mathbf{a}_f\| \, \|\delta\mathbf{s}_k\|} \,. \tag{4.5}$$
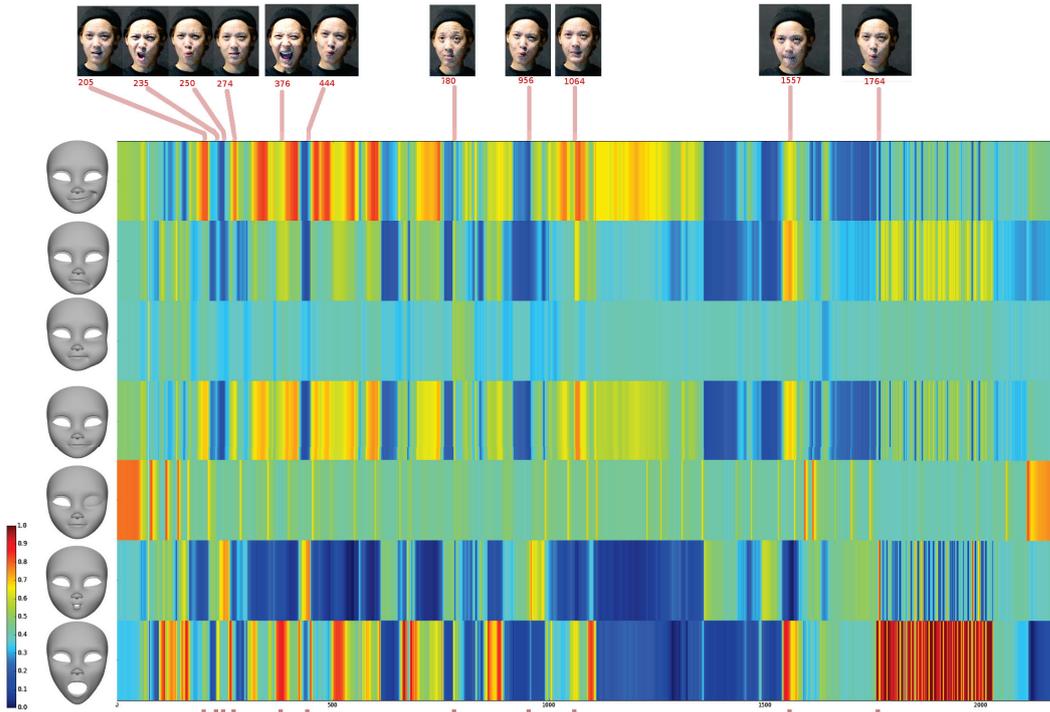
Figure 4.7: Motion space similarity between selected blendshapes (*rows*) and actor's performance (*columns*) consisting of 2150 frames. Please notice the blocking structure, indicating redundancy of information.

©Motion capture: Feel Ghood Music ©Face rig: meryproject.com

Figure 4.7 shows the resulting similarity measures between selected blendshapes and a training sequence. This simple dot-product formulation of Equation (4.5) is effective, because:

- Displacements of blendshapes and actor's expressions in similar directions have a high correlation,

- The locality of blendshapes is considered, because vertices that do not move in blendshape $\delta\mathbf{s}_k$ cancel out the contributions of the corresponding expression $\delta\mathbf{a}_f$.

Contrast Enhancement Two heuristics can further improve the similarity measure: Identifying important blendshapes for a frame in the training sequence (i) is easier for more unique blendshapes and (ii) is easier for blendshapes with the strongest displacement. Both properties are considered by computing a trust value for each sparse blendshape. We first compute the total displacement $d_k = \|\delta\mathbf{s}_k\|$ of blendshape $\delta\mathbf{s}_k$. In a second step, the sparse blendshapes are re-ordered according to their displacements d_k, such that $\delta\mathbf{s}_1$ is the blendshape with the largest displacement and $\delta\mathbf{s}_K$ the one with the smallest displacement. We then build a between-blendshape correlation matrix $\mathbf{P}(k,i) := \rho_{k,i}$ (see Figure 4.11). Finally, the trust value t_k is computed as

$$t_k \;=\; 1 - \frac{\sum_{i=1}^{k-1} \rho_{k,i}^{+}}{\max_{1 \le k \le K}\left(\sum_{i=1}^{k-1} \rho_{k,i}^{+}\right)}\,, \tag{4.6}$$

where the similarity between sparse blendshapes is measured using the positive Pearson Correlation Coefficient $\rho_{k,i}^{+} = \max(0, c_{k,i})$. The sum $\sum_{i=1}^{k-1} \rho_{k,i}^{+}$ adds all $\rho_{k,i}^{+}$ within row k of the strictly lower triangle matrix, and the dominator is the maximum of all row sums. The sparse blendshape with the largest displacement has no entries in the strictly lower triangle matrix, such that $t_1 = 1$. All remaining sparse blendshapes will only have $t_k \approx 1$ if they are uncorrelated to sparse blendshapes with larger displacements. Hence, a blendshape modeling a subtle lip motion will likely get a very low trust value due to high correlations with other, more expressive mouth blendshapes, while a brow-raising blendshape will have a high trust value as it is mostly uncorrelated to other blendshapes.

In practice, most computed similarity values, except the blendshape with maximum displacement, will never reach a trust value of one, because even semantically equal expressions differ for different faces. Therefore, we propose to amplify high correlation values and reduce low correlation values using the following transformation:

$$\chi(\rho_{k,f}) \;=\; \frac{e^{r\rho_{k,f}^{+}}}{e^{r/2} + e^{r\rho_{k,f}^{+}}}\,. \tag{4.7}$$

The steepness r of the function is set to 15 in all our examples. Finally, we linearly interpolate between the original and the modified correlation values, depending on the trust value. The computed similarity $\tilde{\rho}_{k,f}$ replaces then the original correlation from Equation (4.5):

$$\tilde{\rho}_{k,f} \;=\; (1 - t_k)\,\rho_{k,f}^{+} + t_k\,\chi\!\left(\rho_{k,f}^{+}\right)\,. \tag{4.8}$$

4.3.2 Key Expression Extraction

Blendshapes represent peak expressions that we want to match to the actor's most similar expressions. After computing the similarities $\tilde{\rho}_{k,l}$ between blendshapes and the training sequence, we remove the temporal redundancy between consecutive frames following Coleman et al. (2008). The correlations of each blendshape over the whole training sequence (corresponding to a row in Figure 4.7) are first low-pass filtered to remove some superfluous noise. The employed Gaussian kernel is kept small (three frames wide) to avoid over-smoothing fast and peak motions. Finally, all filtered rows are added together by summing over columns. In this cumulative representation of the data, we extract all local peaks in order to obtain a set of most similar facial expressions. Figure 4.8 shows the resulting cumulative function and the extracted peaks.

4.3.3 Manifold Alignment

Given a similarity measure $\tilde{\rho}_{k,f}$ between a sparse blendshape and an actors' expression, our goal is to fit the personalized blendshapes $\delta\mathbf{b}_k$ to the actor's expressions $\delta\mathbf{a}_f$. This means that a closed-eye blendshape should be fitted to an actor's expression with closed eye(s) in the training sequence, without being influenced by
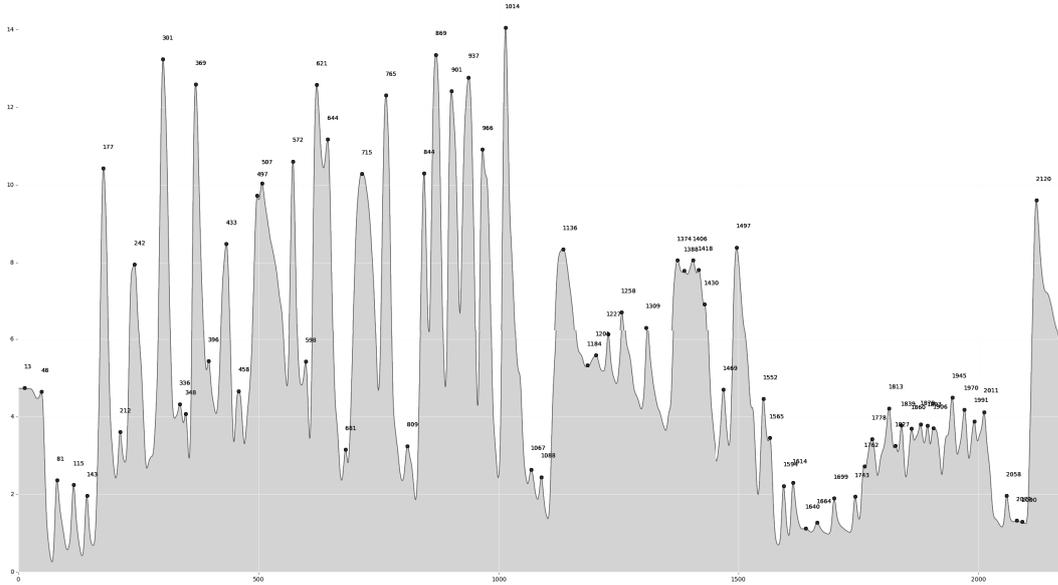
Figure 4.8: Cumulative correlation function (not displacements) for identifying peak expressions within a training sequence. 82 expressions have been identified from the similarities shown in Figure 4.7.

the negatively correlated eye-opening movements. Inspired by manifold alignment techniques we first consider the problem as the minimization of the following energy:

$$E_{\text{Match}}(\delta\mathbf{b}) \;=\; \frac{1}{F}\sum_{f=1}^{F}\sum_{k=1}^{K}\tilde{\rho}_{k,f}\,\|\delta\mathbf{b}_k - \delta\mathbf{a}_f\|^2\;. \tag{4.9}$$

However, this energy formulation would break the local support of blendshapes. Figure 4.9 shows an example of closing the eyes. In general, eyes blink simultaneously, leading to high correlations of both eye-closing blendshapes (left and right). Without special consideration of the local support of each blendshape, the alignment is distributed between both blendshapes. Including a mask of the blendshape displacements effectively disambiguates the displacement distribution (Figure 4.10). Similar to Seol et al. (2016) we encode local support with a soft mask vector ψ_k for blendshape k. The mask entries corresponding to the x/y/z coordinates of marker m of blendshape k are computed as $\|\delta\mathbf{s}_k^m\|\,/\max_{1\le m\le M}\|\delta\mathbf{s}_k^m\|$, with $\max_{1\le m\le M}\|\delta\mathbf{s}_k^m\|$ denoting the largest marker displacement within the blendshape $\delta\mathbf{s}_k$. Our final matching function is then

$$E_{\text{Match}}(\delta\mathbf{b}) \;=\; \frac{1}{F}\sum_{f=1}^{F}\sum_{k=1}^{K}\tilde{\rho}_{k,f}\,\|\delta\mathbf{b}_k - \text{diag}(\psi_k)\cdot\delta\mathbf{a}_f\|^2\;. \tag{4.10}$$

4.3.4 Geometric Constraint

Deformation transfer and similar automatic approaches in general, do not create accurate parallel parametrizations. However, such methods preserve the original semantics and local properties of the transferred expression quite well. A transferred
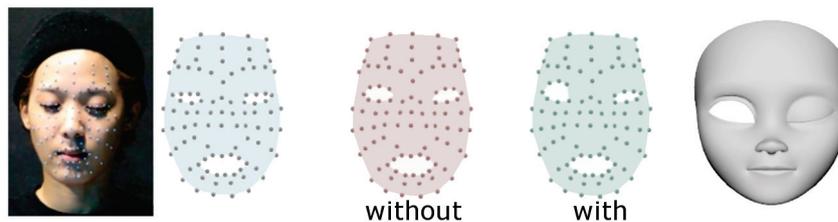
Figure 4.9: Blendshapes local support. *From left to right:* A closed-eye frame and corresponding marker positions (*blue*), fitting results of the eye-closed blendshape using E_{Match} without local support (*red*) and with local support considered (*green*), and the original blendshape for reference.
©Motion capture: Feel Ghood Music ©Face rig: meryproject.com

Figure 4.10: Local deformations of blendshapes. Red encodes the strongest displacement within a blendshape. *From left to right::* Left eye closed, kiss, open smile, right smile, and jaw drop. ©Face rig: meryproject.com

smile will remain a recognizable smile, although it might be too dampened or too exaggerated. Preserving the local features of an expression, e.g., the o-shape of the lips for the kiss expression is the intention of the geometric constraint.

Based on the sparse correspondences between the neutral expressions $\mathbf{s}_0$ (of the character rig) and $\mathbf{a}_0$ (of the actor), we create an initial guess $\mathbf{g}_k$ for each personalized blendshape $\mathbf{b}_k$ using RBF-deformations (Orvalho et al. 2008, Seol et al. 2012). To this end, we first compute an RBF thin-plate spline that transforms the neutral expression $\mathbf{s}_0$ to $\mathbf{a}_0$, by placing an RBF center at every marker $\mathbf{s}_0^m$ and solving for the RBF weights. The resulting RBF function is then used to convert all delta-blendshapes $\delta\mathbf{s}_k$ to the initial guesses $\delta\mathbf{g}_k$ of the actor blendshapes. We also tried deformation transfer (Sumner & Popović 2004) for this process and can confirm the similarity of the results (Seol et al. 2012). We preferred the RBF deformation method due to higher stability in case of degenerate meshes.

The goal of the geometric prior is to preserve the local shape properties of the initial guesses $\mathbf{g}_k$ while computing the personalized blendshapes $\mathbf{b}_k$. Local shape properties can be encoded well using per-vertex Laplacians (Botsch & Sorkine 2008), but this requires a triangle mesh, and so far the sparse blendshapes have been defined as sets of M marker points only. We, therefore, triangulate the marker points (in the *uv*-domain given by the parametrization) and add edges connecting upper/lower eyelids and lip markers to benefit from contact relationships (Saito 2013).

We formulate our geometric constraint as an energy that penalizes the change of the Laplacians between the unknown vertices of the personalized blendshape $\delta\mathbf{b}_k$ and the initial guesses $\delta\mathbf{g}_k$. This formulation is equivalent to a physically-inspired energy that minimizes bending (Bickel et al. 2007, Saito 2013):
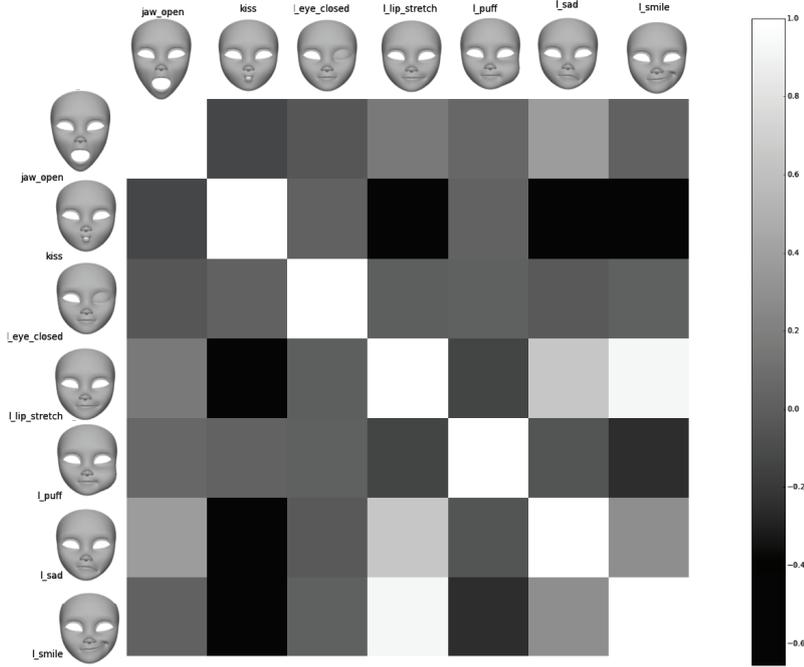
Figure 4.11: Computed Pearson Correlation Coefficient between an exemplary blendshape set of the Mery character. Signed correlations are used in Section 4.3.5 to preserve both similarities and dissimilarities between blendshapes. Only positive correlations between animation frames and blendshapes are required for Section 4.3.3. ©Face rig: meryproject.com

$$E_{\mathrm{Mesh}}(\delta\mathbf{b}) \; = \; \frac{1}{M} \sum_{k=1}^{K} \sum_{m=1}^{M} \| \Delta \left(\delta\mathbf{b}_k^m - \delta\mathbf{g}_k^m \right) \|^2 \; . \tag{4.11}$$

As a discretization of the Laplace operator we employ the standard cotangent weights (Pinkall & Polthier 1993).

4.3.5 Cross-Expression Constraint

The last energy term is responsible for maintaining the relationship between different blendshapes. If, for example, the mouth-open expression is corrected, this correction should also partly apply to the o-viseme. Inspired by application and relationship of graph and mesh Laplacians (Belkin & Niyogi 2005), we construct a Cross-Expression Graph that connects all blendshapes with each other. In this graph, each sparse blendshape becomes a node with edges to all other sparse blendshapes and edge weights are encoded using the signed similarity measure from Equation (4.5). We use signed correlations as edge weights (Figure 4.11) to preserve both similarities and dissimilarities between blendshapes. The weighted signed graph Laplacian (Kolluri et al. 2004, Kunegis et al. 2010) of one node is then defined as the weighted sum over its neighborhood:

	Loki	Osipa	Mery	Billy
blendshapes	46	33	54	39
key expressions sequence 1 (2150 frames)	73	71	82	76
key expressions sequence 2 (740 frames)	46	43	49	48
time (s)	18.71	5.37	23.25	7.57

Table 4.1: Details about the sequences, extracted key expressions, and the blendshape rigs, together with timings for automatic blendshape transfer. Both sequences were captured with 99 optical markers.

ⓒFace rigs: Mark Pauly, Jason Osipa, meryproject.com, Jana Bergevin

$$\Delta(\delta\mathbf{s}_k) \; = \; \frac{1}{\sum_{i \neq k} |\rho_{k,i}|} \sum_{i \neq k} \rho_{k,i} \left(\delta\mathbf{s}_i - \delta\mathbf{s}_k \right) . \tag{4.12}$$

The graph Laplacian $\Delta(\delta\mathbf{b}_k)$ for the personalized blendshapes is defined equivalently. Similar to Equation (4.11), the resulting energy term penalizes dissimilarity of correlated blendshape displacements:

$$E_{\text{CEG}}(\delta\mathbf{b}) \; = \; \frac{1}{M} \sum_{k=1}^{K} \| \Delta \left(\delta\mathbf{b}_k - \delta\mathbf{s}_k \right) \|^2 . \tag{4.13}$$

4.3.6 Numerical Optimization

Optimizing for the personalized blendshapes $\delta\mathbf{b}_k$ requires minimizing E_{Align} from Equation (4.9) consisting of the three quadratic energies E_{Match}, E_{Mesh}, and E_{CEG}. The energies are all separable in the x/y/z coordinates of $\delta\mathbf{b}_k$, but are coupled between all blendshapes through E_{CEG}. This leads to three linear systems of size $(MK \times MK)$, which is more efficient to solve than one big linear system of size $(3MK \times 3MK)$. The entire framework was implemented within Maya[1] using the Maya Python API, and all tests were performed on a computer with an Intel I7 3.4 GHz processor and 8 GB of memory. The linear systems were solved with the SciPy sparse solver (Jones et al. 2001). Details and timings are listed in Table 4.1.

4.4 Geometric Prior

After computing the personalized blendshapes $\delta\mathbf{b}_k$, the blendshape weights $\mathbf{w}_f$ can be estimated for each frame $\delta\mathbf{a}_f$ of the actor's performance. Despite accurately personalized blendshapes, the flexibility provided by the facial rig might not be sufficient to faithfully reproduce the actor's expressions, leading to artifacts in the retargeted face $\mathbf{v}(\mathbf{w}_f)$. We observe that one essential property of a valid expression is a surface free of fold-overs, i.e., without strong local bending. Thus, instead

---

[1]www.autodesk.com/maya

of formulating an arbitrary criterion for the values of the blendshape weights (see Section 4.2), we derive a physically-inspired prior similar to Equation (4.11) that penalizes surface bending and thereby eliminates fold-overs (Bickel et al. 2007, Saito 2013). We formulate the prior energy in terms of blendshapes and their weights, instead of in terms of vertex positions, so that our prior can be incorporated into any blendshape-based retargeting framework.

For faces, it seems natural to select the neutral pose $\mathbf{v}_0$ as the original, undeformed state. Thus our geometric prior penalizes bending between an expression $\mathbf{v}$ and the neutral face $\mathbf{v}_0$:

$$E_{\text{Prior}}(\mathbf{v}) \ = \ \frac{1}{N} \sum_{n=1}^{N} \|\Delta\left(\mathbf{v}^n - \mathbf{v}_0^n\right)\|^2 \ . \tag{4.14}$$

Based on Equation (4.1) the deformed face $\mathbf{v} = \mathbf{v}(\mathbf{w})$ can be written in terms of the blendshape weights $\mathbf{w}$. Analogously, the displacement $\mathbf{v} - \mathbf{v}_0$ can be written as $\delta\mathbf{V}\,\mathbf{w}$. Plugging this into the above prior energy and writing the Laplacian as a $(3N \times 3N)$ matrix $\mathbf{L}$, again using the cotangent weights (Pinkall & Polthier 1993), leads to the formulation of the prior energy in terms of $\mathbf{w}$:

$$E_{\text{Prior}}(\mathbf{w}) \ = \ \frac{1}{N} \|\mathbf{L}\,\delta\mathbf{V}\,\mathbf{w}\|^2 \ . \tag{4.15}$$

The combination of the fitting energy (4.2), the geometric prior energy (4.15), and a sparsity regularization $E_{\text{Sparse}}(\mathbf{w}) = \frac{1}{K}\|\mathbf{w}\|_1$, leads to our objective function for facial retargeting:

$$E_{\text{Retarget}}(\mathbf{w}) \ = \ \underbrace{E_{\text{Fit}}(\mathbf{w})}_{\text{ActorSpace}} + \underbrace{\mu\,E_{\text{Prior}}(\mathbf{w})}_{\text{RigSpace}} + \underbrace{\nu\,E_{\text{Sparse}}(\mathbf{w})}_{\text{WeightSpace}} \ . \tag{4.16}$$

Unlike previous approaches, our energies operate in the spaces where modifications of the input data are minimal. First, E_{Fit} is computed in the actor space, such that the incoming animation is not modified. In contrast, E_{Prior} is computed in the rig space based on the original blendshapes. Finally, in accordance with common practices in manual key-framing (Seol et al. 2011), weight activation sparsity is directly enforced on the blendshape parameter space, which simplifies any subsequent manual editing of the animation. Under the assumption that the actor's markers are saved in cm and the face rig has been uniformly scaled to roughly match the actor's head, we recommend setting $\mu = 0.3$ and $\nu = 0.6$.

The retargeting was implemented as a Maya command plugin using the Maya C++ API. We pre-compute the matrix product $\mathbf{L}\delta\mathbf{V}$ as it remains constant over time. Due to the L_1 sparsity term, the retargeting solves an Iteratively Re-Weighted Least Squares problem of size $(K \times K)$ using the Eigen library (Guennebaud et al. 2016). On average we achieve $105\,\text{fps}$, which confirms that the proposed prior is very suitable for real-time applications.

4.5 Comparison to Existing Facial Retargeting

In order to evaluate our blendshape transfer algorithm and our geometric prior, we first compare each part to common alternative formulations in a realistic but simple scenario. Doing so has the advantage that contributions can be evaluated individually and limitations of each method become clearly visible.

4.5.1 Automatic Blendshape Transfer

As described previously, RBF-deformation and deformation transfer are the most common techniques for personalizing blendshapes. We compare our algorithm to the RBF-proportion matching (Seol et al. 2012) and example-based facial rigging (Li et al. 2010). For the example-based facial rigging we select 15 distinctive facial expressions from our captured sequence as training examples. Distinctiveness is guaranteed by clustering all expressions of the training animation using k-means. The expressions closest to the cluster centers are then chosen as examples. We initialize the corresponding blendshape weights as the result of our retargeting algorithm for these specific frames. Since this method alternatingly optimizes for blendshape geometries and blendshape weights, it requires only approximated blendshapes weights at the beginning. No peak expressions are required as examples.

We retarget the input animation to both, realistic and stylized characters. For this comparison, we only use E_{Sparse} for regularization and exclude any other prior. As shown in Figure 4.12, our blendshape transfer outperforms existing approaches, in particular, if facial proportions differ significantly. In all cases, plausible results are obtained, and in nearly all cases expression intensity is restored faithfully. In contrast, the quality of the RBF-proportion matching (Seol et al. 2012) and example-based facial rigging (Li et al. 2010) degrades with a higher degree of stylization. Figure 4.13 demonstrates that other retargeting algorithms, like space-time facial cloning (Seol et al. 2012), also benefit from our more accurate personalized blendshapes.

4.5.2 Geometric Prior

Next, we compare the effectiveness of the proposed geometric prior to different well-established facial priors. To illustrate the effect of the different priors on the retargeting results, we generate personalized blendshapes using the RBF proportion matching (Seol et al. 2012), since this method yielded the most artifacts in the evaluation of Section 4.5.1. The actual retargeting is then computed as described in Section 4.4, where we replace the prior energy E_{Prior} of Equation (4.16) by several options. Note that we employ the sparsity energy E_{Sparse} for all examples.

As a simple prior we incorporate an L_2 weight regularization $E_{\text{Prior}}(\mathbf{w}) = \|\mathbf{w}\|^2$. The combination of L_2 regularization and L_1 regularization (E_{Sparse}), also known as Elastic Net (Zou & Hastie 2005), enforces small weights but is less restrictive than non-negativity constraints. This is beneficial since there exist valid facial expressions with negative weights. In addition, we compare our geometric prior to the model-specific PCA prior (Seol et al. 2012). The PCA is constructed from the given blendshapes only, in order to have equal input conditions.

For the comparison, we identify the frames with strongest artifacts and step-wise increase μ until the geometric artifacts are removed. Figure 4.14 shows the effect of each prior for one representative example. Our prior converges the fastest to an
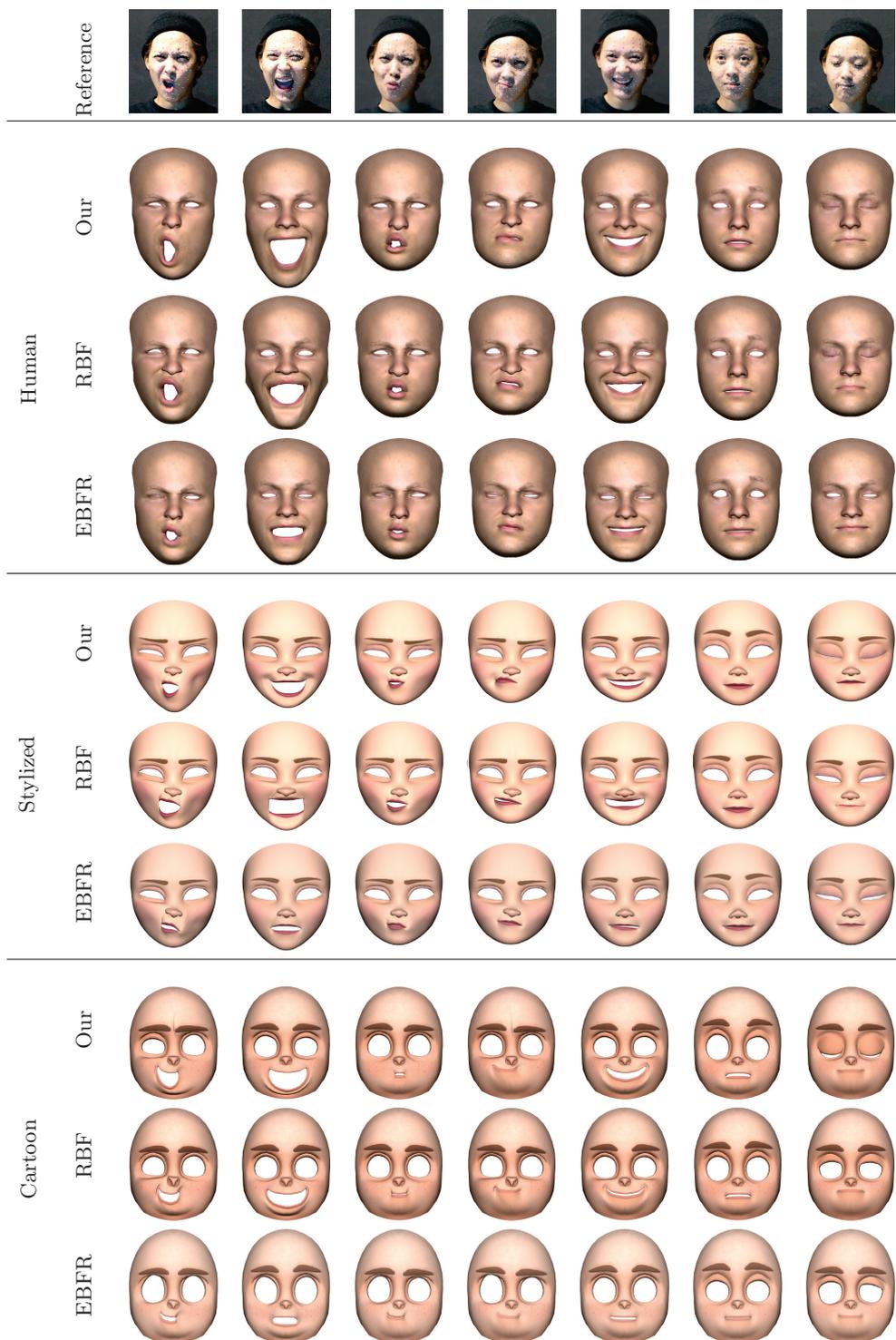
Figure 4.12: Retargeting using personalized blendshapes created by our algorithm (*Our*), RBF proportion matching (*RBF*), and example-based facial rigging (*EBFR*). In all cases, facial semantics are restored faithfully, and in particular, for stylized characters, our algorithm outperforms the other approaches.

©Motion capture: Feel Ghood Music ©Face rigs: Mark Pauly, meryproject.com, Jana Bergevin
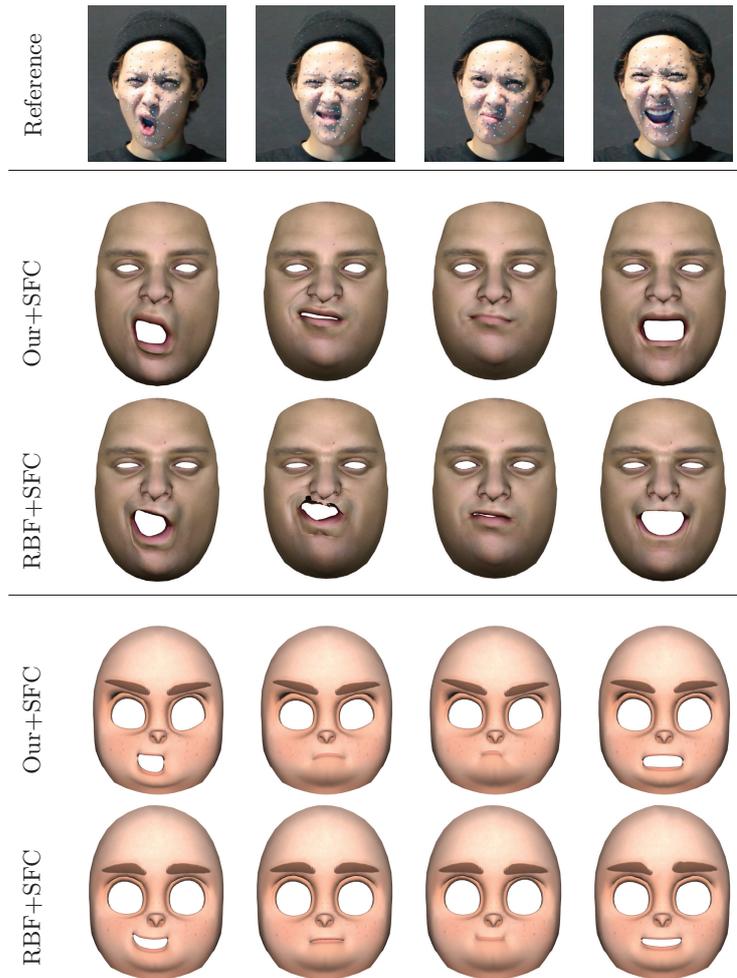
Figure 4.13: Retargeting using space-time facial cloning with our personalized blendshapes (*Our+SFC*) and with the RBF proportion matching (*RBF+SFC*). Please notice the improved expressivity and fewer artifacts for our method. ©Motion capture: Feel Ghood Music ©Face rigs: Jason Osipa, Jana Bergevin

artifact-free expression because it minimizes bending at the vertex level, without being limited to the given examples as in the PCA case. In addition, Figure 4.15 compares our real-time capable retargeting with the offline method of Seol et al. (2012).

4.5.3 Discussion and Limitations

In rare cases, we observed inaccuracies during retargeting for our algorithm, RBF-proportion matching, and example based facial rigging. While different methods might work better for specific frames, our method performs better when considering the overall sequence. A particular strength of our method is that it also works for highly stylized characters, as long as the blendshapes approximately reflect natural facial movements.

Like any data-driven approach, the performance of our method depends on the availability of good training data, which in our case is an expressive facial sequence.
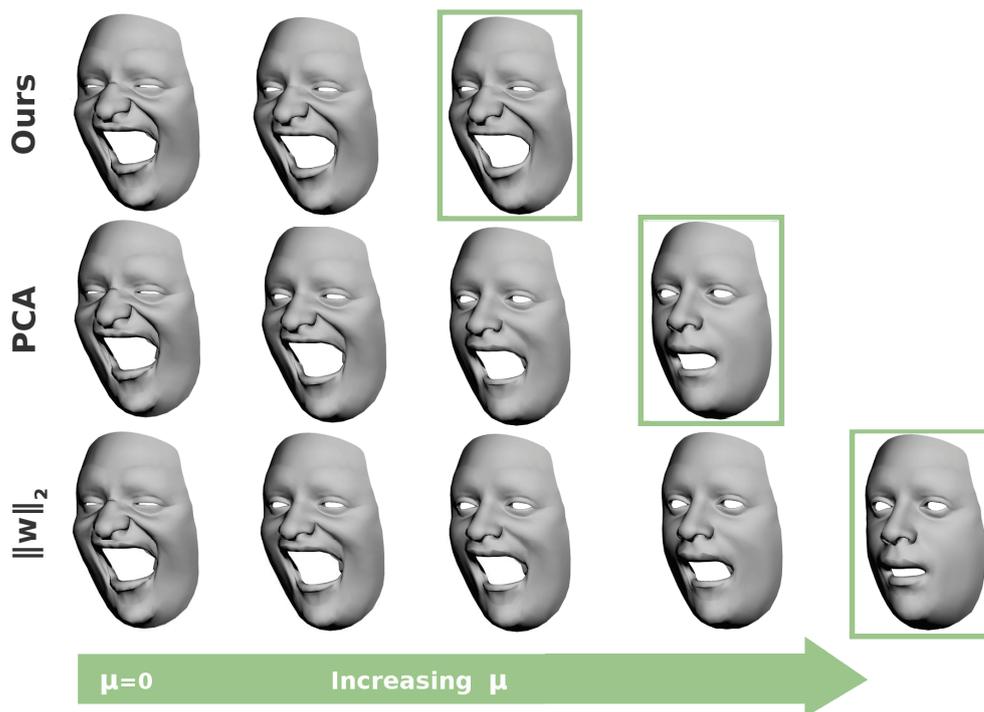
Figure 4.14: Comparison of our geometric prior (*top*), the model-specific PCA prior (*middle*) and the L_2 regularization (*bottom*). *From left to right:* Starting with a retargeting without regularization, we successively increase prior activation until a solution without self-intersections is obtained. ©Face rig: Jason Osipa
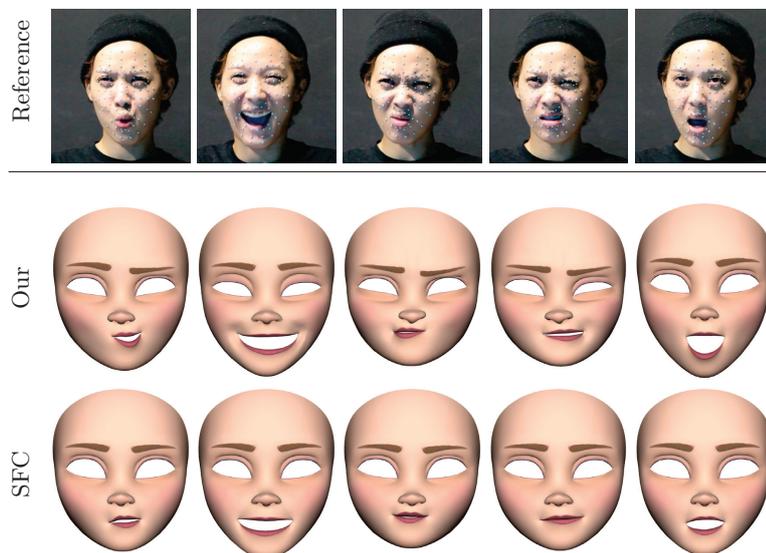
Figure 4.15: Retargeting using our personalized blendshapes in combination with the geometry prior (*Our*) and space-time facial cloning (*SFC*) for offline retargeting. Semantics are restored faithfully and the high expressivity of the original actor is maintained. Especially for stylized characters, our real-time algorithm outperforms even offline state of the art approaches.

©Motion capture: Feel Ghood Music ©Face rig: meryproject.com

However, because we do not enforce to capture isolated FACS expressions, we consider this requirement as minor: In an offline scenario, the entire animation sequence can be used for training, while for real-time applications it takes only a few seconds to create such a sequence. Our results were computed using the full animation sequence as the training sequence for a fair comparison with the offline retargeting of Seol et al. (2011). In the absence of expressive data for certain expressions, mesh constraints (Section 4.3.4) ensure plausible blendshapes that are similar to the well-established expression transfer methods, while blendshape constraints (Section 4.3.5) balance the changes occurring to similar blendshapes.

We observed that mouth expressions are highly similar (e.g., o-viseme, kiss, mouth-open), making it very difficult to identify equivalent blendshapes across different identities in the training sequence (Figure 4.16). In such cases, E_{Match} tends to under-estimate the transferred blendshapes. Although this is addressed by enhancing the contrast of the similarities to obtain a better fit to the actor's expressions (Equations (4.7), (4.8)), the adjusted blendshapes might be more subtle than the actor's expressions, resulting in slightly exaggerated expressions (Figure 4.16, bottom) at the retargeting step. However, the user can always return to the default RBF-based deformation transfer using only E_{Mesh} and E_{CEG} (Figure 4.17).

While our approach shows convincing results on the tested blendshape rigs and animations, our alignment might not reproduce the desired result if blendshapes model a very cartoony behavior (e.g., popping eyes) or vary in the number of facial features (e.g., different number of eyes). Cross-mapping approaches to facial animation retargeting are better suited for this type of animation. Our method does not address the transfer of fine-scale details. In practice, it is not often desired, as fine-scale details are already encoded in the facial rig and should remain consistent when the animation is transferred from different actors.

By replacing the automatically computed correlations with manually selected weights for specific examples, our method can easily be extended to a semi-supervised method. Because our similarity measure is limited to the range $[0, 1]$, which is equivalent to the recommended weight space for blendshapes, such an extension would be straightforward.
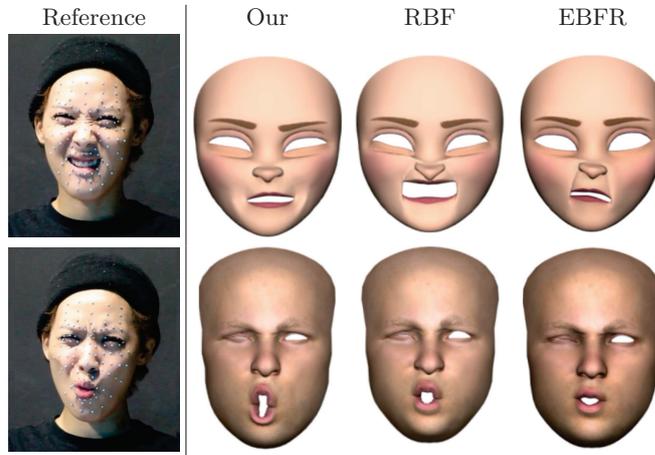
Figure 4.16: Semantic changes for subtle expressions appear for all approaches. Our method allows the user to seamlessly blend between our and RBF blendshape transfer by increasing the weights of E_{Mesh} and E_{CEG}.

ⒸMotion capture: Feel Ghood Music ⒸFace rigs: meryproject.com, Mark Pauly.
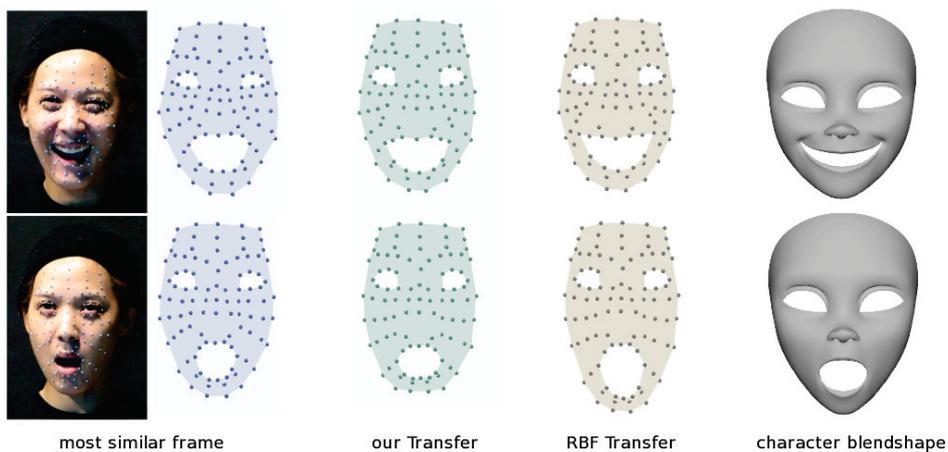
Figure 4.17: Sparse blendshape comparison of the initial guess (RBF transfer) and our blendshape transfer method. Sparse blendshapes of the character rig (*right*) are adjusted to match the actor's proportions (*left*).

ⒸMotion capture: Feel Ghood Music ⒸFace rigs: meryproject.com

Part II

Perception of Realistic and Stylized Faces

Over the last years, we are facing an increasing presence of robots and computer animated characters in daily life. However, the impact of many design decisions is still not fully understood. We, therefore, investigate the impact of aspects like shape, material, shading and lighting on the perceived appeal and realism of stylized faces. In addition, we conduct an experiment using event-related potentials (ERPs) to explore the processing of stylized stimuli by the brain. To this end, we create stimuli consisting of several stylization levels of two real people (Chapter 5).

In total, we conduct four experiments using two different experimental designs. In Experiments 1a–1c (Section 6.3) we investigate step-by-step the influence of shape and material on the overall appearance of a virtual character using rating scales. For validation of the obtained results, we significantly extend the number of stimuli for Experiment 2 (Section 6.4) and focus in Experiment 3 (Section 6.5) on the influence of different expressions across all tested stylization levels. The experiment design is motivated by differing design choices of recent animation films, ranging from cartoon shapes with cartoon materials (e.g., *Despicable Me*), over stylized shapes with realistic material (e.g., *The Adventures of Tintin*), up to very realistic shapes and material (e.g., *Beowulf*).

As rating scale experiments reflect conscious decisions of participants, we consider in Experiment 4 (Chapter 7) an alternative procedure using ERPs, which unveils the brain processes involved when looking at stylized characters. In addition, this experiment will also provide information whether computer-generated characters are suitable for replacing photographs of real people in face perception research in psychology (de Borst & de Gelder 2015). So far, little is known about how stylized faces are processed by the brain and which facial details are important in particular. Even less is known about how a presented emotional expression interacts with the stylization of a face. Due to the small amount of comparable work using ERP, only the most distinctive stimuli were chosen for this experiment. The stimuli consist of two characters, five stylization levels in addition to photographs and three facial expressions: anger, happy, neutral.

5 Stimuli Creation

Figure 5.1: Our initial stimuli set consists of two identities, five stylization levels and five basic expressions – from the most abstract (*left*) to the most realistic versions (*right*).

Our experiment setup requires virtual doubles of real people together with stylized versions of the same characters. The initial Experiments 1a–1c (Section 6.3) required the design of three levels of stylization of the same character. This has been later extended to five stylization levels (Figure 5.1) for Experiments 2 and 3 (Sections 6.4 and 6.5) and the ERP Experiment (Chapter 7). The realistic characters are based on real people without ethnic bias to the group of participants. For the acquisition and creation of the virtual doubles, we relied on the character replication pipeline as described in detail in Chapter 2. In addition to the neutral expression, we captured for each character four of the universal facial emotions: anger, happiness, sadness, and surprise (Ekman 1972). 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 the original description of the basic emotions (Ekman 1972). We discarded disgust and fear because their status as basic expressions was questioned recently (Jack et al. 2014) and they are harder to identify by observers.

While a realistic character can be obtained from 3D scans of a real person, no automatic solution exists to generate increasingly stylized versions. Present stylization algorithms are limited in most cases to 2D images as inputs, where stylization is achieved using image processing techniques (for an overview see, e.g., Kyprianidis et al. (2013)) or by creating corresponding pairs of matching facial parts (Zhang et al. 2014). An exception for the stylization of 3D models is the work of Jachnik et al. (2015). However, in all cases, the stylization capabilities remain too limited for our purpose.

We, therefore, asked different professional 3D artists to produce the required stylized shapes and materials based on our realistic characters, taking inspiration from commercial animation films (see examples in Figure 5.2). For our first set of experiments (Section 6.3), we used three stylization levels for shape and material. For the later experiments (Section 6.4) two additional stylization levels have been created, such that the realism scale is sampled accordingly. For all stylized characters, 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. As the instructed artists worked indepen-
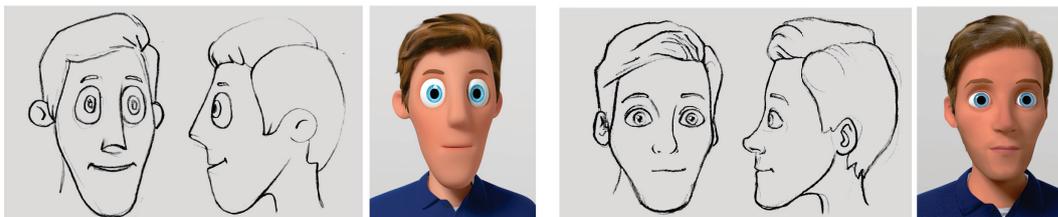
Figure 5.2: Two of the stylizations created for the study, showing the original sketches on the left and their resulting stylized 3D models on the right. The designs are inspired by the films *Cloudy with a Chance of Meatballs* (*top*) and *Toy Story* (*bottom*).

dently of each other, some of the stylized characters required further adjustments to match better our requirements: (i) similarity with the original people, (ii) roughly linear sampling of the realism scale and (iii) consistent expressions. Furthermore, we created the hairstyles and clothing for the different stylization levels.

All stimuli for the following perception experiments have been created based on the initial set of two characters, five stylization levels in addition to photographs and five basic expressions (Figure 5.1). Overall, 406 distinct images were rendered. While in some cases the required changes were small, e.g., changing the camera perspective, other changes have been very comprehensive. Transferring, e.g., the material of the realistic characters to the stylized version, involved not only the transfer of the albedo, specular and bump textures but also of textures for the eyes, teeth and clothes. In other cases, apparently simple modifications revealed problematic setups and required re-adjustments of specific parameters for all affected stimuli, e.g., changing from global illumination lighting to simple, direct illumination caused rendering artifacts due to the selected material parameters. For rendering all our realistic and stylized stimuli, we used the same rendering pipeline and very similar setups as described in Chapter 2. The entire dataset is available online[1] for research purposes.

[1]http://graphics.uni-bielefeld.de/publications/sigasia2015/

6 Rating Scales Experiments
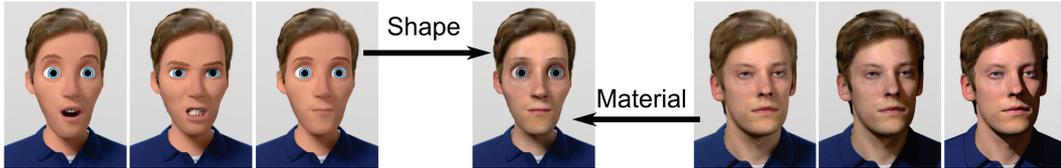
Figure 6.1: We test the influence of expression (*left*), shape and material stylization (*center*) and shading and lighting (*right*) on facial perception.

In our work, we focus 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 facial expressions: anger, happiness, sadness, surprise, and neutral. We then create all combinations along these dimensions (shape, material, expression, Figure 6.1) 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.

The design of our rating scales experiments 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 photo-realistic materials in *The Adventures of Tintin*. We use static pictures as stimuli, as it has been found that much of the information that people use to evaluate virtual characters are available in a still image (McDonnell et al. 2012).

Contribution   This work was created under the supervision of Mario Botsch (Bielefeld University), Diego Gutierrez (Universidad de Zaragoza) and Rachel McDonnell (Trinity College Dublin). In this project I was the primary investigator and contributed to various aspects of the project, starting from the initial idea up to writing up the final paper. My biggest work packages were carrying out and analyzing the perceptual experiments and the creation of the stimuli. The latter included scanning the actors and post-processing the data, creating hairstyles, adjusting the stylized characters, and setting up and rendering all variations of the stimuli. I received support from Katja Zibrek with analyzing the experiments, and Carlos Aliaga with adjusting and setting up the initial character set. Carlos Aliaga also created the most abstract stylization level.

6.1   Overview of Facial Perception Experiments Using Rating Scales

Stylization   Some of the first attempts to measure the likability of stylized and realistic characters were 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 & Nagayama 2007, Burleigh et al. 2013). Green et al. (2008) concluded that there is less tolerance to deviations from original proportions in cases where faces are more attractive and human-like. Others focused on the boundary when characters are perceived as real, by morphing between photographs and puppet faces (Looser & Wheatley 2010) or virtual faces (Cheetham et al. 2011). 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 reducing photorealism 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 et al. (2012) found that rendering style affects the appeal and trustworthiness of the characters.

(a) Rough   (b) Environment   (c) Chrome   (d) Environment
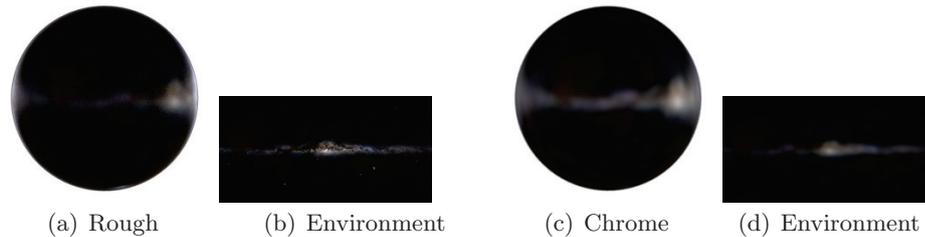
Figure 6.2: A special case demonstrating the ambiguity of light and material. *Left:* A sphere with a rough surface lit by a star environment map. *Right:* A chromatic sphere lit by a blurred environment map. Despite different materials and lighting, the appearance of both spheres is nearly identical.

Additionally, a character rendered in an appealing style can be perceived to have more positive personality traits (Zibrek & 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 occur 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 effects. 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 concerning 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 & Cunningham 1991). Several studies confirmed that texture changes do result in a significantly more attractive face (Benson & Perrett 1992, Little & Hancock 2002). Publications in the cosmetics domain also help explain the observed effects on appeal: Fink et al. (2006) created textures from photographs of 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 & 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 & Rose 2011)).

Lighting and Shading Several researchers have studied material perception beyond the skin. We focus at this point on glossy, diffuse and translucent surfaces and omit transparent materials. Studying the perception of materials is challenging due to the strong interaction with lighting condition, making the reasons for the visual appearance ambiguous (Figure 6.2). Consider e.g. a perfectly polished chrome

ball within a closed box covered with velvet. In the presence of indirect light, the ball would mirror the velvet surface, making it impossible to distinguish it from the velvet material of the box. It is therefore not surprising that participants inconsistently approximate parameters for glossy surfaces, especially in case of unnatural lighting (Fleming et al. 2003). Also, accuracy on identifying equal materials (Pont & te Pas 2006) or determination of roughness (Ho et al. 2006) vary for different lighting setups. Besides light even the shape of an object influences the perception of glossiness (Vangorp et al. 2007, Vangorp 2009, Olkkonen & Brainard 2010, 2011). In case of translucent materials, the lighting direction has a fundamental impact on perceived translucency. While frontally lit translucent objects lack many visual cues (e.g., blurred features, soft shadows, low contrast), these features are enhanced when illuminated from the back (Fleming & Bülthoff 2005).

Rather than focusing on accuracy in material perception, the question remains what makes surfaces and light look realistic. Based on real photographs, Rademacher et al. (2001) identified that surface smoothness and shadow softness affect realism, but not the number of objects and lights. However, the effect size was bigger for surface smoothness than for light (Rademacher 2002). The fact that the human visual system is tolerant to inaccuracies in lighting or shading was considered to speed-up rendering, e.g., approximating indirect light between frames through spherical harmonics (Jarabo et al. 2012). Another well-established example is the replacement of the computationally intensive inter-reflections between surfaces through several simpler light sources (Keller 1997). Respective perception parameters have been systematically studied in Křivánek et al. (2010). Finally, Wisessing et al. (2016) measured the impact of render style and lighting on the intensity and appeal of expressions in short animation sequences. Lighting had no or very little influence on perceived intensity, but soft shadows have been rated as more appealing, even though the effect size was small.

Expression   In our experiments, 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 looking at 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, Grimshaw, Nabi & Williams 2011). Additionally, given different hypotheses that iconic representation of faces increases 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.

6.2   General Experiment Design

Since both the design and the analysis of our experiments in this section as well as in Sections 6.4 and 6.5 share many similarities, we describe the general setup now and later only mention deviations. Following previous work (e.g., (McDonnell et al. 2012, Ho & 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 it is gloomy and leaves 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 have not seen anything like before.

- *Extremely unattractive—Extremely attractive*: Indicate if you find the character unattractive and ugly or beautiful and attractive.

We modeled 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 have chosen 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.

The user's task and the rating scales were explained in a written document to the participants before the experiment. Afterwards, all stimuli were presented in a random order and shown for 3 s 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 cm $\times$ 20.0 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 chapter.

For statistical analysis of each rating scale, we conducted an n-way repeated measures Analysis Of Variance (rm-ANOVA). We run Mauchly'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 explore the results further (Cunningham & Wallraven 2011).

6.3 Experiment 1: Importance of Shape, Material and Lighting

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 having a potentially different influence on appearance. This makes the experiment design an extremely challenging 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. 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.

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 6.3). The inter-surface mapping for the texture transfer was computed based on a dense correspondence map established using the non-rigid registration technique from Chapter 3.

6.3.1 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 6.3). 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. Figures 6.4 and 6.5 compare 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.

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 6.4b), 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.
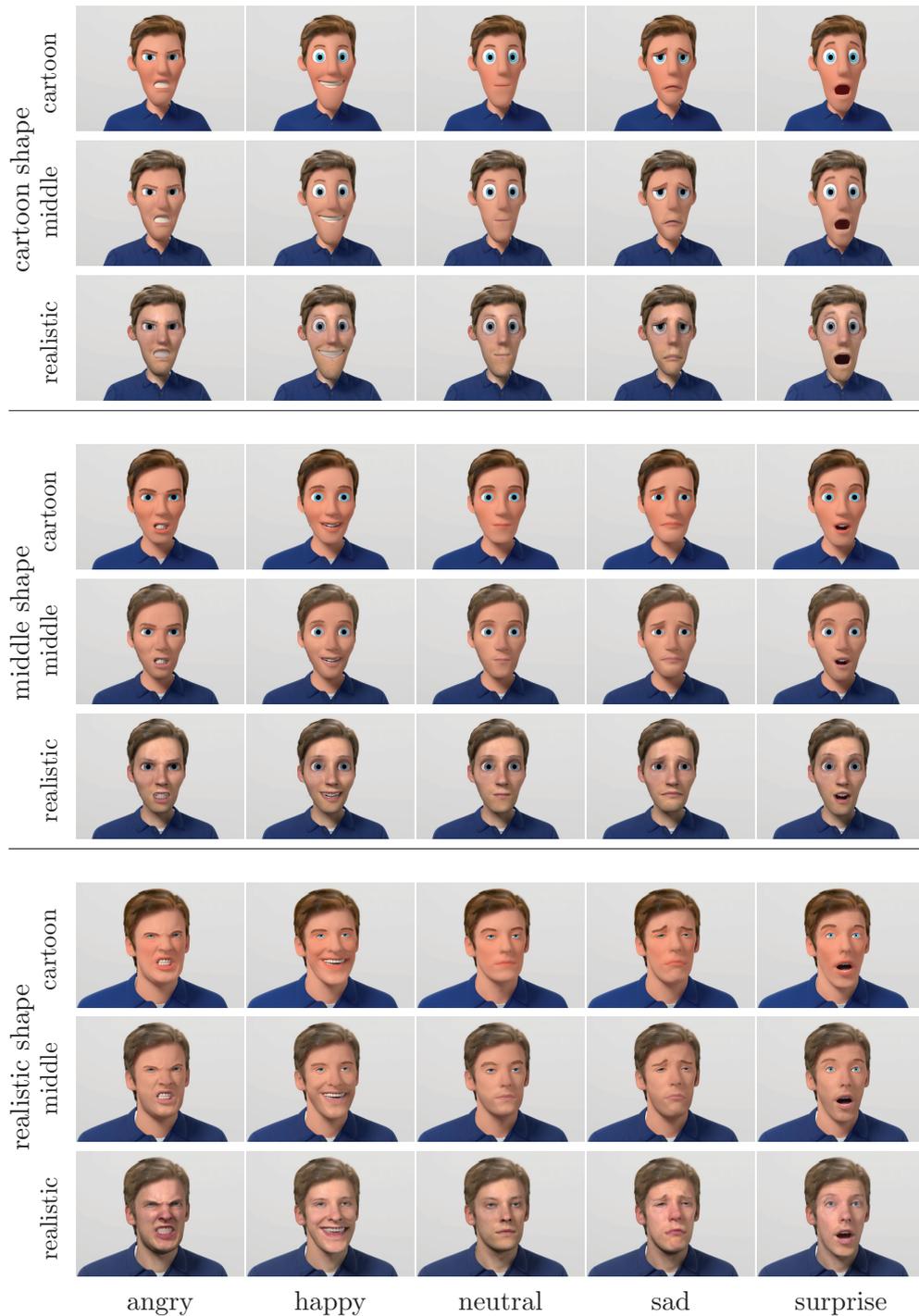
Figure 6.3: Stimuli used in Experiment 1a: three levels of shape and material stylization (*y-axis*) and five expressions (*x-axis*). Textures of characters with matching material and shape stylization level have been transferred to the other shapes for testing stylization levels of shape and material independently.

(a) Realism – neutral

(b) Realism – all

(c) Familiarity – neutral
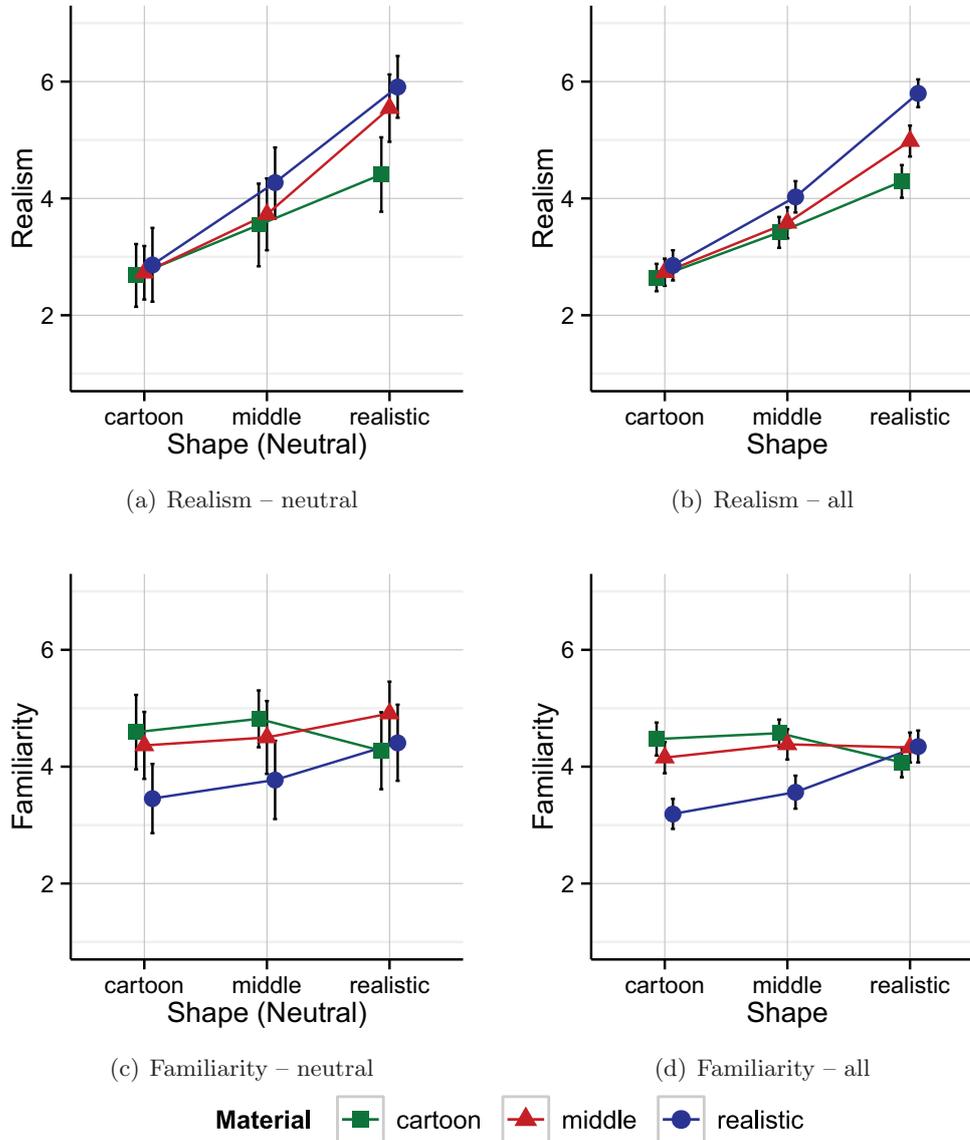
(d) Familiarity – all

Figure 6.4: Results of Experiment 1a: Ratings for perceived realism and familiarity, for different shape and material stylizations. (a) and (c) show results for the neutral expression only, (b) and (d) are averages over all expressions. Errorbars denote 95% confidence levels. Individual per-expression results are discussed in Section 6.5.

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 6.4d).

Appeal We found a main effect of material for 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 6.5b) 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 in shape-material combinations that reduce appeal as well, which have been 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 6.5b, d).

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

In an experiment similar to the previous one, we tested the initial baseline characters (three matching shape/material stylization, five expressions) with two different

(a) Appeal – neutral

(b) Appeal – all

(c) Reassurance – neutral
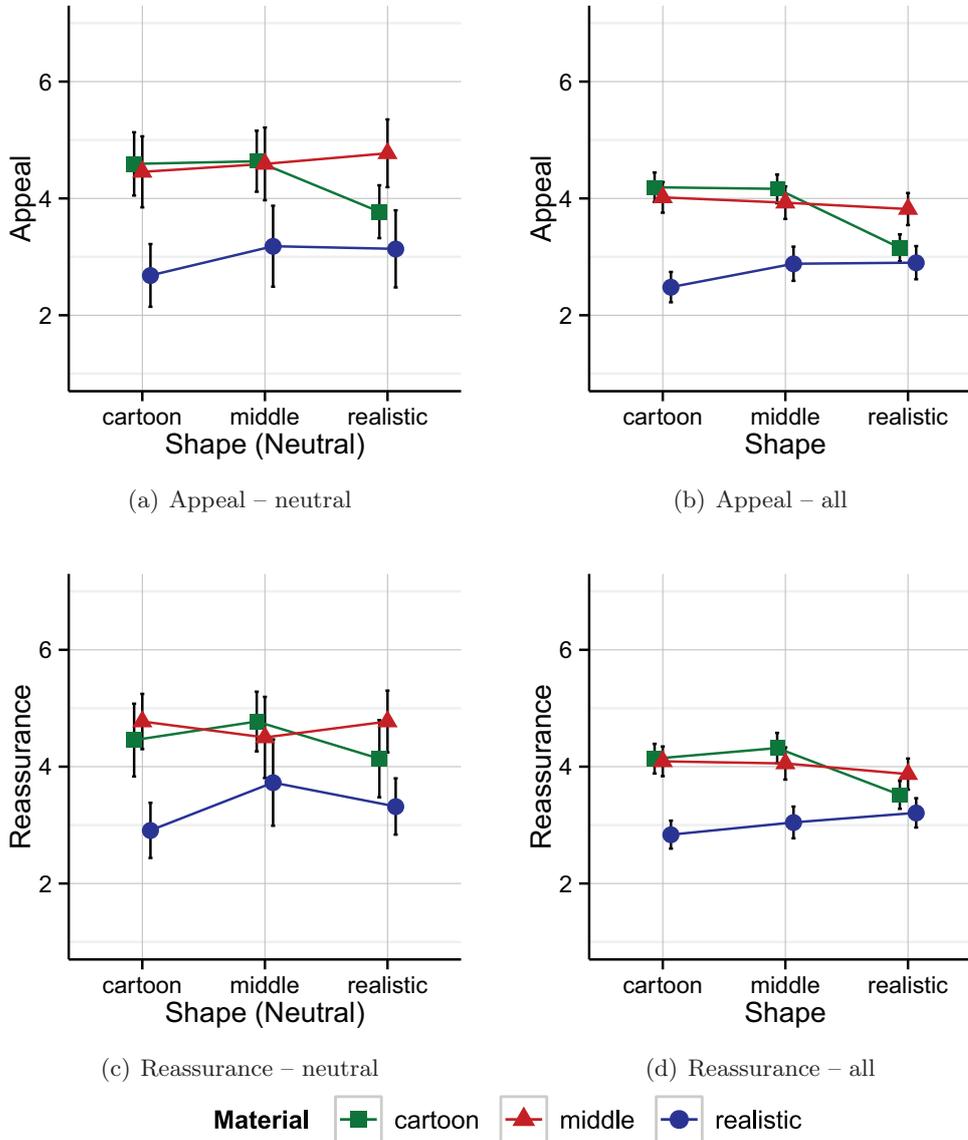
(d) Reassurance – all

Figure 6.5: Results of Experiment 1a: Ratings for perceived appeal and reassurance/eeriness for different shape and material stylizations. (a) and (c) show results for the neutral expression only, (b) and (d) are averages over all expressions. Errorbars denote 95% confidence levels. Individual per-expression results are discussed in Section 6.5.
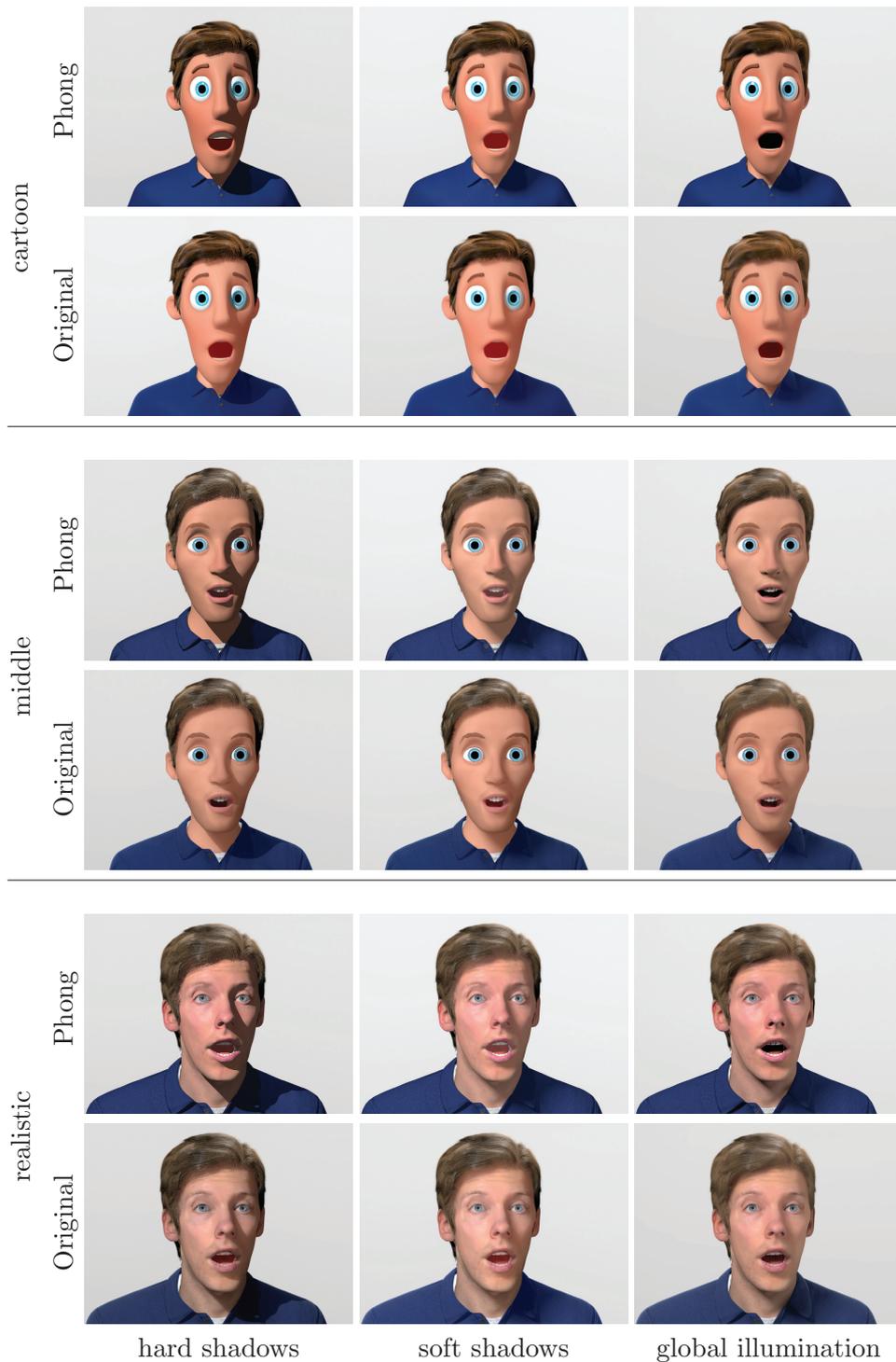
Figure 6.6: Stimuli for Experiment 1b: Three stylization levels of the same identity showing the surprised expression. Each character was rendered with the simple Phong material (*upper row*) and original sophisticated shaders for skin, clothes, etc. (*lower row*) under three lighting setups: Direct light with hard shadows, direct light with soft shadows and global illumination with soft shadows (*from left to right*).
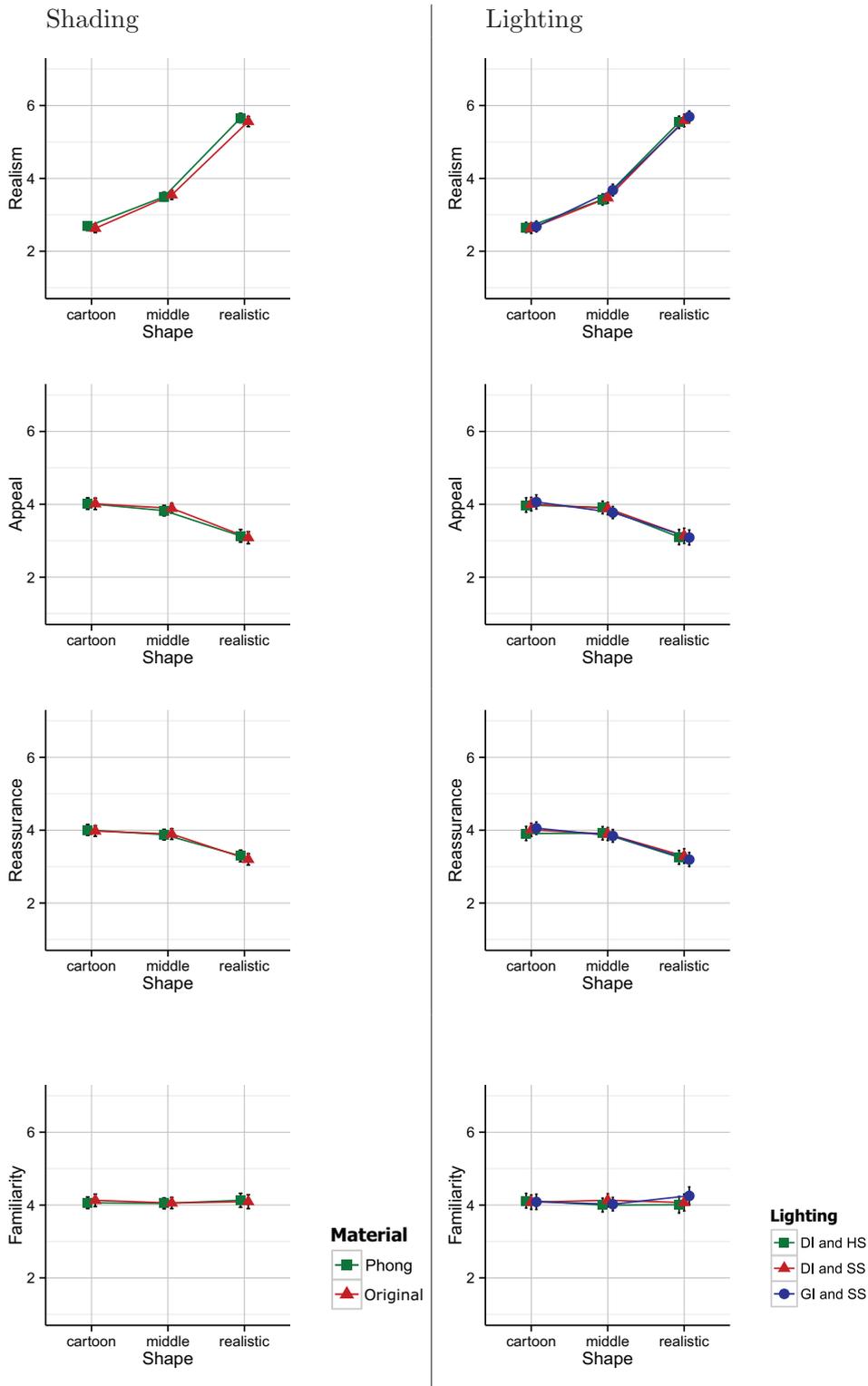
Figure 6.7: Results for Experiment 1b: Comparison of ratings for perceived realism, appeal, reassurance and familiarity for different shading and lighting setups. Error-bars denote 95% confidence levels. Meaning of the acronyms: DI–Direct Illumination, GI–Global Illumination, HS–Hard Shadows and SS–Soft Shadows

shaders and three illumination methods. For shading, we tested a simple Phong shader in addition to the original sophisticated shaders for skin, cloth, etc. The lighting categories were (i) global illumination and soft shadows, (ii) direct light and soft shadows, (iii) direct light and hard shadows. One key light together with a rim light illuminated the scene. Instead of using a fill light, a concave background reflected light in case of the global illumination setting and a low amount of constant ambient light was added for the direct light setups. 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 with soft shadows being rated more realistic than ambient light with soft shadows ($p = 0.020$) and hard shadows ($p = 0.004$), the difference was minimal (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 (Figure 6.7). These results suggest that textures have more influence than shader parameters on appearance, and therefore we explore them more in depth in the following.

6.3.3 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 & Matts (2008). In order to analyze whether their findings on attractiveness can also explain our effects on appeal and reassurance, we designed a variation of Experiment 1a from Section 6.3.1. Our remaining evaluation experiment should then:

- test whether it is possible to influence appeal or realism by changing only the albedo texture,

- show a possible correlation between appeal, reassurance and attractiveness,

- 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 five materials, which were also transferred to the middle and cartoon shapes, as shown in Figure 6.8.

For this experiment we tested these five materials on the three 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
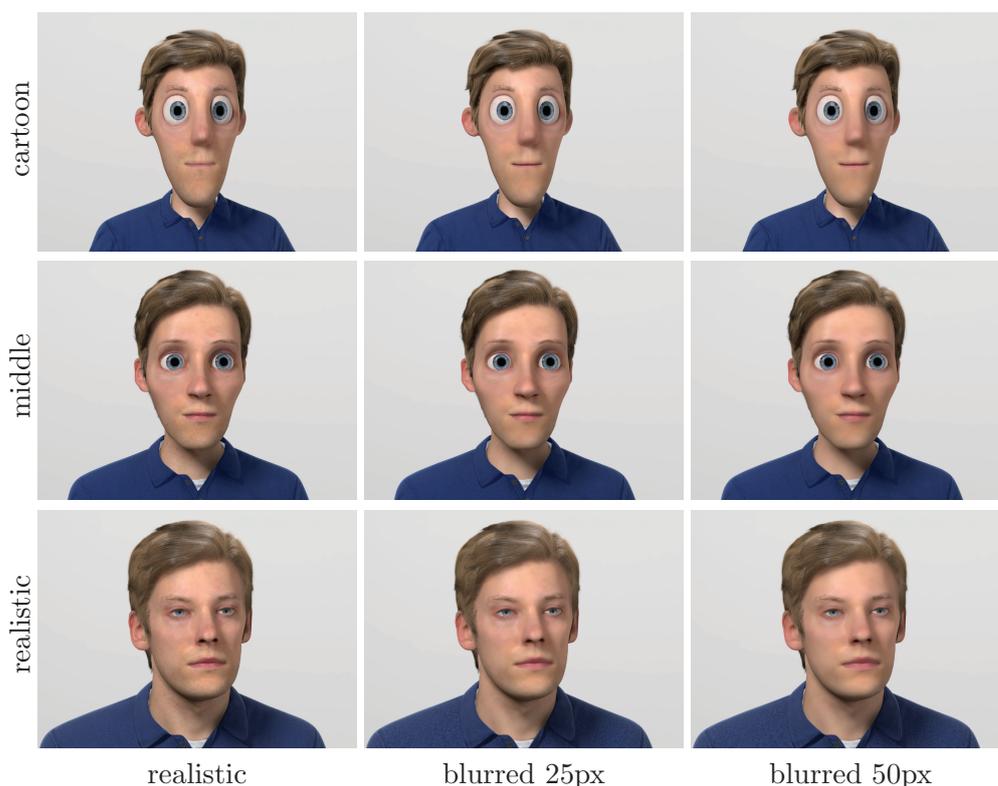
Figure 6.8: Stimuli for Experiment 1c: Realistic material with a realistic texture and two variants with blurred textures (Gaussian kernels of 25 and 50 pixels), for the three shape stylizations.

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 6.3.1 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 6.9 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 & Matts (2008): We generalize their findings to character shapes of different stylization levels.

Reassurance Although the graphs of reassurance and appeal are similar (Figure 6.9; $\alpha = 0.89$), 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.

6.3.4 First Conclusions on the Importance of Shape, Material and Lighting

The three experiments 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 negatively affect the appeal and attractiveness of the characters and make them eerier.

- Texture has a stronger influence on appeal and attractiveness than shading or illumination models. Blurring a realistic texture does not significantly reduce realism but increases 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).

(a) Realism

(b) Appeal

(c) Attractiveness

(d) Reassurance

Material ■ cartoon ▲ middle ● realistic ◆ 25px blur ✕ 50px blur
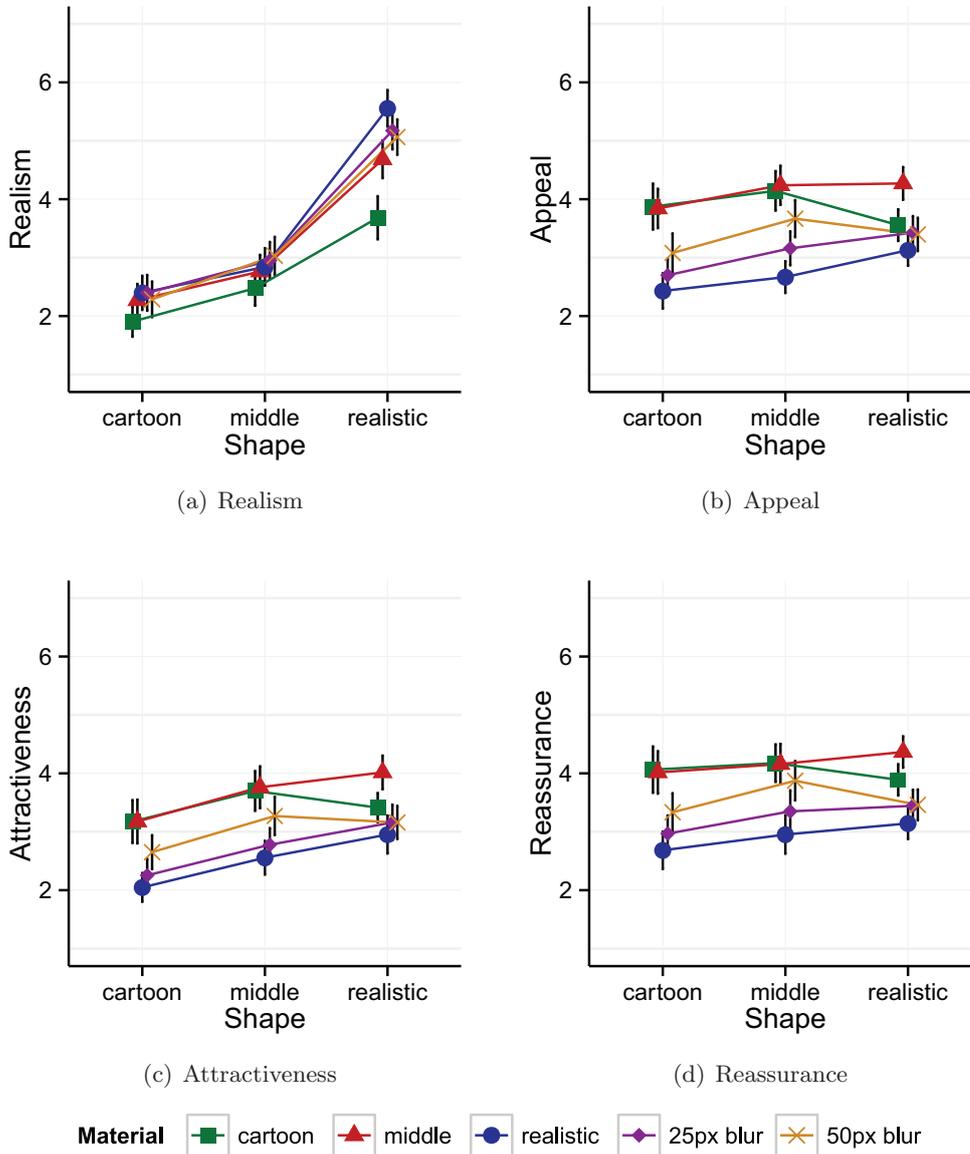
Figure 6.9: Results of Experiment 1c: While there is nearly no difference between the realistic and blurred textures for the realism scale, the blurred textures increase appeal and attractiveness and reduce eeriness.

6.4   Experiment 2: Further Investigation of Shape and Material

The experiments in Section 6.3 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 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 two characters times five shape stylizations times five material levels times five expressions, leading to a total of 250 images. Figure 6.10 shows five expressions and matching shape/material levels of the male character, while Figure 6.11 shows the 25 combinations of material and shape for the female character. Both figures visualize a representative subset of the stimuli.

Procedure   The largely extended stimuli require a reduction of the scales 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 a 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 volunteers (17 female, 4 male) took part, average age 23.4 years.

91
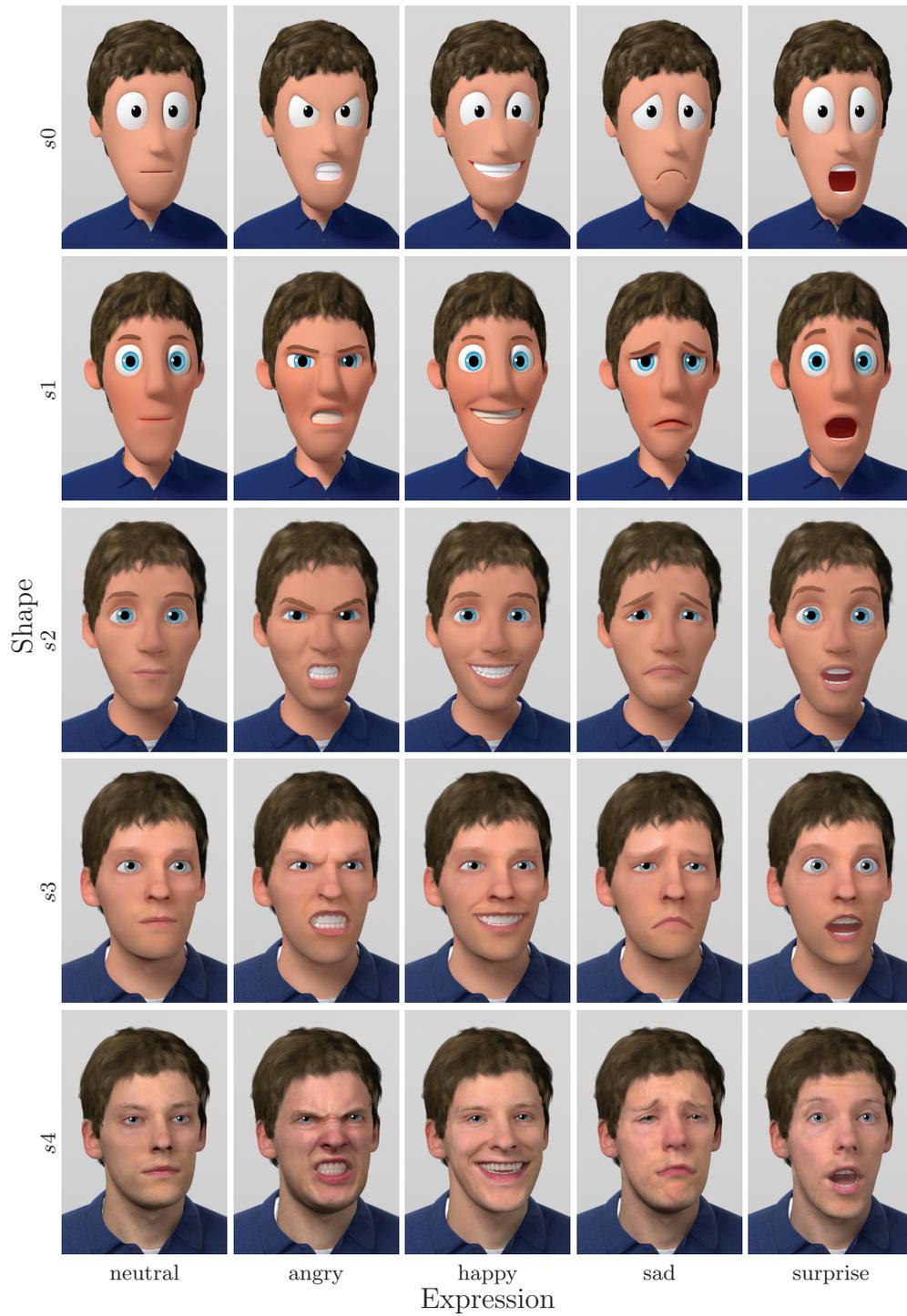
Figure 6.10: Stimuli for Experiment 2: Renderings of the male character for different stylizations (rows) and basic emotions (columns).
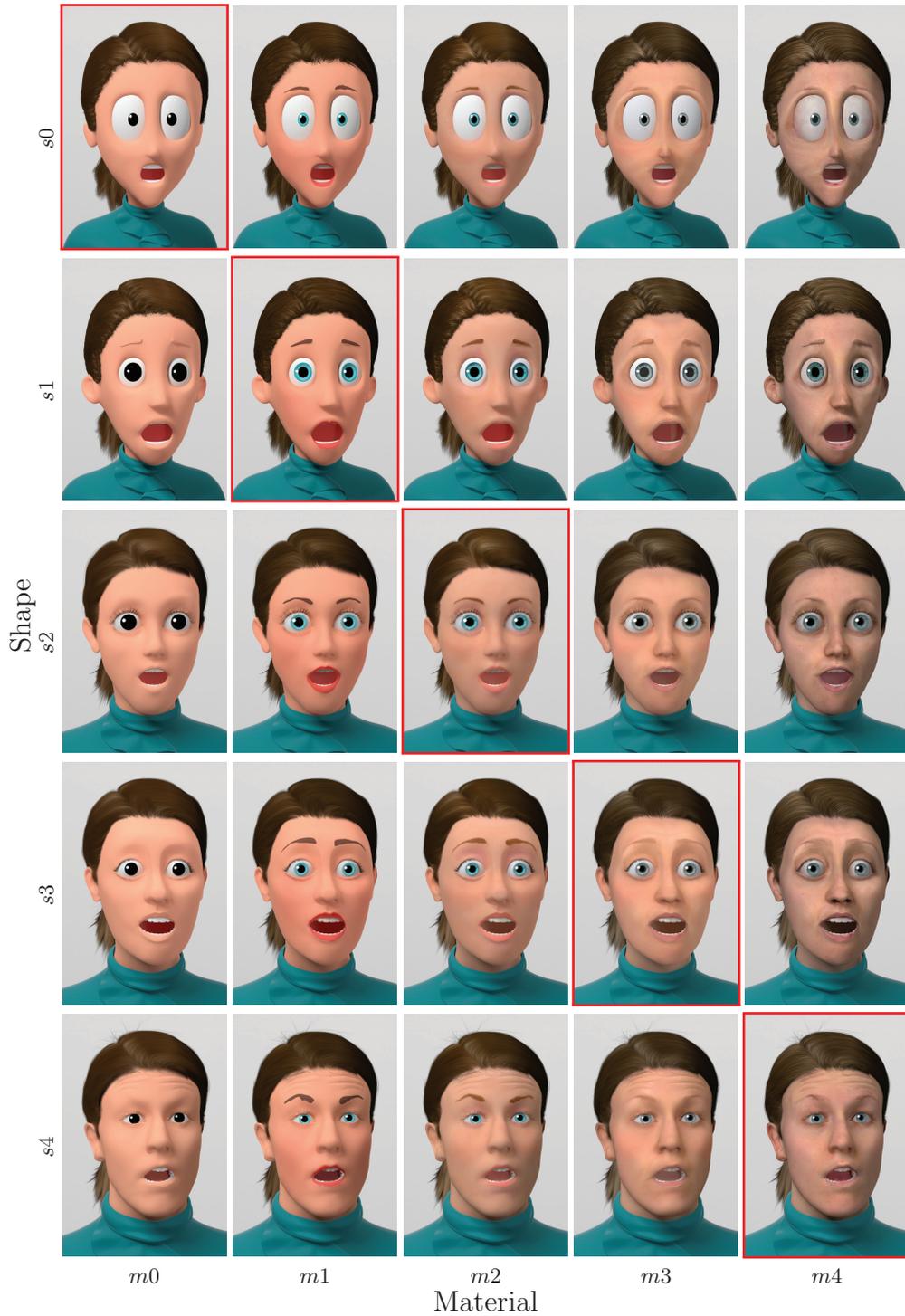
Figure 6.11: Stimuli for Experiment 2: Combinations of shape and material stylization for the female character (surprise expression), with baseline stimuli on the diagonal.

Our results are summarized in Figures 6.12 and 6.13 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 6.5.

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 also differ significantly in more than 80% of the cases. Most non-significant comparisons can be found for the shape level $s0$ (see Figure 6.12). 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 6.3.1, 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 6.3.3 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 substantially 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 actors in a neutral pose. As expected, the average realism rating is very high (6.98, $SD = 0.15$). The average appeal rating of both actors 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 6.16 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.
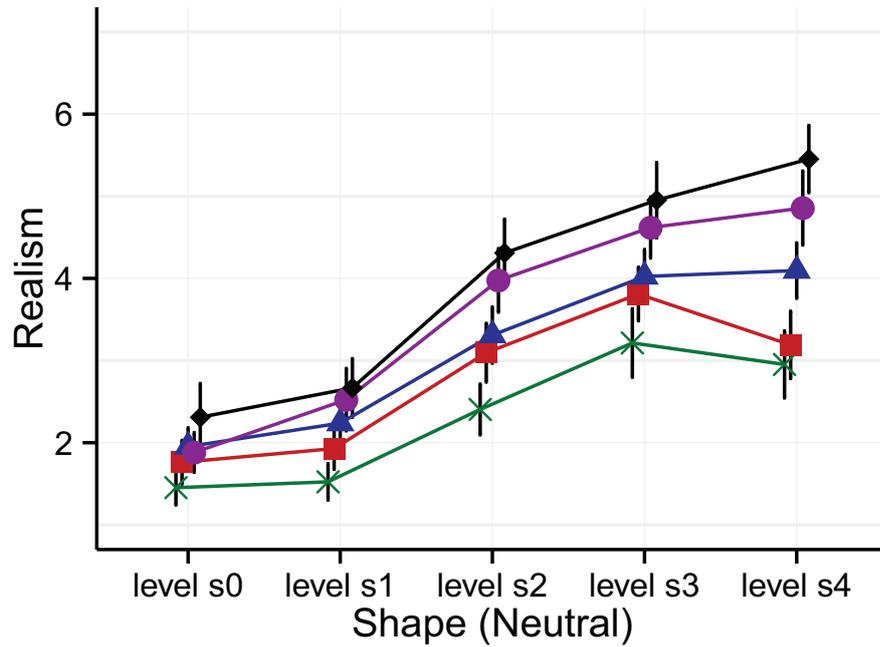
(a) Realism - neutral
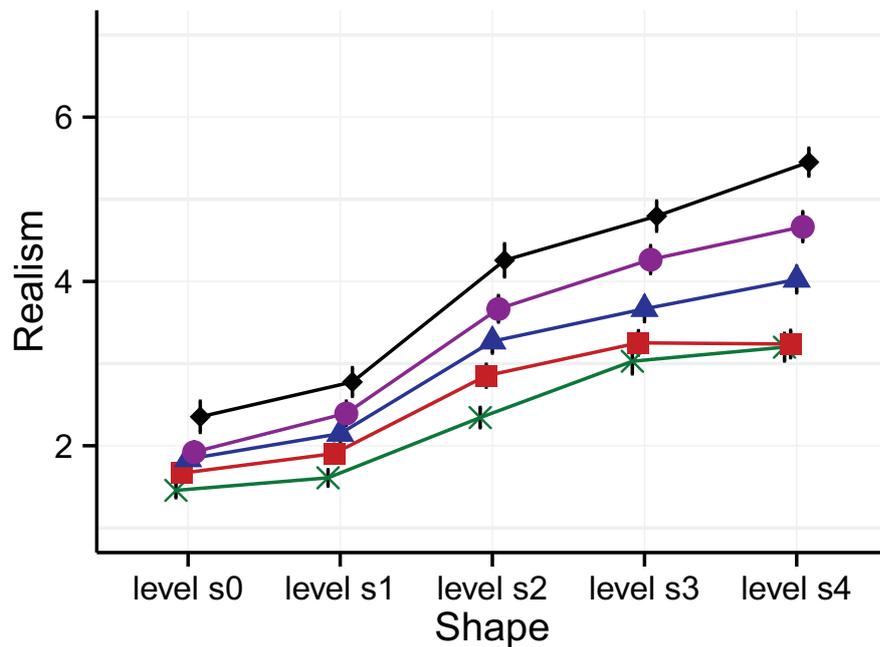
(b) Realism - all

Figure 6.12: Results of Experiment 2: Ratings for perceived realism for different shape and material stylizations. *Upper row*: neutral expression averaged over male and female characters. *Bottom row*: averaged over all expressions and characters.
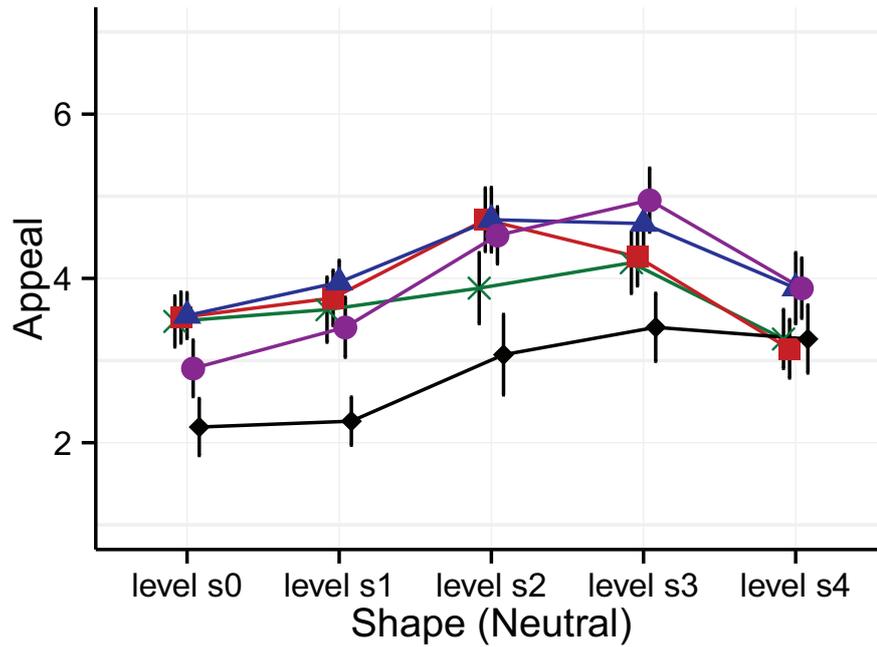
(a) Appeal - neutral
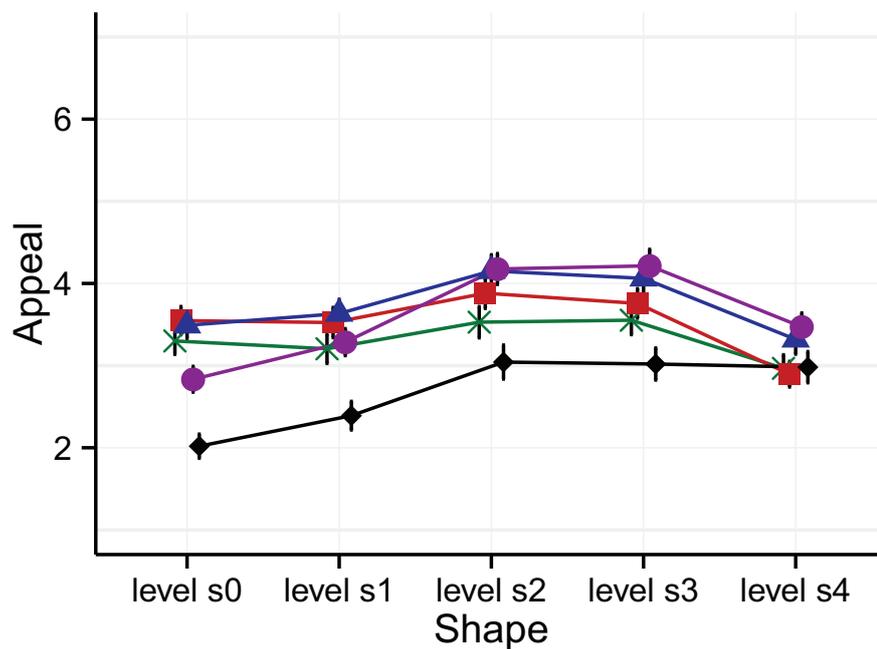
(b) Appeal - all

Figure 6.13: Results of Experiment 2: Ratings for perceived 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.

6.5   Experiment 3: Effect of Expressions

In the 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 6.11). This is interesting since previous findings suggest that valence of emotion affects character perception (Calder 1996, Tinwell, Grimshaw, Nabi & Williams 2011), 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?

6.5.1   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 intensity* 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, 23.6 years old on average) took part in this experiment. Results are shown in Figure 6.14 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 ($F^*(4.56, 104.9) = 41.3$, $p < 0.0001$, $\epsilon = 0.285$), material ($F(16, 368) = 4.97$, $p < 0.0001$) and character ($F^*(2.3, 51.26) = 4.23$, $p = 0.016$, $\epsilon = 0.557$). 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 expression. 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%.
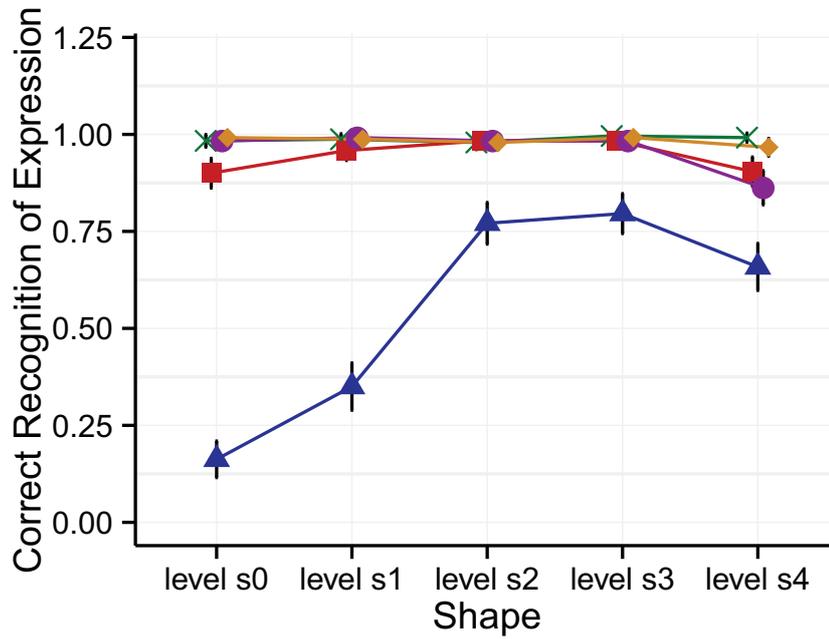
(a) Recognition
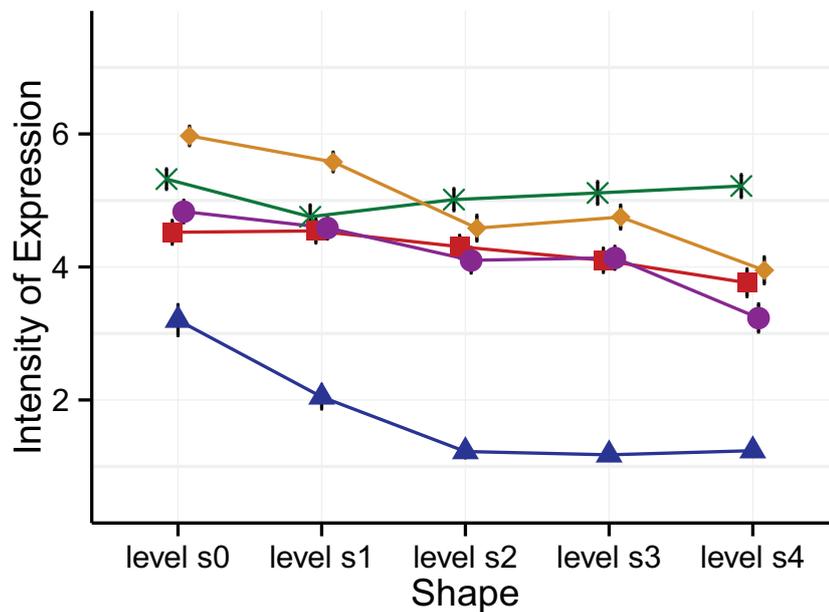
(b) Intensity

Figure 6.14: 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.

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 for shape levels $s2$ and $s3$ the intensity remains 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, 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 character ($F(4, 92) = 19.55$, $p < 0.0001$) 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 $s4$ ($p < 0.01$), 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 adequately stylizing features helps increase expressivity. The neutral expression is hard to read for very stylized character shapes, suggesting that low-intensity subtle expressions are harder to convey in abstract characters designed to enhance expressivity. 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.

6.5.2   Effect of Expressions on Realism and Appeal

In our previous experiments on material and shape with the five basic expressions (Section 6.3), 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. We omit details on the effects of expressions for Experiments 1a and 1b (Appendix A.1.1) as they do not provide additional insights. Figure 6.15 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 ($p < 0.006$). As 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 ($p < 0.001$). Overall, 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.
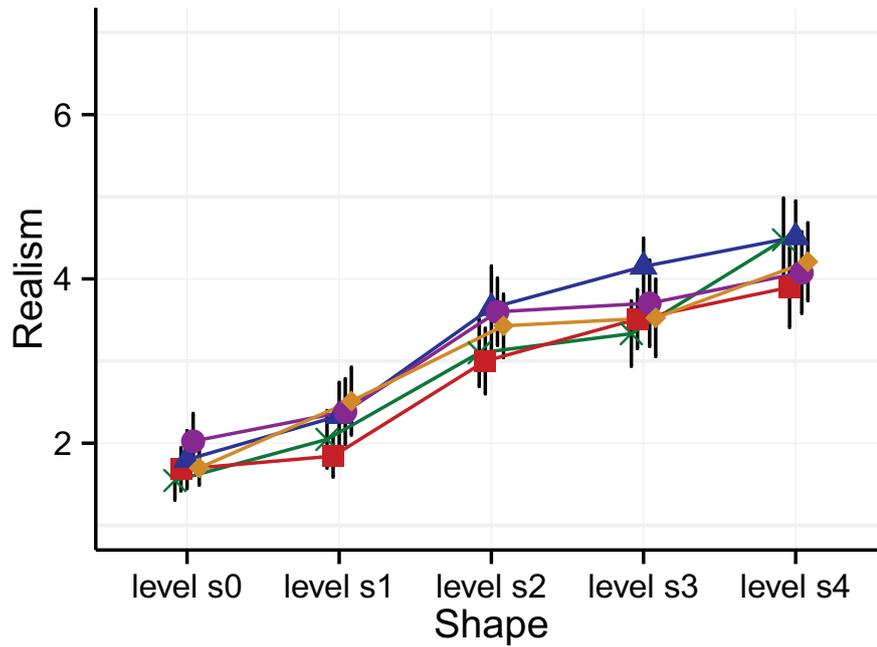
(a) Realism
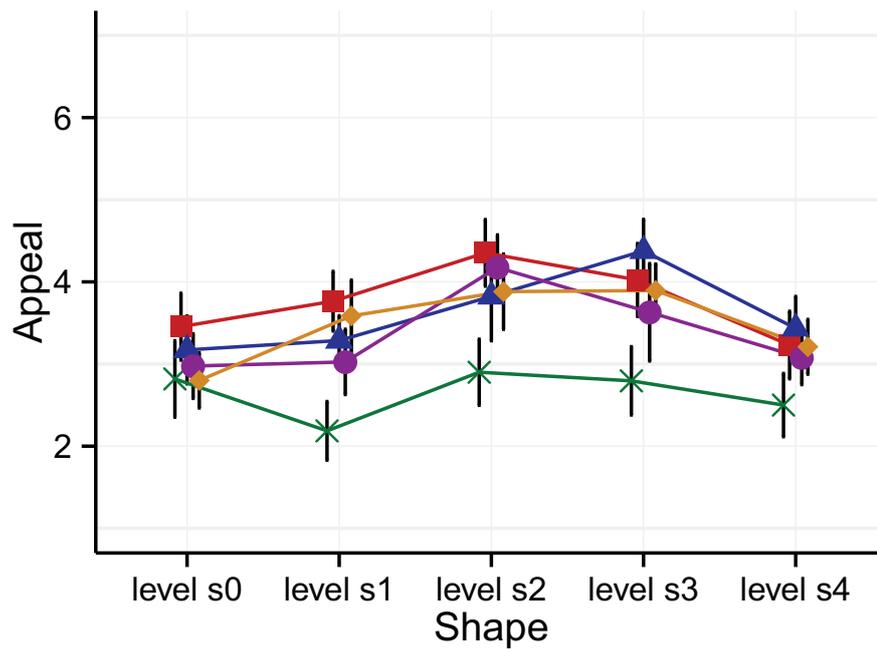
(b) Appeal

Figure 6.15: 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.
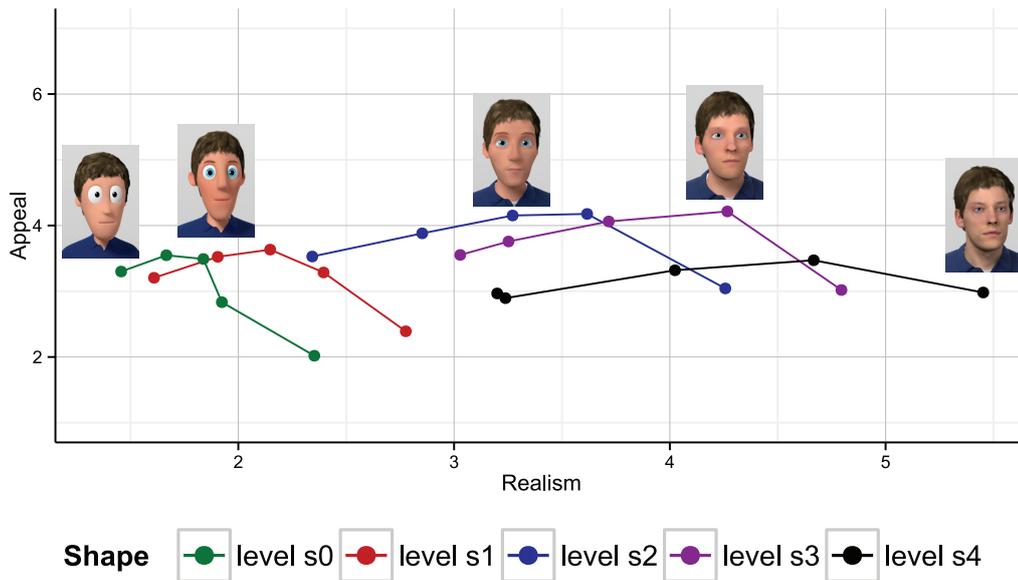
Figure 6.16: Participant ratings for our male stimuli plotted on a realism-vs-appeal diagram, similar to Tinwell, Grimshaw & Williams (2011), MacDorman et al. (2009). Each graph corresponds to one shape stylization, while graph nodes correspond to material levels. The icons are placed above the nodes of matching shape/material levels. The diagram reveals that realism is a bad predictor for appeal. Instead, it is the compatible degree of realism of both shape and material that matters. Results of the female character are similar.

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. This might also be the primary reason for the variations in our ratings. We rule out recognition as an error source since all the expressions were recognized outstandingly well.

6.6   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 to 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 6.12, 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 with empirical knowledge regarding the effect of make-up. 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.

Variations in lighting and shading in Experiment 1b remained unnoticed by our participants. Given the amount of research that is dedicated to realistic rendering, this result is surprising. The combination of three aspects might have contributed to this result. First, texture and shape have been the most relevant features for the task. Second, all tested lighting setups, ranging from direct illumination with hard shadows to global illumination with soft shadows, can be achieved with real light setups (Rademacher 2002) and remained plausible to the participants. Third, we discussed in Chapter 2 that the Phong BRDF can approximate more advanced BRDFs up to a certain level. This fact, paired with perception inaccuracies of the human visual system for frontally illuminated objects with translucent materials (Fleming & Bülthoff 2005), might have sufficiently masked the shortcomings of the simpler material and lighting setups.

Our results are consistent across all tested expressions, except for anger, which was consistently rated less appealing and eerier. 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 across 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 6.16), 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 on stylized shapes. This confirms previous knowledge on modeling or drawing expressive, stylized characters, where the global shape of the character mainly determines expressivity. 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 challenging to connect emotionally with the virtual character. These results are consistent with previous work (Wallraven et al. 2007,

---

[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 the viewer to connect with the story.

2008) and may explain the conscious disturbing effect of stylizing hyper-realistic characters in some movies (e.g., *A Scanner Darkly* or *Renaissance*).

Interestingly, Fleming et al. (2016) investigated on the perceived appeal and realism of body stylization and parallels exist between their and our work. They found no difference between natural lighting embedded in the texture and CG lighting. Furthermore, appeal increased and realism remained constant for moderate shape stylization. One explanation could be that moderate body stylization leads to shapes similar to idealized body shapes in western societies. In our case, moderate stylization of albedo textures had a similar effect as (digital) make-up, that removes skin impurities in order to align better with existing beauty standards. Both observations show that moderate stylization can enhance appeal and can plausibly be explained by traits that increase attractiveness of real people.

As a limitation of the present work, it has to be noted that only two different identities were used. Therefore, our results might not generalize across all conceivable characters. In order to increase validity of the obtained results, our characters have been stylized based on popular 3D characters. Furthermore, the stimuli have been designed to match the quality of current animation movies as much as possible. To our best knowledge, our results are unique in that they provide many different stylization levels for the same characters.

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. As 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. 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.

7 Event-Related Potential Experiment
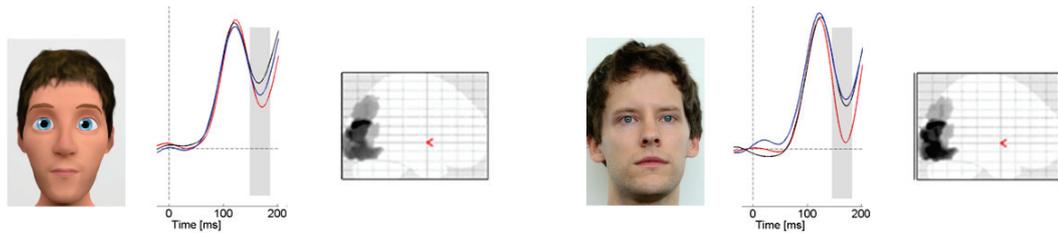
Figure 7.1: For our stimuli (*left*), we measure brain processing in the temporal domain (*center*) and reconstruct the source of the electrical signal (*right*).

We mentioned already that Event-Related Potential (ERP) experiments are an excellent way to identify brain processes and active brain regions during the perception of visual stimuli (Figure 7.1). Furthermore, facial perception of real people has a long history in neuroscience (Section 7.2), which helped to identify face-specific brain regions. In this context, it is especially interesting to investigate whether the brain similarly processes stylized characters. Our stylized characters (Chapter 5) are unique as they stylize the same identity and keep variations small between different stylization levels, e.g., color, view direction, background or lighting. While the electrical activity of the brain can be measured accurately in ERP experiments, it is difficult to connect the measured signal with a meaningful interpretation. Having obtained detailed information in the previous experiments (Chapters 6.3, 6.4 and 6.5) of what participants thought and felt after watching the stimuli allowed us to draw more precise conclusions from the measured data. In the following, we first give a short introduction on understanding ERP experiments and then report our results.

Contribution   This project was part of an interdisciplinary collaboration with the Affective Neuropsychology research group at Bielefeld University. For this study, I created the stylized stimuli, which make the study unique in comparison to previous work. Furthermore, I contributed to the study design and played an essential role in connecting the measured brain responses with results of the rating scale experiments and other ERP studies.
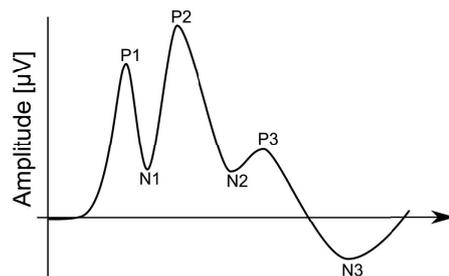
Name	Ordinal	Latency	Scalp distribution
N170	N1	150–190	occipito temporal
EPN	between P2 and N2	250–400	temporo parieto occipital
LPP	after N3	400–600	front parieto central

Figure 7.2: An idealized waveform illustrating the most basic naming convention for amplitudes and an overview of the main components used in the following study.

7.1 Understanding Event-Related Potential Experiments

Empirical studies in the form of perceptual experiments or usability tests are widespread in computer graphics. However, this is not the case for studies that directly measure brain response. Thus before describing the following experiment, a very brief overview of ERP experiments will be given, such that unfamiliar readers can understand the core of this chapter. For a more complete description, we recommend publications dedicated to introduce this topic (Luck 2005, Woodman 2010).

Hans Berger (1929) discovered the possibility of measuring the electrical activity of the brain by placing electrodes on the sculpt. However, the bare electrical signal, called the electroencephalogram (EEG), is rarely of interest for perceptional experiments. Instead, some post-processing of the obtained EEG measurement is required. The beginning of every trial is well defined by displaying the stimulus, and all electrical activity measurements can be aligned according to this starting point. By averaging all EEG signals across participants, equal trials and a set of electrodes for condition A a grand average waveform is obtained (Figure 7.2). This grand average waveform is then compared to the grand average waveform for condition B within specific time intervals (Table Figure 7.2, table). It is important to note that differences are only reported between grand average waveforms. Individual waveforms might vary due to many of reasons like time since the last meal or body temperature (Polich & Kok 1995) and the influence of such side-effects vanishes when data of the entire group is considered.

Every grand average waveform consists of specific peaks that are named either according to the order of appearance (e.g., P1, P2, etc.) or the latency (e.g., N170) after the initial event. The letters P and N indicate positive or negative voltage. In addition to this general nomenclature, specific names for components exist that are defined according to timing, scalp distribution or specific tasks. The N170, EPN (early posterior negativity) and LPP (late positive potential) components are three such examples that are often analyzed in face-related literature (Figure 7.2, table).
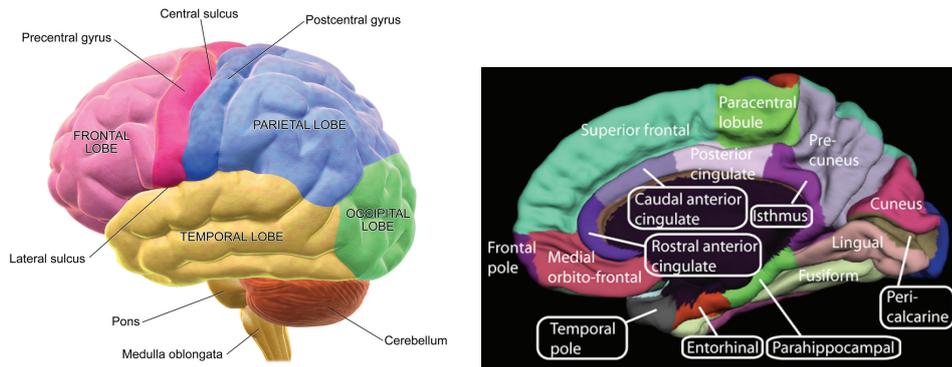
Figure 7.3: *Left*: The division of the cerebral cortex into four different lobes (Blausen 2014). *Right*: Further division of the cerebral cortex into gyri (Hagmann et al. 2008).

At this point the advantage of using ERP experiments becomes obvious. Due to the long research history, it is not only possible to identify a specific reaction of the brain to the stimuli, but also to relate it to previous studies.

Compared to direct feedback techniques like questionnaires or Likert scales, ERP experiments have the advantage of providing an unbiased response to stimuli. However, this comes at the cost that the measured electrical activity does not allow a definite conclusion. Instead, one must rely on a combination of interpretations of previous research for different components in conjunction with rating scale experiments. Unfortunately, as we will see later for the LPP component, there might be several reasons for a stronger activation of a specific component. Besides EEG, brain response can be measured using other techniques like functional Magnetic Resonance Imaging (fMRI). While fMRI is excellent in the spatial resolution and accurately locates brain activation, ERP outperforms fMRI in the temporal domain, allowing a differentiation between sensory and cognitive processing. Nevertheless, it is possible to compute the localization of ERP signals. However, it should be noted that the localization is an ill-posed problem that is only solvable by using additional constraints. Thus source reconstructions of ERP signals are much less reliable than the localization through fMRI.

For completeness, we shortly review the location and functionality of relevant brain structures for our experiment and refer the reader to e.g., Vanderah & Gould (2016) for a more detailed overview. The biggest part of the human brain is the cerebrum, consisting of two symmetrical hemispheres. Each hemisphere consists of the white matter (inside) and the cortex that covers the surface of the cerebrum. The cortex is folded, consisting of gyri[1] and sulci which specify the ridges and the grooves respectively. The cerebrum is divided into four areas, the frontal, temporal, parietal and occipital lobes (Figure 7.3). Part of the occipital lobe is the visual cortex, responsible for the initial processing of visual signals. Different stimuli can also activate nearby areas of the visual cortex, like the cuneus, lingual gyrus or the fusiform cortex. For face processing especially, two dedicated regions exist: the occipital face area (Pitcher et al. 2011) located in the occipital lobe and the fusiform face area (Kanwisher et al. 1997), which is part of the fusiform cortex.

---

[1]singular: gyrus

7.2 Review of Facial Perception in Neuroscience

So far, little is known about how stylized faces are processed by the brain and which facial details are essential in particular. Even less is known about how a presented emotional expression interacts with the stylization of a face. Event-related potentials (ERPs) are an excellent way to analyze face perception, as the measured brain responses can distinguish between highly automatic and more controlled stages of processing and do not require overt responses that may be distorted by experimental demand characteristics. Neuroscientific research has shown that face processing strongly activates dedicated areas in the visual cortex, in particular, the occipital face area (Pitcher et al. 2007, Steevesa et al. 2006) and the fusiform face area (Kanwisher et al. 1997). The fusiform face area is also involved in generating the face-specific N170 component in ERP studies (Deffke et al. 2007). Previous work showed that the N170 component peaks selectively for faces (Bentin et al. 1996). Real faces elicit a stronger N170 compared to abstract sketches of faces, but the difference is not statistically significant compared to schematic faces (Sagiv & Bentin 2001), suggesting a gradual increase. Larger N170 amplitudes are found for human faces compared to those of other species (Gajewski & Stoerig 2011). Further, a recent meta-analysis indicated that larger N170 amplitudes are observed for angry, fearful, and happy than for neutral facial expressions (Hinojosa et al. 2015). It has been hypothesized that the N170 amplitude peak reflects sufficient visual information processing for classifying an emotional expression (Schyns et al. 2007). Moreover, emotion effects at the N170 can be found for faces of medium stylization (Mühlberger et al. 2009) and even for robots with rather schematic faces (Dubal et al. 2011). Attractiveness also seems to induce small modulations in the N170 for characters of the same stylization level (Lu et al. 2014, Marzi & Viggiano 2010). Finally, it has been found that baby faces cause a stronger N170 amplitude than adult faces, most likely due to neotenic features, like proportionally larger eyes (Proverbio et al. 2010).

Enhanced processing of emotional faces is also reflected in an enhanced Early Posterior Negativity (EPN), as well as a larger Late Positive Potential (LPP) (e.g. (Schupp, Öhman, Junghöfer, Weike, Stockburger & Hamm 2004)). The EPN indicates early attention mechanisms, whereas LPP is viewed as reflecting higher-order evaluation and episodic memory encoding (for an overview see (Schupp et al. 2006)). Both EPN and LPP amplitudes are sensitive to voluntary attention deployment to a stimulus and can be increased by explicit attention instructions (Codispoti et al. 2006, Eimer et al. 2003, Schindler & Kissler 2016b, Schupp et al. 2007). Both ERP components are also sensitive to the emotional content of various stimuli types, including faces (e.g., (Flaisch et al. 2011, Schindler & Kissler 2016a, Schupp, Öhman, Junghöfer, Weike, Stockburger & Hamm 2004, Steppacher et al. 2015, Wieser & Keil 2013)). This emotion sensitivity is thought to reflect intrinsically motivated attention to emotional stimuli and often co-varies with stimulus intensity or arousal (Cuthbert et al. 2000, González-Roldan et al. 2011, Olofsson et al. 2008, Schupp, Junghöfer, Weike & Hamm 2004). As, in the absence of relevant social context, participants typically report less subjective arousal for emotional faces than for emotional pictures, differences in arousal-level might account for generally smaller ERP emotion effects in response to faces compared to pictures (Thom et al. 2013). Apart from emotional expressions, larger EPN (Werheid et al. 2007) and LPP amplitudes (Johnston & Oliver-Rodriguez 1997, Marzi & Viggiano 2010, Werheid et al. 2007)

are reported for attractive faces compared to unattractive faces, also applying to moderate stylizations (Lu et al. 2014). Other than larger LPP amplitudes, a better recognition performance has been observed for attractive faces (Marzi & Viggiano 2010). However, enhanced memory performance for attractive faces has been questioned. The effect may reflect a response bias and higher subjective familiarity of attractive faces (Potter et al. 2007).

Interestingly, in a recent study contrasting real neutral faces with neutral faces of puppets, no differences at the N170 level were observed, while from 400 ms onwards a larger LPP was found for real faces (Wheatley et al. 2011). This was attributed to the salience and unique identity of a real face and subsequent mentalizing about the depicted individual (Wheatley et al. 2011). In this regard, perceived uniqueness or distinctiveness by shape or reflectance manipulations of initially non-distinctive real faces have been found to result in a larger late positivity as well as in better memory performance (Itz et al. 2014, Schulz et al. 2012). Further, biographical information also has been found to increase the processing of real faces at early and late ERP components (Abdel Rahman 2011, Abdel Rahman & Sommer 2012). Recently, larger LPP amplitudes have been reported for realistic computer-generated faces when biographical information was added (Taylor et al. 2016). The combined evidence suggests that a unique identity, either pre-existing or ascribed, enhances processing, particularly at late stages.

Against the above background, the present research systematically addresses how brain responses, known to reflect distinct stages of face processing, vary across different levels of character stylization and for different emotional expressions. Based on previous research, we expected to find strongest responses for real faces at the N170, EPN and LPP (Bentin et al. 1996). For very abstract faces, the proportionally larger eyes and uniformly coloured textures might also cause a strong N170 (Proverbio et al. 2010). For angry and happy faces, we expected to find larger N170, EPN and LPP amplitudes. Interactions between the level of stylization and emotional expression were explored on all components. Moreover, source analyses were employed to uncover the cortical generators of stylization effects, addressing the possibility of differential recruitment of visual areas such as fusiform or occipital face areas.

7.3 Experiment Design

Participants Thirty-three participants were recruited at Bielefeld University. They gave written informed consent and received 11 Euros or course credit for participation. The Bielefeld University ethics committee approved the study (EUB number 2016-112). All methods were performed in accordance with the guidelines and regulation at Bielefeld University. The participants (22 females) were 23.30 years on average ($SD = 3.68$), all of them right-handed and had normal or corrected-to normal vision. Upon structured interview, no participant reported a current neurological or psychiatric disorder or relevant medication intake.

Stimuli For the current study, all stylized characters from Chapter 5 were used with adjusted camera view and aspect ratio, such that all faces are nearly of the same size and the eyes are located at similar positions (Figure 7.4). Background planes with a 50% grey were inserted in the 3D scenes before rendering. In addition, photographs of the real actors have been added. To keep the experiment tractable, only three emotional expressions (happy, angry, neutral) per stylization level were

used without interchanging shape and material. In addition, facial neotenic cues have been computed by measuring the relative size of facial parts (Cunningham 1986).

Procedure The faces were randomly presented for 600 ms, followed by a fixation cross with variable latencies (randomly between 400 ms and 500ms). All faces were repeated fifteen times, while it was prevented that the same stimulus was presented twice in a row. Faces were presented on a 15.4-inch screen (Dell Latitude D830) with 1600×1200 pixel resolution (image width: 800; height: 1142). The background colour was 50% grey. The stimulus presentation lasted for about ten minutes, while the whole session took approximately 50 minutes. Participants had no task but were instructed to attend to the presented faces, while moving as little as possible. Participants were encouraged to reduce their eye-movements by focusing on the fixation cross.

EEG recording and analyses EEG was recorded from 128 BioSemi active electrodes[2]. The recorded sampling rate was 2048 Hz. During recording, Cz was used as reference electrode. Biosemi uses two separate electrodes as ground electrodes: A Common Mode Sense active electrode (CMS) and a Driven Right Leg passive electrode (DLR). The two electrodes form a feedback-loop which enables to measure the average potential close to the reference in the AD-box[3], where also information about extra functions of the CMS/DRL loop can be retrieved. Four additional electrodes (EOG) measured horizontal and vertical eye-movement. These were placed at the outer canthi of the eyes and below the eyes.

Pre-processing and statistical analyses were done using BESA[4], EMEGS (Peyk et al. 2011) and SPM8[5] for EEG data. Offline, data were re-referenced to the average reference and then filtered with a forward 0.16 Hz high-pass and a zero-phase 30 Hz low-pass filter. Filtered data were segmented from 100 ms before stimulus onset until 600 ms after stimulus presentation. The 100 ms before stimulus onset were used for baseline correction. Eye-movements were corrected using the automatic eye-artifact correction method implemented in BESA (Ille et al. 2002). Additionally, trials exceeding a threshold of 120 μV were rejected. Overall, 4.04% of all electrode measurements were interpolated. On average, 5.41% of all trials were rejected, leaving 28.38 trials per cell, leading to 85 trials per realism condition and 170 trials per emotion condition.

Cortical source reconstructions of significant ERP differences were generated and statistically assessed with SPM8 for EEG (Litvak & Friston 2008), following recommended procedures. First, a realistic boundary element head model (BEM) was derived from SPMs template head model based on the Montreal Neurological Institute (MNI) brain. Electrode positions were then transformed to match the template head, which is thought to generate reasonable results even when individual subjects' heads differ from the template (Litvak et al. 2011). Average electrode positions as provided by BioSemi were co-registered with the cortical mesh template for source reconstruction. This cortical mesh was used to calculate the forward solution. For the inverse solution, the group inversion algorithm was used (Litvak

---

[2]www.biosemi.com

[3]http://www.biosemi.com/faq/cms&drl.htm

[4]http://www.besa.de

[5]http://www.fil.ion.ucl.ac.uk/spm/
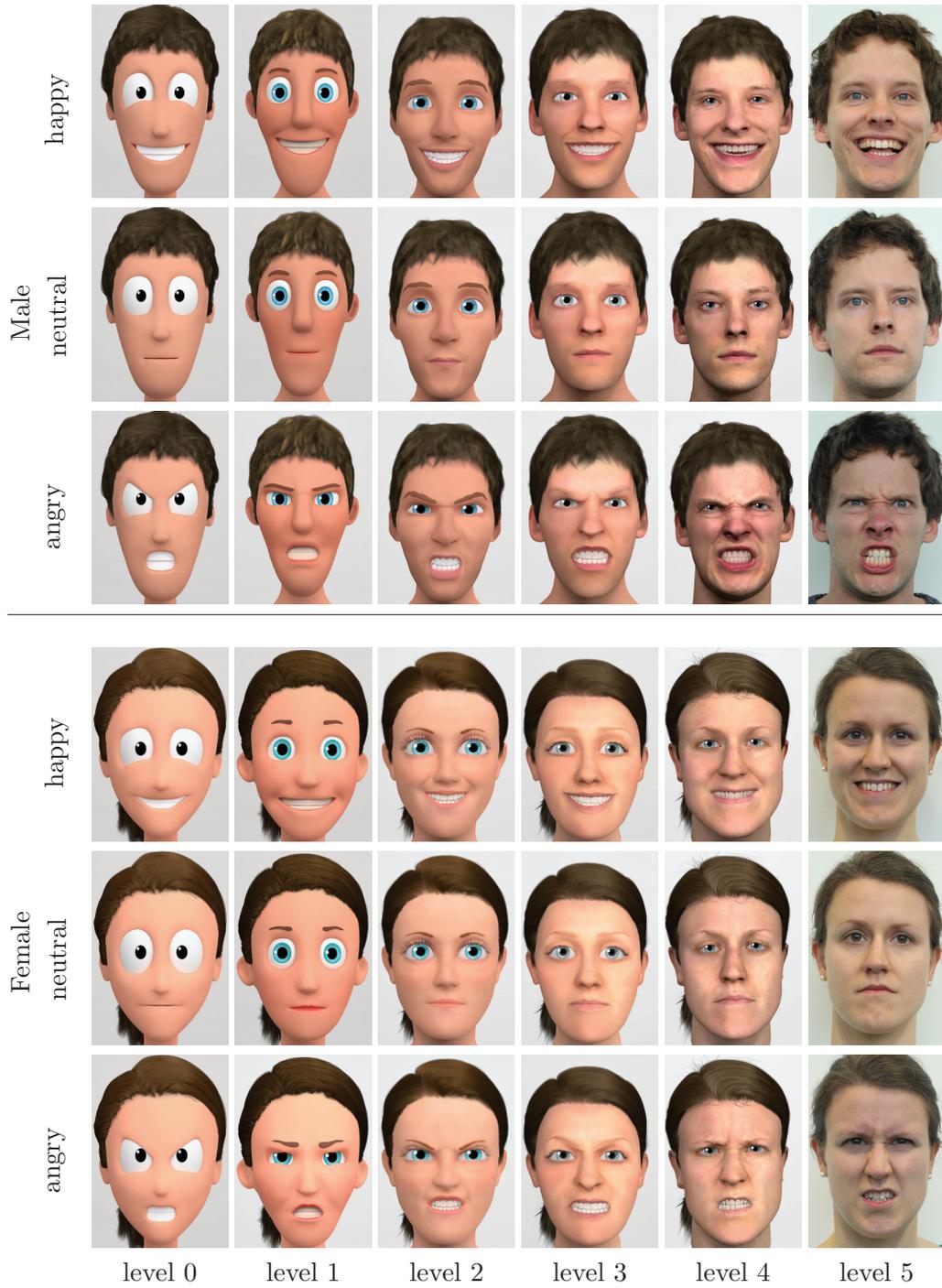
Figure 7.4: The stimulus set consists of the characters (male and female), three emotions (angry, happy, neutral) and six stylization levels.

& Friston 2008) and the solution was calculated from 100 ms pre-baseline to 600 ms after stimulus onset.

Statistical analyses   EEG scalp-data were statistically analyzed with EMEGS. Six (stylization: level 0-5) by three (emotional display: angry, neutral, happy) repeated measure ANOVAs were set-up to investigate main effects of the communicative sender, emotion and their interaction in time windows and electrode clusters of interest. Partial eta-squared (η_P^2) was estimated to describe effect sizes, where $\eta_P^2 = 0.02$ describes a small, $\eta_P^2 = 0.13$ a medium and $\eta_P^2 = 0.26$ a large effect (Cohen 1988). When Mauchly's Test detected a violation of sphericity, degrees of freedom were corrected according to Greenhouse-Geisser (Cunningham & Wallraven 2011). For significant main effects, linear compared to u-shaped contrasts were calculated. Time windows were segmented from 150 to 190 ms to investigate the N170, from 250 to 400 ms to investigate the EPN, and from 400 to 600 ms to the LPP component, after collapsing all conditions and visual inspection of the ERP components. For the N170 and EPN time windows, two large symmetrical temporo-occipital clusters of fifteen electrodes each were examined (left: I1, OI1, O1, POO3, PO9, PO9h, PO7, PO7h, P9, P9h, P7, P7h, TP9h, TP7, TP7h; right: I2, OI2, PO10, POO4, PO10, PO10h, PO8, PO8h, P10, P10h, P8, P8h, TP10h, TP8, TP8hT8). For the LPP time windows, a large parietal cluster was investigated (twenty-six electrodes: CCPz, CP5, CP5h, CP3, CP1, CPz, CP2, CP4, CP6, CPP5h, CPP3, CPPz, CPP4, P3, P1, Pz, P2, P4, PPO3, PPO1, PPOz, PPO2, PPO4, PO3, POz, PO4). Results did not change qualitatively when selecting different literature-based electrode clusters for the N170, EPN or LPP (Dubal et al. 2011, Bublatzky et al. 2014, Itz et al. 2014).

Source reconstructions were performed for the main effects of face-stylization. For each analyzed time window in scalp space, three-dimensional source reconstructions were generated as NIFTI images (voxel size = 2 mm × 2 mm × 2 mm). These images were smoothed with a Gaussian kernel using an 8 mm full-width half-maximum. The statistical comparisons used in source space were based on significant differences on the scalp. In line with previous studies (Schindler & Kissler 2016*a,b*, Schindler et al. 2015), we describe statistical differences in source activity of voxels differing at least at an uncorrected threshold of $p < 0.005$ and a minimum of twenty-five significant voxels per cluster. Additionally, in all tables results are shown applying a family-wise error corrected threshold of $p < 0.05$. The identification of activated brain regions was performed using the LONI atlas (Shattuck et al. 2008).

7.4   Measured Brain Activity

N170   For the N170 component, significant modulations of stylization ($F(5, 160) = 15.93$, $p < 0.001$, partial $\eta_P^2 = 0.332$; see Figure 7.5), emotion ($F(2, 64) = 50.33$, $p < 0.001$, partial $\eta_P^2 = 0.611$; see Figure 7.6), as well as an interaction of stylization with emotion ($F(10, 320) = 2.44$, $p = 0.008$, partial $\eta_P^2 = 0.071$) were found over the two large symmetrical occipital sensor groups (see Figure 7.5). For stylization levels, we tested a linear ($F(1, 32) = 0.09$, $p = 0.765$, partial $\eta_P^2 = 0.003$), compared to a u-shaped contrast ($F(1, 32) = 68.50$, $p < 0.001$, partial $\eta_P^2 = 0.682$), and found a u-shaped form for the face-specific N170 component. Although these u-shaped forms are visible and highly significant for all emotions, we identified with increased
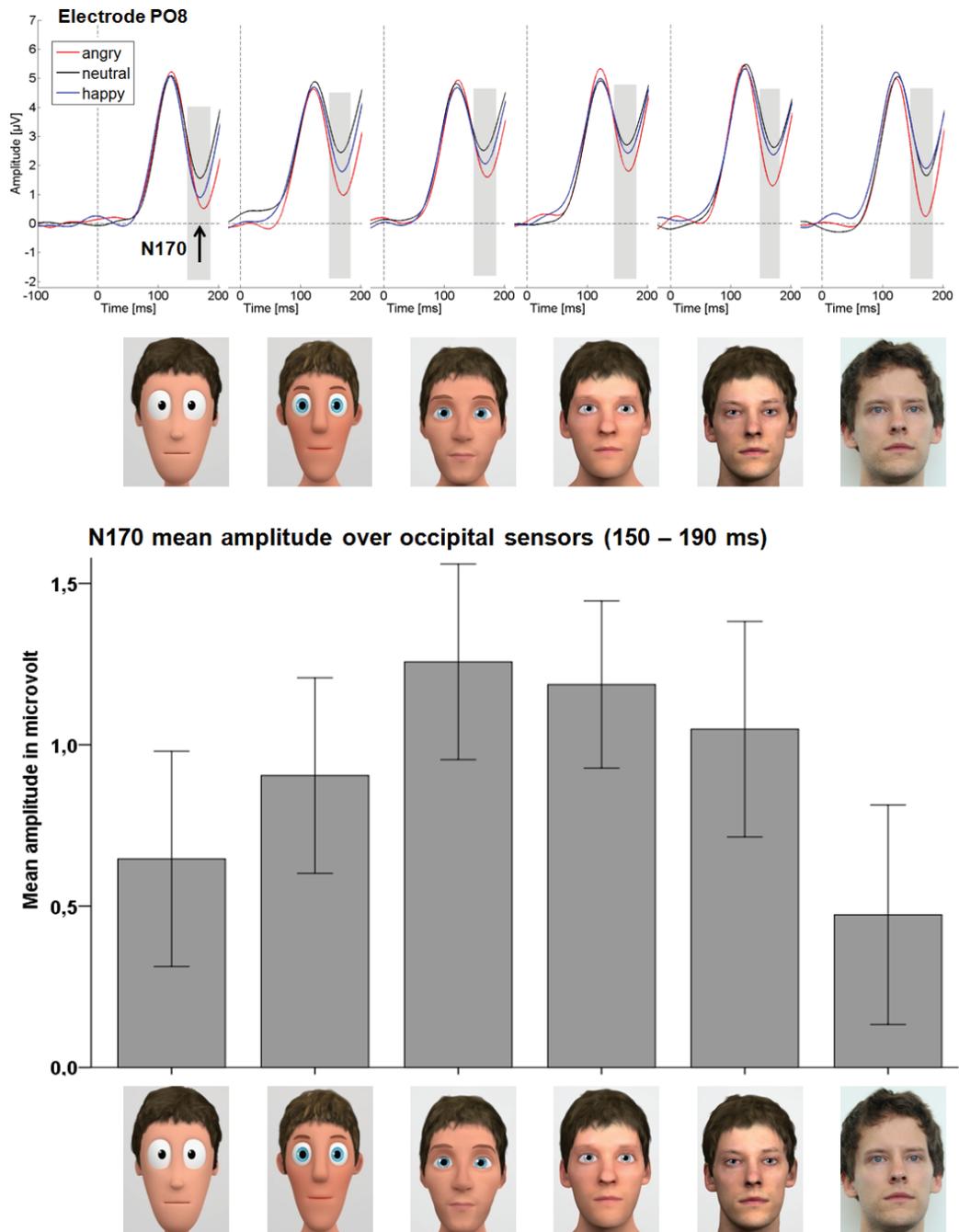
Figure 7.5: N170 ERP results. The upper panel displays the N170 at electrode PO8. From this panel, the interaction between emotion and realism can be observed. The lower panel shows the mean N170 over the occipital sensor cluster. Error bars are +/- one standard deviation of the mean. Note that, while negative-going, the N170 peak is still in the positive range (see top panel). Therefore, smaller bars represent higher N170 amplitudes.
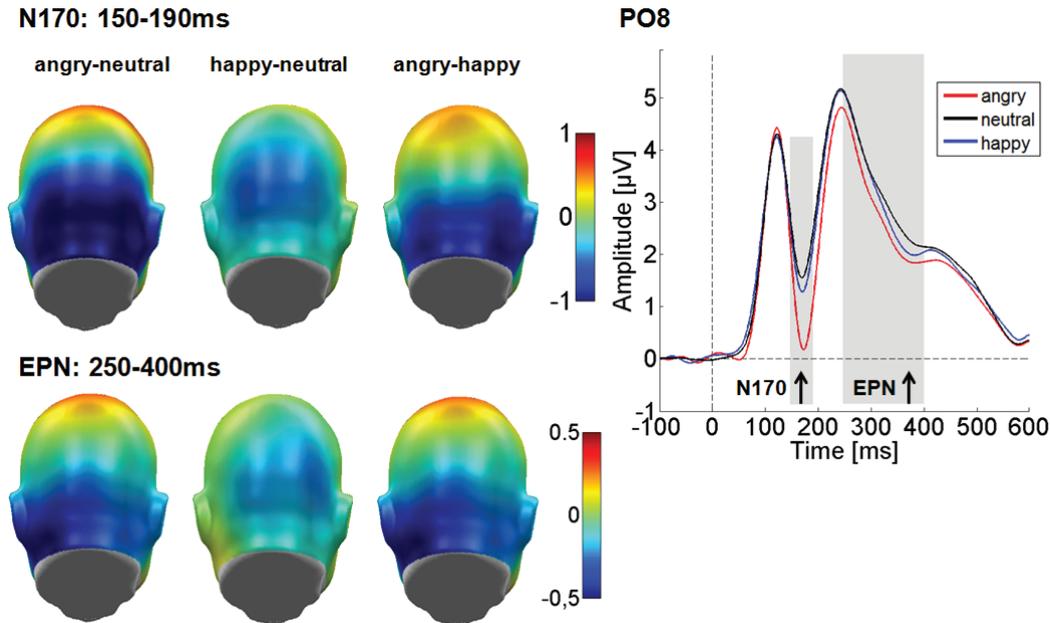
Figure 7.6: Emotion main effects for the N170 and the EPN. *Left*: Displayed are the difference topographies for the main effects of emotion across all stylization levels. Blue colours indicate a relatively larger negativity and red colours a larger positivity. *Right*: Displayed is the time course for angry, happy and neutral faces at electrode PO8.

realism less intense reactions towards happy expressions and more intense reactions towards angry faces (see Figure 7.5, upper panel). Moreover, emotional modulations resulted in a stronger quadratic ($F(1, 32) = 59.81$, $p < 0.001$, partial $\eta_P^2 = 0.651$), compared to a linear contrast ($F(1, 32) = 41.64$, $p < 0.001$, partial $\eta_P^2 = 0.565$), showing the most pronounced N170 for angry faces, smallest for neutral faces and in between happy expressions (all $ps < 0.001$).

Additionally, there was a trend for a main effect of channel group ($F(1, 32) = 3.37$, $p = 0.075$, partial $\eta_P^2 = 0.095$) and an interaction of stylization with channel group ($F(5, 160) = 3.30$, $p = 0.025$, partial $\eta_P^2 = 0.087$), showing in tendency a u-shaped distribution with two maxima over the left and a u-shaped distribution with one maximum over the right sensor cluster. The interaction of emotion with channel group ($F(2, 64) = 0.58$, $p = 0.532$, partial $\eta_P^2 = 0.018$) and triple interaction were both insignificant ($F(10, 320) = 1.03$, $p = 0.471$, partial $\eta_P^2 = 0.031$).

EPN In the EPN time range, over the same cluster, main effects of emotional content ($F(2, 64) = 19.89$, $p < 0.001$, partial $\eta_P^2 = 0.383$) and laterality were observed ($F(1, 32) = 33.83$, $p < 0.001$, partial $\eta_P^2 = 0.514$; see Figure 7.6). The EPN effect was somewhat better explained by a u-shaped form ($F(1, 32) = 22.73$, $p < 0.001$, partial $\eta_P^2 = 0.415$), compared to a linear contrast ($F(1, 32) = 17.59$, $p < 0.001$, partial $\eta_P^2 = 0.355$), where the strongest responses were found for angry faces, least responses for neutral faces and happy faces in between. The effect of laterality showed a larger EPN on the right compared to the left sensor group. All other main and interaction effects were insignificant ($ps > 0.10$).
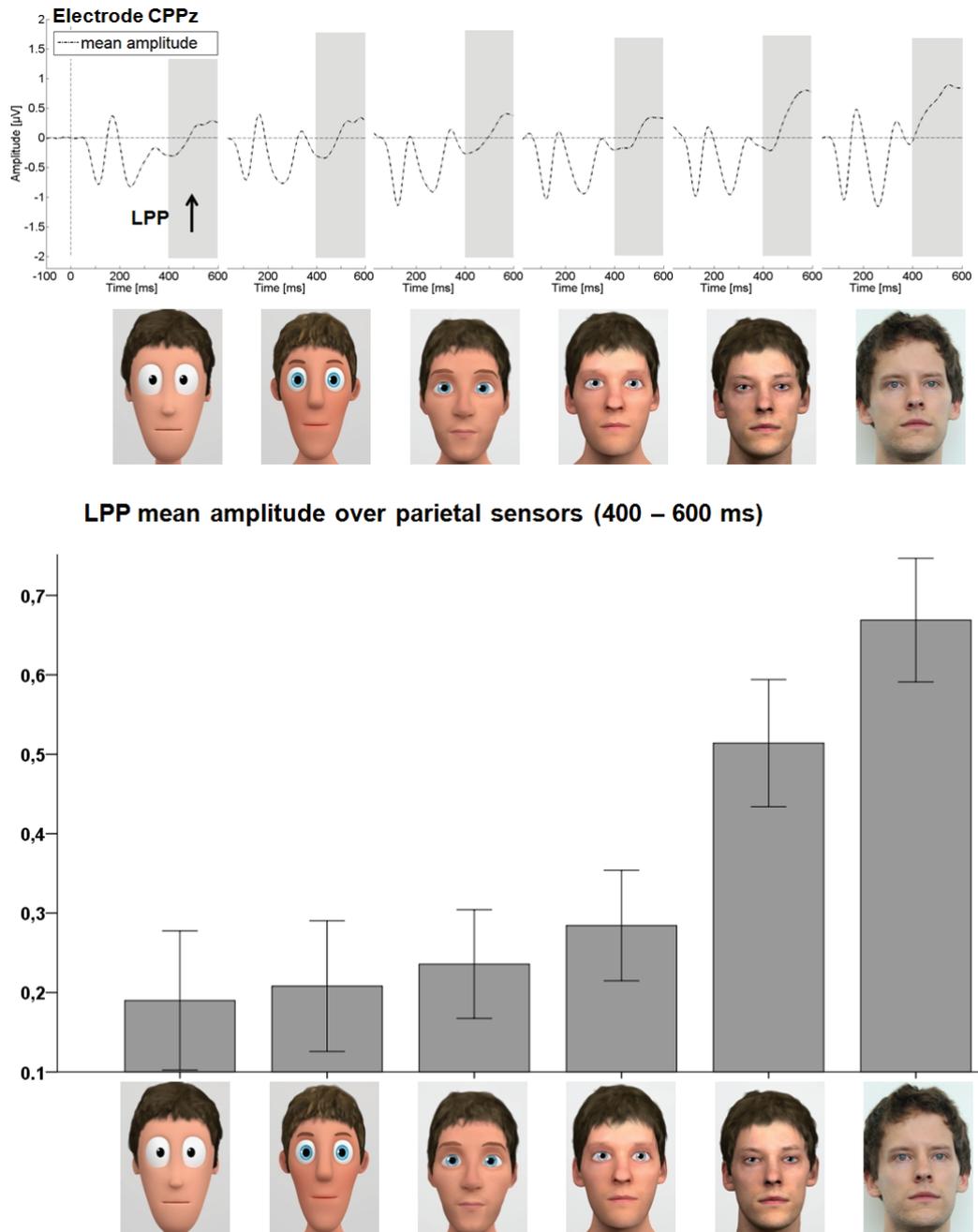
Figure 7.7: LPP ERP results. The upper panel shows the time course for electrode CPPz. Here an increase of LPP amplitudes can be seen for increasing realism. The lower panel displays the mean LPP over the parietal sensor cluster. Error bars are +/- one standard deviation of the mean.

There was no effect of stylization ($F(5, 160) = 1.56$, $p = 0.215$, partial $\eta_P^2 = 0.046$), and no interaction of stylization with emotion ($F(10, 320) = 1.04$, $p = 0.412$, partial $\eta_P^2 = 0.031$), or channel group ($F(5, 160) = 0.80$, $p = 0.555$, partial $\eta_P^2 = 0.024$). Further, there was no interaction of emotion with channel group ($F(2, 64) = 0.76$, $p = 0.474$, partial $\eta_P^2 = 0.023$), and no triple interaction ($F(10, 320) = 1.59$, $p = 0.108$, partial $\eta_P^2 = 0.047$).

LPP In the LPP time window, over a large parietal sensor group, we found only a main effect of stylization ($F(5, 160) = 12.62$, $p < 0.001$, partial $\eta_P^2 = 0.283$). Here, a linear increase in the LPP amplitude was observed (see Figure 7.7). The linear contrast ($F(1, 32) = 31.28$, $p < 0.001$, partial $\eta_P^2 = 0.494$) accounted for substantially more variance explained than did a u-shaped contrast ($F(1, 32) = 11.39$, $p = 0.002$, partial $\eta_P^2 = 0.262$). There was no effect of emotion ($F(2, 64) = 0.04$, $p = 0.965$, partial $\eta_P^2 = 0.001$), and no interaction of stylization with emotion ($F(10, 320) = 1.00$, $p = 0.444$, partial $\eta_P^2 = 0.030$).
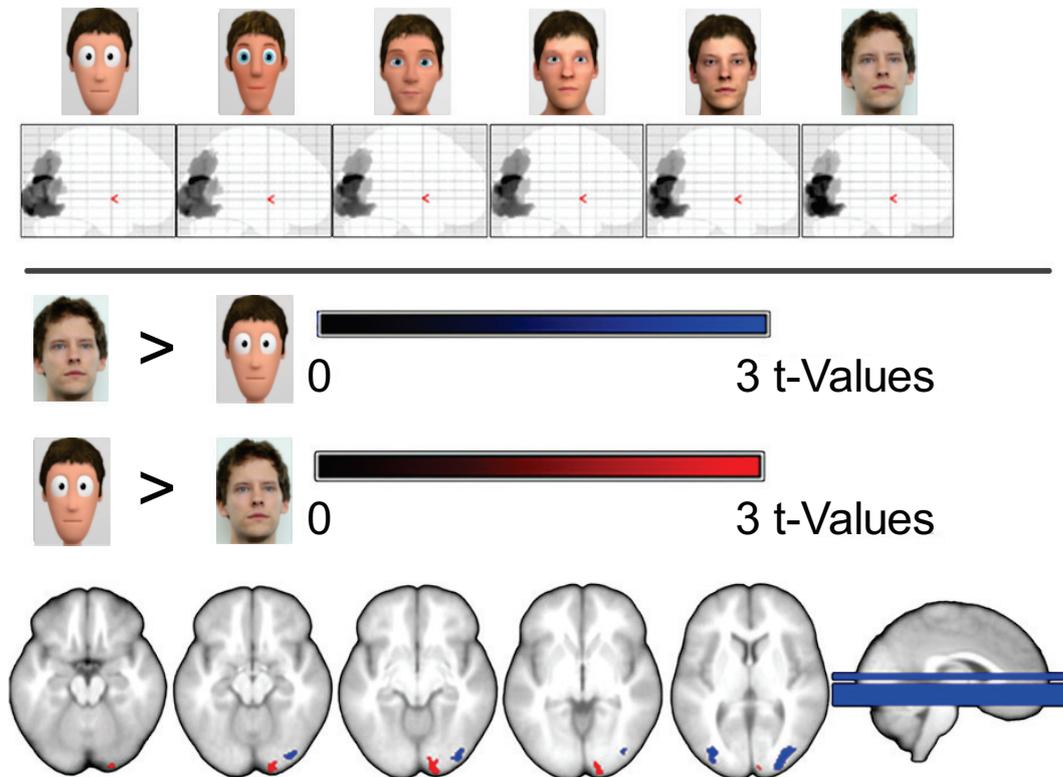
Figure 7.8: Source estimations for main effects of stylization level for the N170. *Upper panel*: Displayed are the family-wise error corrected ($p < .05$) source activations for each stylization level. For all faces, significant generators can be observed for the N170 in bilateral superior, middle and inferior occipital gyri, as well as in the bilateral fusiform gyri. *Lower panel*: Displayed are the differences in source activity between stylization levels (post-hoc contrasts, uncorrected $p < .005$). In the N170, the real faces lead to larger middle and inferior occipital activations, while the most stylized faces are processed more intensely in the right inferior occipital gyrus / cuneus / lingual gyrus.

Cluster-level	peak-level		MNI coordinates			LONI
Number of significant voxels	Peak $t(1, 192)$	Peak p-uncor	x (mm)	y (mm)	z (mm)	Area
Level 5 > Level 0						
408	3.35	$< .001$	28	-90	2	Mid OG R
86	2.70	$< .005$	-34	-90	8	Mid OG L
Level 0 > Level 5						
169	3.31	$< .001$	18	-98	-14	Inf OG R

Table 7.1: Source estimations for stylization level main effects for the N170 (150-190 ms). Results show differences only between very stylized faces (level 0) and real faces (level 5). Real faces elicited more activity in middle occipital regions, while cartoon faces elicited more activity in the right inferior frontal gyrus / cuneus / lingual gyrus.

Notes: Number of significant voxels = voxels which differ significantly between both conditions. Peak p-uncor. = uncorrected p value. For each significant peak, respective coordinates (x, y and z) are displayed in MNI space. If a cluster exhibited more than one peak, only the largest peak is reported. Area = peak-level brain region as identified by the LONI atlas. Mid = middle. Inf = inferior. OG = Occipital Gyrus. R / L = laterality right or left.

Source reconstruction Source reconstructions were calculated for significant effects of stylization level. All presented faces elicited strong visual responses (see Figures 7.8, and 7.9 upper panel). However, the extreme poles (photographs compared to cartoon faces) showed a distinct processing already in the N170: While real faces led to larger inferior and middle occipital activations, highly stylized faces caused stronger responses in the right cuneus / lingual gyrus (see Figure 7.8, Table 7.1). Thus, despite similar N170 peaks for the extreme poles, the cortical generators differ. For neither of the extreme poles did we find statistical differences in source localization compared to moderately stylized characters.

Later, in the LPP stronger superior occipital and superior parietal activations are observed for real compared to cartoon faces (see Figure 7.9, Table 7.2). These source estimations mirror the linearly increasing LPP for higher realism. With increasing realism of the faces, the differences to real faces become smaller and finally disappear in middle occipital areas. For the reverse contrasts, no differences were found.

7.5   Discussion

The results are striking in that they reveal a dissociation of stylization effects on the N170 and LPP: Taking into account that (i) the N170 amplitude is larger for faces than for objects (Kanwisher et al. 1997, Deffke et al. 2007, Proverbio et al. 2010), (ii) larger for real compared to schematic faces (Sagiv & Bentin 2001), and (iii) cuteness and baby-like features have been associated with a larger N170 (Proverbio et al. 2010), we suggest neoteny and perceived face realism to drive the u-shaped N170 modulation. Analysis of stimulus properties (see Figure 7.10) indicates that neotenic features, such as eye size, decrease non-linearly and very quickly for stylization levels 0 to 2. On the other hand, perceived realism increases linearly for the tested characters (Section 6.4).
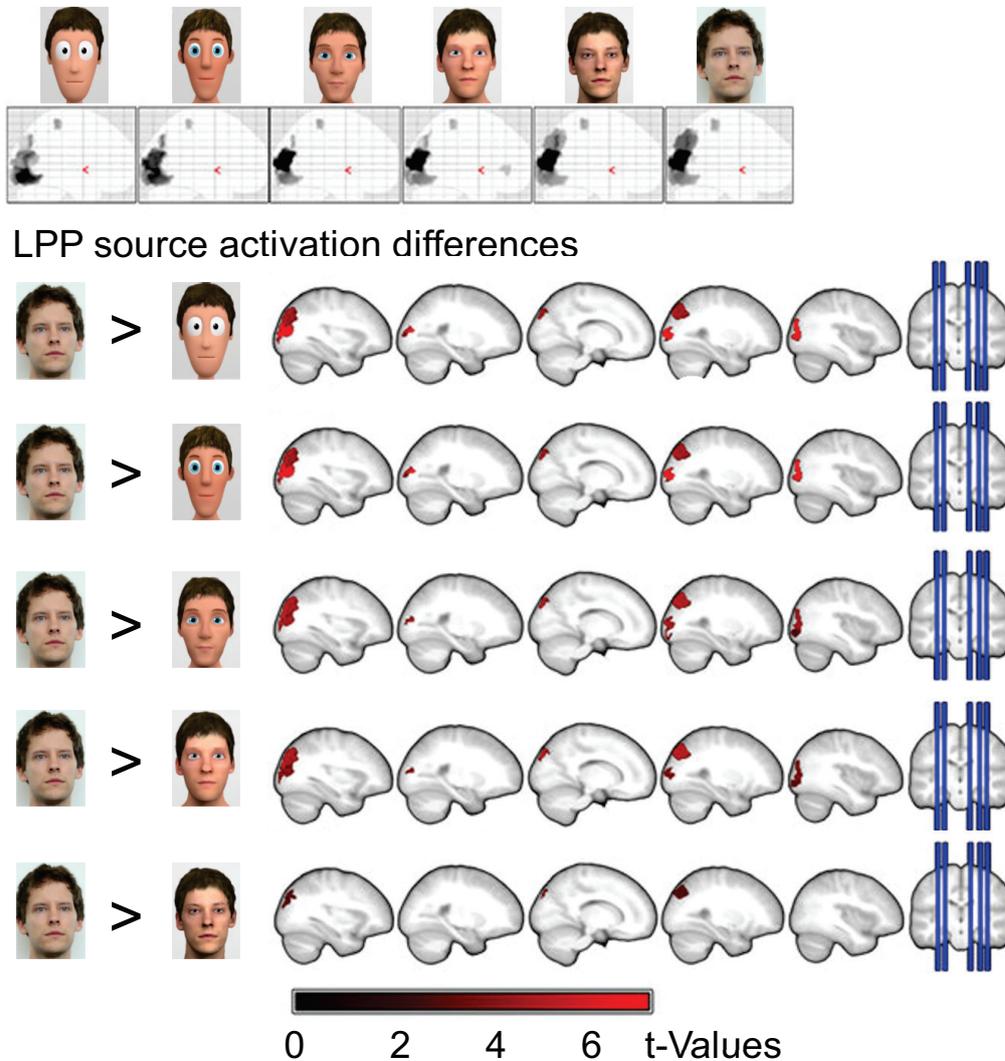
Figure 7.9: Source estimations for main effects of stylization for the LPP time window. *Upper panel*: Displayed are the family-wise error corrected ($p < .05$) source activations for each realism level. For all faces, significant generators can be observed for the LPP in bilateral superior, middle and inferior occipital gyri, as well as in the bilateral fusiform gyri and bilateral superior parietal areas. *Lower panel*: Displayed are the differences between realism levels (post-hoc contrasts, uncorrected $p < .005$). In the LPP, real faces are processed more intensely in bilateral middle and superior occipital and superior parietal areas. However, with increased realism, these differences become smaller and finally disappear in middle occipital regions.

Cluster-level	peak-level		MNI coordinates			LONI
Number of	Peak	Peak	x	y	z	Area
significant voxels	$t(1, 192)$	p-uncor	(mm)	(mm)	(mm)	
Level 5 > Level 0						
1331 (707^a)	7.47	$< .001$	-36	-86	18	Mid OG L
700 (483^a)	7.09	$< .001$	28	-90	2	Mid OG R
673 (209^a)	4.74	$< .001$	20	-82	30	Sup OG R
Level 5 > Level 1						
1192 (639^a)	6.51	$< .001$	-38	-84	20	Mid OG L
617 (356^a)	5.51	$< .001$	34	-88	16	Mid OG R
685 (301^a)	4.94	$< .001$	20	-82	30	Sup OG R
Level 5 > Level 2						
1215 (610^a)	6.06	$< .001$	-36	-86	20	Mid OG L
684 (345^a)	5.13	$< .001$	20	-82	30	Sup OG R
749 (53^a)	4.98	$< .001$	34	-92	2	Mid OG R
Level 5 > Level 3						
1167 (348^a)	5.63	$< .001$	-34	-88	20	Mid OG L
705 (369^a)	5.24	$< .001$	20	-82	30	Sup OG R
565	4.10	$< .001$	36	-86	0	Mid OG R
Level 5 > Level 4						
243	3.18	$< .001$	-38	-82	22	Mid OG L
329	2.91	$< .005$	20	-82	30	Sup OG R

Table 7.2: Source estimations for stylization level main effects for the LPP (400-600 ms). Results show enhanced activity for real faces compared to stylized faces. Real faces (level 5) elicited more activity in middle and superior occipital regions. Differences become smaller with increasing realism of the stylized faces.

Notes: ^{a}Resulting cluster size with FWE-corrected threshold of $p < .05$ ($geq25$ significant voxels). Number of significant voxels = voxels which differ significantly between both conditions. Peak p-uncor = uncorrected p Value. For each significant peak, respective coordinates (x, y and z) are displayed in MNI space. If a cluster exhibited more than one peak, only the largest peak is reported. Area = peak-level brain region as identified by the LONI atlas. Inf = inferior, Mid = middle, Sup = superior. OG = Occipital Gyrus. R / L = laterality right or left.
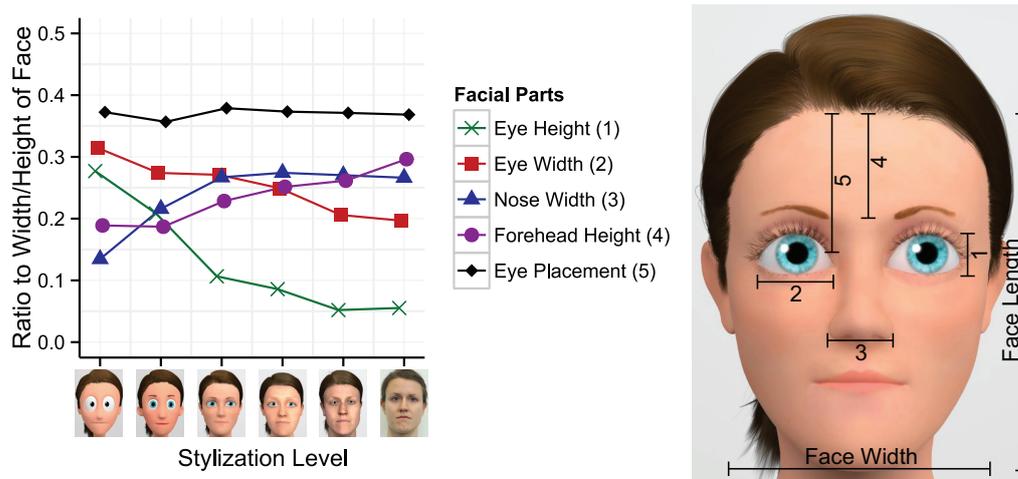
Figure 7.10: Length and width of different facial parts, proportionally to the length and width of the face – averaged across both characters. For more stylized characters, the bigger eyes and smaller nose follow neotenic characteristics. However, this is not the case for vertical eye placement or forehead height.

N170 generators were found to differ between very abstract and realistic faces: Although all faces activated extended visual regions, including the right fusiform gyrus, cartoon faces elicited stronger early visual cortex activations (cuneus, lingual gyrus, inferior occipital gyrus), while for real faces, stronger activations were found in middle occipital regions. Results suggest that processing of highly stylized faces relies more on structural analysis, associated with the so-called occipital face area, whereas realistic faces activate to a greater extent holistic processing, associated with the fusiform face area (Haxby et al. 2000). Fusiform responses have been found for a variety of face stimuli (Tong et al. 2000). However, within computer-generated characters, stronger fusiform responses were found when these looked and acted naturally and meaningfully (Shultz & McCarthy 2014). The present U-shaped modulation could result from an interaction of perceived realism activating fusiform-dependent holistic processing and neoteny features activating feature-based processing in more occipital face areas.

Emotion effects on the N170 and EPN are in line with previous work (Schupp, Öhman, Junghöfer, Weike, Stockburger & Hamm 2004, Mühlberger et al. 2009, Hinojosa et al. 2015). They indicate that emotional expressions modulate the N170 and EPN responses across stylization levels, while, also in line with the literature, in real faces, angry expressions had the largest impact on ERPs (Hinojosa et al. 2015).

It is remarkable that stylization and emotional expressions interact on the N170, indicating an early interplay of structural analysis and emotional classification, rather than dual processing routes for identity and expression. The present data suggest that with increasing realism more resources are captured by cues signaling threat (Bishop 2008). Accordingly, more realistic angry faces were rated more intense compared to happy faces (see Section 6.5.1). On the other hand, for very stylized faces, relatively stronger processing of happy expressions was observed. Neotenic features may selectively enhance the processing of happy expressions. With the exception of the forehead, all neotenic features decrease or remain nearly unchanged

with increasing realism (Figure 7.10). Especially the big eyes and a small nose highly contribute to a cute perception of the character which possibly amplifies processing of positive expressions.

Unlike the N170 amplitude, LPP amplitude increases in line to perceived face realism across the tested stylization levels. This might be due to the uniqueness of a real face, prompting a multitude of ad hoc social inferences (Todorov et al. 2008). Bruce and Young's influential model suggests that after initial perceptual and structural analyses, the observed stimuli are compared with face representations stored in memory, and if there is a match, person-specific knowledge is retrieved (Bruce & Young 1986). Indeed, person-related semantic information enhances LPP amplitudes (Abdel Rahman & Sommer 2012). Recently, it has been further shown that biographical information can increase the LPP to computer-generated faces (Taylor et al. 2016). Although we did not provide participants with explicit biographical information, as a result of social inferences, the more realistic faces might be perceived as having a unique biographical identity. Behavioral evidence showed that computer-generated faces are harder to remember, possibly because they are not encoded as a unique person (Balas & Pacella 2015, Crookes et al. 2015). The noticeable discontinuity between levels 0–3 and 4–5 could also imply a categorical change between realistic and non-realistic characters as shown by classification tasks at a similar stylization level (Looser & Wheatley 2010, Cheetham et al. 2011).

Typically, distinctiveness is achieved by exaggerating specific spatial differences between an individual and an average face (Lee & Magnenat-Thalmann 2000). Distinctiveness by shape or reflectance manipulations has been found to result in a larger EPN and LPP as well as a steeper learning curve and better memory trace for initially non-distinctive faces (Schulz et al. 2012, Kaufmann & Schweinberger 2012, Itz et al. 2014). However, naturally distinctive faces lead to the largest LPP and are remembered even better (Schulz et al. 2012). Regarding the creation of the currently used face stimuli, spatial differences were not overexaggerated in comparison to an average face. However, rated face-realism and distinctiveness might not be uncorrelated, as more realistic faces, for example, exhibit a more detailed texture. Other studies have related the enhanced LPP for real compared to doll faces to the unique identity of the real face, generating an impression of personal social presence (Wheatley et al. 2011). Future studies should aim to disentangle effects of face-distinctiveness from face-realism.

Other factors can also modulate the LPP, but are unlikely to play a role in the current experiment: For instance, LPP responses increase with higher perceived familiarity (Eimer 2000). However, in Section 6.3.1 familiarity remained quite constant across stylization levels. Similarly, facial attractiveness enhances the LPP (Werheid et al. 2007, Marzi & Viggiano 2010). For our stimuli, appeal, which is conceptually similar to attractiveness, was rated highest for medium-stylized faces (see Section 6.4). Therefore, it is unlikely that attractiveness is responsible for LPP modulations in our experiment. In source space, the linear modulation of the LPP was reflected in larger and broader activations in occipito-parietal areas. The localization of this increase is in line with both enhanced perceptual processing of more realistic faces and, in particular, also memory-related processes.

Although higher LPP amplitudes have been reported for more emotional than for neutral stimuli e.g., Flaisch et al. (2011), Bublatzky et al. (2014) and Steppacher et al. (2015), we found no differences between emotional and neutral expressions on this component. In general, during passive viewing, emotion effects are smaller for

faces compared to complex scenes, and participants typically report less subjective arousal for faces (Thom et al. 2013). Nevertheless, large emotion effects were present for the N170 and EPN time window. Similarly, Thom et al. (2013) found emotion effects for the N170, while for the EPN emotion effects were only descriptively visible and no differences were found for the LPP. This suggests that without an explicit task, a highly automatic response towards emotional facial expressions modulates early components without affecting late stages of processing. In this vein, it has been shown that at late stages emotion effects benefit more from attention to the emotional category than do early responses (Schupp et al. 2007, Schindler & Kissler 2016*b*).

Overall, we demonstrated that stylized characters elicit neural effects that are different from the ones elicited by real faces. Importantly, the pattern changes qualitatively across different processing stages, although the measured facial features changed continuously across similar stylization levels. For face perception experiments, which use computer-generated stimuli, this means that, unless a high level of realism is achieved, results cannot be transferred directly to real humans. Thus, computer-generated stimuli may be suitable to test initial hypotheses, but require final validation with real photographs.

So far, it is unclear why or when exactly realism is beneficial in practical applications like games or perceptual studies, but the present study, in demonstrating that realism affects different processing stages in a distinct manner, may offer some clues: On the one hand, highly stylized faces, with their neotenic features, and very realistic faces, influence early stage processing and are equally efficient in transient attention capture. On the other hand, only more realistic faces induce the kind of post-processing necessary to build an individual identity representation and likely facilitating identification with the character. Beyond gaming, these findings have implications for the design of virtual reality therapy settings, for instance of social phobia. They underscore that depending on the overall goal optimal character design will differ. If so, the uncanny valley phenomenon may also arise from a perceived mismatch between situational expectations resulting from a given virtual scenario and character appearance.

In character design, the main problem for artists is that, for adult characters, neotenic and realistic features often exclude each other. For instance, skin smoothness is a neotenic feature, but detailed pores and skin-impurities are required to achieve full realism for a virtual character. Similarly, big eyes are considered as cute, but realistic characters that have unnaturally big eyes are perceived as creepy (Seyama & Nagayama 2007), inverting the intended effect. This dichotomy between realistic characters on one side and rather cute characters on the other side, could also explain the plausibility of the uncanny valley concept, and the present data indeed reveal a neural dissociation that might support it. In practice, considering two independent scales, one for realism and one for neotenic features, seems to be a promising future direction to predict whether a stylized character will be perceived positively or negatively when used in a game or as an interactive agent. Moreover, these parameters can be controlled more easily than appeal or attractiveness, which depend on many aspects and are more subjective.

8 Conclusion

This thesis offers new and practically relevant insights for the creation of appealing realistic and stylized characters by investigating in the fields of geometric modeling, facial animation, face perception and involved processing by the brain. The first part of the thesis focuses on technical aspects of creating realistic characters and highly controllable stimuli. In addition, the presented algorithms improve existing solutions of dense correspondence estimation and facial retargeting, which are fundamental problems in computer graphics. The second part proves the benefit of the technical contributions by enabling new experiments with gained insights that offer a set of guidelines for character design in general. In the following, we shortly list our contributions.

We started with describing a highly practical approach for creating realistic digital faces based on real people. The suggested pipeline in combination with the reported alternatives for specific components can be implemented within a reasonable time. Furthermore, we showed in the template fitting subsection how to obtain best results out of the scanned 3D data. With ELASTIFACE we presented a novel method for establishing correspondences between textured face models. The strength of our approach is its simplicity, robustness, and performance. We have shown that our method is more suitable for the fitting of non-isometric objects than previously published non-rigid registration techniques. Additionally, we have demonstrated practically relevant extensions for perception studies like matching arbitrary texture layouts and part-based morphing. Our contribution to facial retargeting addresses primarily the pre-processing step of (sparse) blendshape transfer. We exploit the inherent similarities between facial expressions of different proportions to generate, through a combination of statistical and geometric methods, a parallel parametrization that fits the range of motion of the actor and preserves the semantic relationships and geometric properties of the facial rig's blendshapes. Furthermore, we introduced a new prior that takes advantage of the differential mesh properties.

For the perceptual experiments, we created to our knowledge the first stimulus set for two identities with a linear modulation of shape and material stylization. This required, especially for the creation of realistic characters, a pipeline that combines recent scientific results with best-practices within the industry. By applying the texture transfer feature of ELASTIFACE, we investigated the impact of shape and material stylization independently. Based on rating scale experiments, we identified shape as the primary attribute for perceived realism, while material, especially the albedo texture, affects more the perceived appeal. Our results have been consistent across all tested expressions, with the exception of the angry expression that was always perceived as less appealing. Furthermore, we measured EEG response while

participants watched a subset of our stimuli. Our results indicate that face realism has a strong influence on the N170 as well as the LPP component. We observed an interaction already at the N170: While neotenic features, perceived realism and emotions influence the N170 component, the LPP component is only influenced by perceived realism. The increased LPP was caused by enhanced occipito-parietal activity and could base on enhanced processing in face specific subregions. Overall, our results suggest that realism is a bad predictor for the perception of appeal, and that the intuition behind the uncanny valley hypothesis might be rather caused through the combination of neotenic features and perceived realism.

By testing our character creation pipeline and ELASTIFACE on a real project, namely the creation of our stimuli, we identified possible directions for further development. Building our 3D assets often required manual interventions. This process could be improved either by further automatization or by simplifying the editing process. While current research focuses on automatization, the second aspect is often ignored. However, being able to quickly fix corrupt data or combine several scans to obtain the best results is of high practical relevance if the captured person or technology is only available for a short time. For establishing dense correspondences between profoundly different shapes, we consider improvements on user interaction and speed as the most promising directions. We believe that an automatic solution is out of scope in such a case because even humans are not able to select consistently corresponding feature points between different shapes. Instead, "correct" correspondences often depend on the context, and optimal correspondences are achieved faster for the target application by quickly iterating different configurations.

In the context of facial animation, the main challenge to solve is the automatic creation of blendshapes for different stylization levels. Such a feature would be highly desirable in order to animate our characters. Existing technical limitations were one of the main reasons why investigating the effect of realistic and stylized animation was outside the scope of this work. 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, it is interesting to note that blendshape based facial animation separates pose (shape) from timing. This raises the fundamental question of how a stylized animation differs if facial retargeting algorithms reliably compensate for individual facial expressions. Another difficult but beneficial issue to solve, in order to extend the stimuli set, would be an automatic or semi-automatic character stylization. This would require a more in-depth understanding of how to maintain similarity across different stylization levels. Finally, we tested in our ERP experiment only a subset of the available stimuli. Using the existing or extending the current stimuli set for new ERP or fMRI studies is an obvious future direction.

A Supplemental Material for Perceptual Experiments

A.1 Additional Diagrams and Analysis

A.1.1 Effect of Expressions in Experiments 1a and 1b

In this section, we discuss the results of Experiment 1a (shape and material) and Experiment 1b (shading and lighting), analyzing the particular *expressions* of the stimuli. For the analysis, we conducted a three-way repeated measure ANOVA. A Tukey HSD test was used for pairwise comparisons within each experiment. Figure 6.15 shows the results, which we proceed now to analyze.

Realism A main effect of expression for realism was found in the shading and lighting experiment (Exp. 1b: $F(4, 76) = 3.78$, $p = 0.007$), but not in the shape and material experiment. The effect could mainly be attributed to the sad expression, which has been rated slightly more realistic ($p = 0.005$) than others. Because the means (3.90 ± 0.15) of all groups are within a small range, we classify this effect as noise and omit similar examples for the rest of this section.

Appeal and Reassurance A main effect was found for appeal in both experiments (Exp. 1a: $F^*(2.57, 54.01) = 33.14$, $pp < 0.0001$,$\epsilon = 0.643$; Exp. 1b: $F^*(1.53, 29.10) = 22.22$, $p < 0.0001$, $\epsilon = 0.383$), which is primarily caused by the anger expression ($p < 0.0001$). Similarly, there is a main effect of expression for reassurance in both experiments (Exp. 1a: $F^*(2.61, 54.77) = 24.61$, $p < 0.0001$,$\epsilon = 0.652$; Exp. 1b: $F^*(1.68, 31.92) = 18.61$, $p < 0.0001$, $\epsilon = 0.420$), again mainly caused by the anger expression ($p < 0.0001$).

Familiarity A similar main effect is obtained for familiarity (Exp. 1a: $F(4, 84) = 8.80$, $p < 0.0001$; Exp. 1b: $F^*(2.04, 38.84) = 5.15$, $p = 0.001$,$\epsilon = 0.511$). In this case anger is the only reason for the significant differences of the means. But the anger expression is only significantly different from the happy ($p = 0.003$) and neutral expressions ($p = 0.001$).

(a) Realism

(b) Appeal

(c) Reassurance

(d) Familiarity

Expression anger happy neutral sad surprise
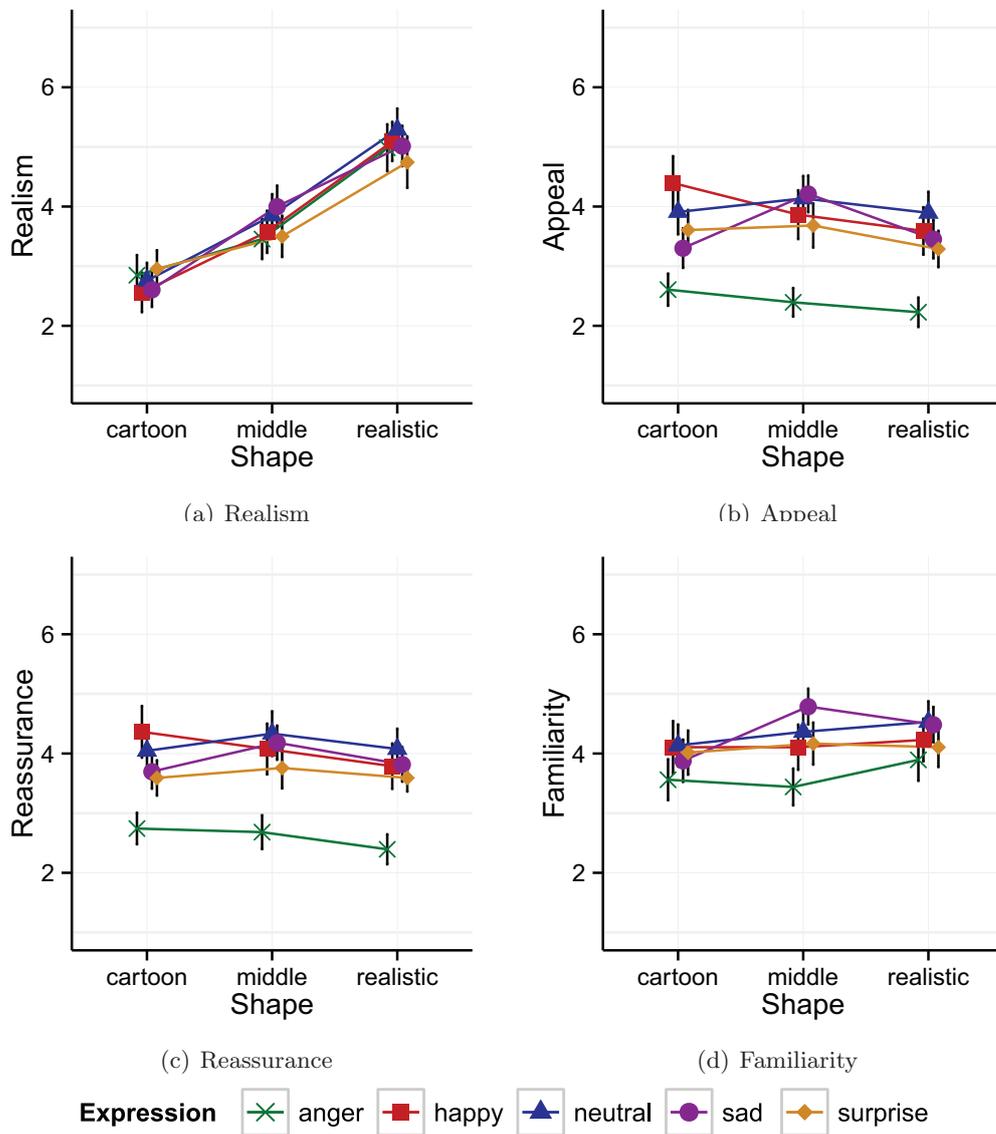
Figure A.1: Results of expressions in Experiments 1a: While emotions do not differ in realism, the anger expression was perceived as more eerie and unappealing for all stylization levels.

(a) Realism

(b) Appeal

(c) Reassurance

(d) Familiarity

Expression ✕ anger ■ happy ▲ neutral ● sad ◆ surprise
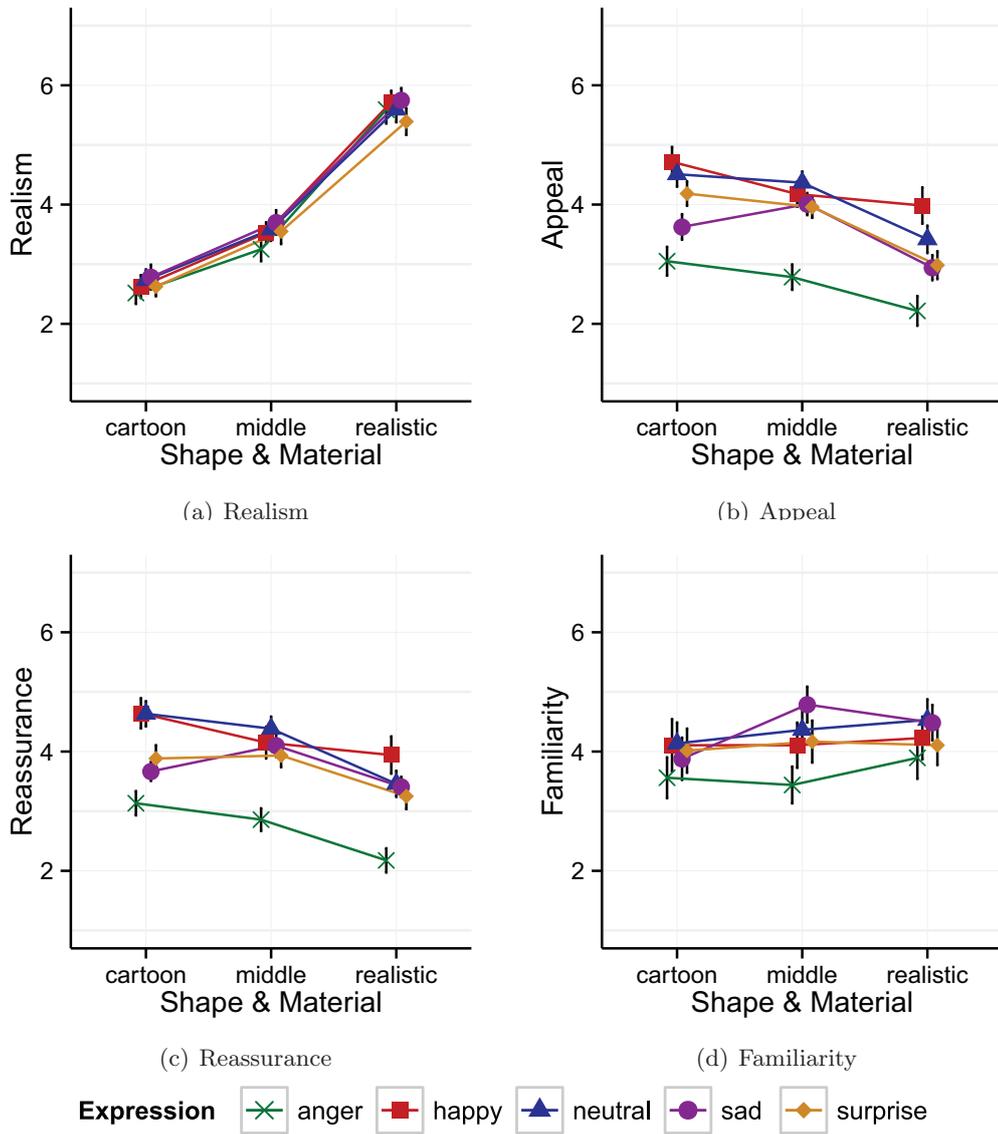
Figure A.2: Results of expression in Experiment 1b: Ratings for perceived realism, appeal, reassurance and familiarity grouped by stylization level and expression.

A.1.2 Results of Experiment 2 Itemized by Character
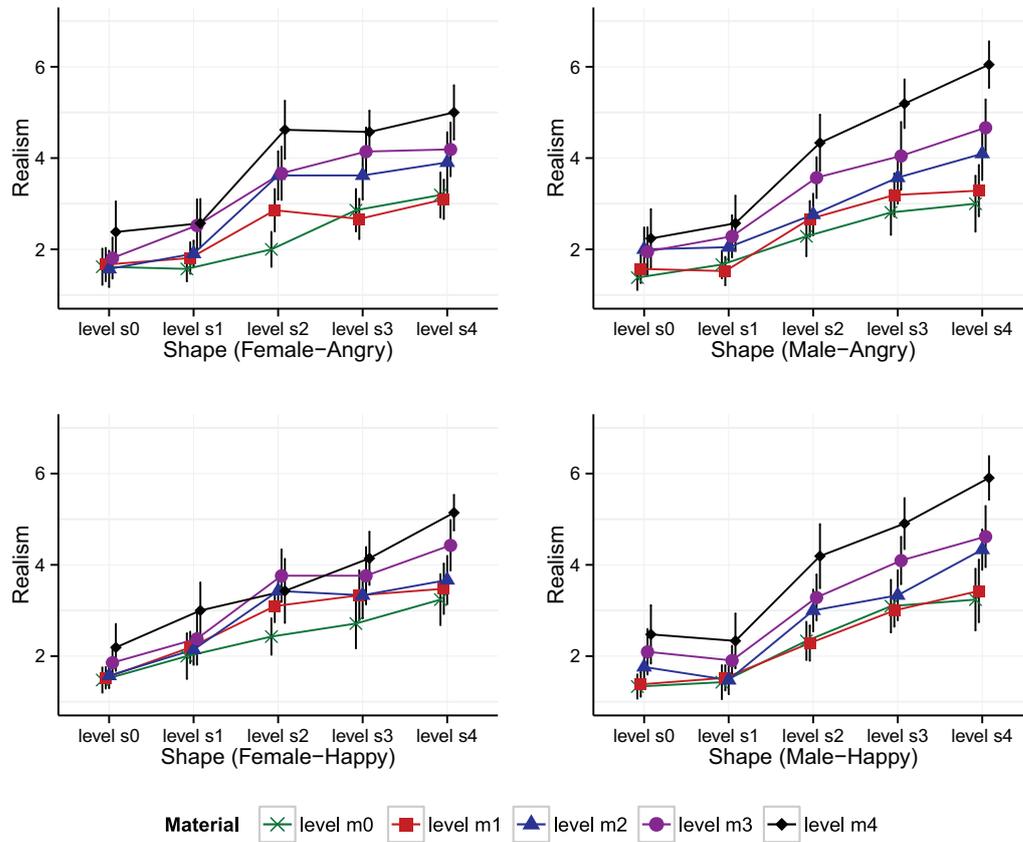
Figure A.3: Results for Experiment 2: Ratings for perceived realism, separated by expression, male and female.
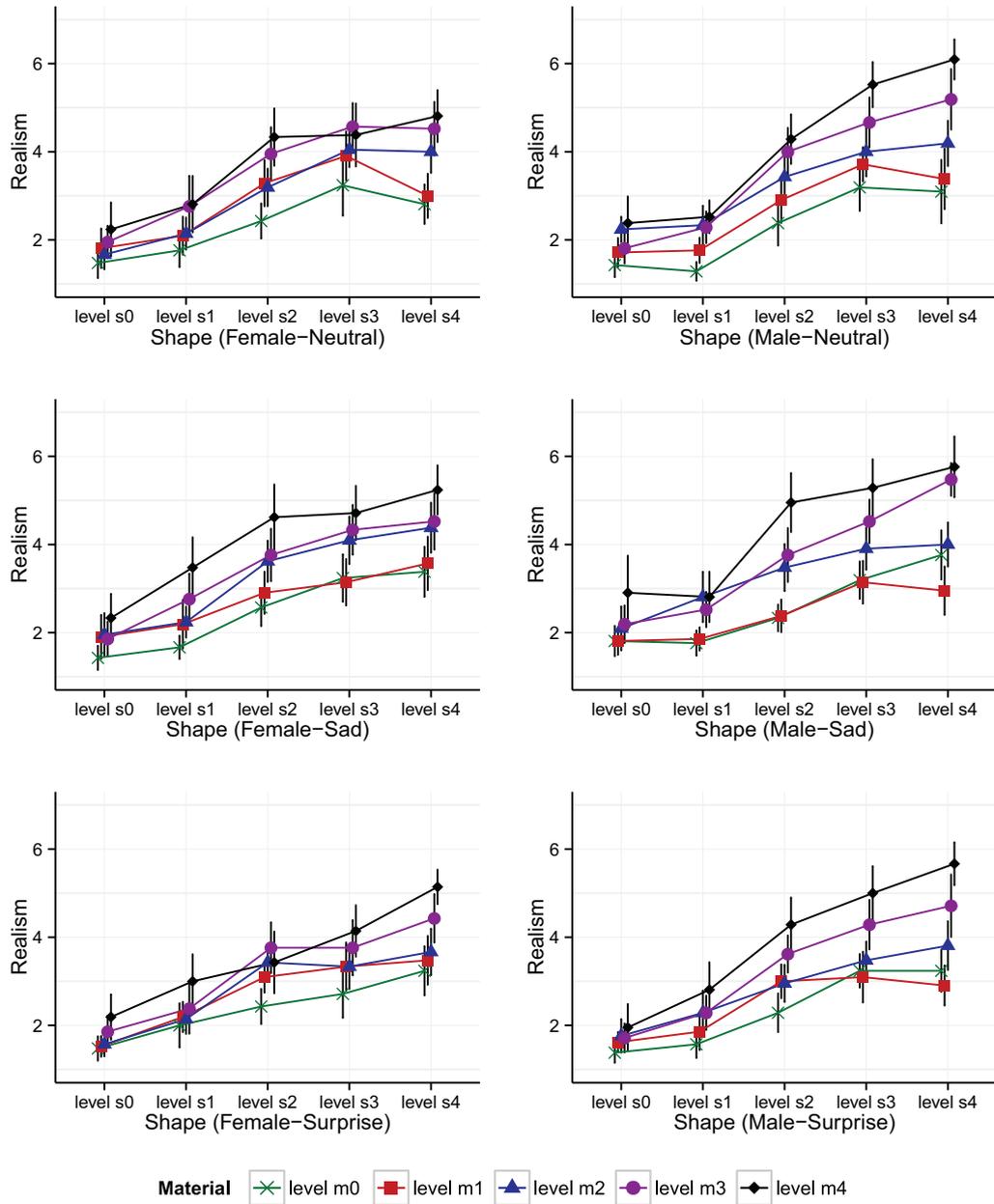
Figure A.4: Results for Experiment 2: Ratings for perceived realism, separated by expression, male and female.
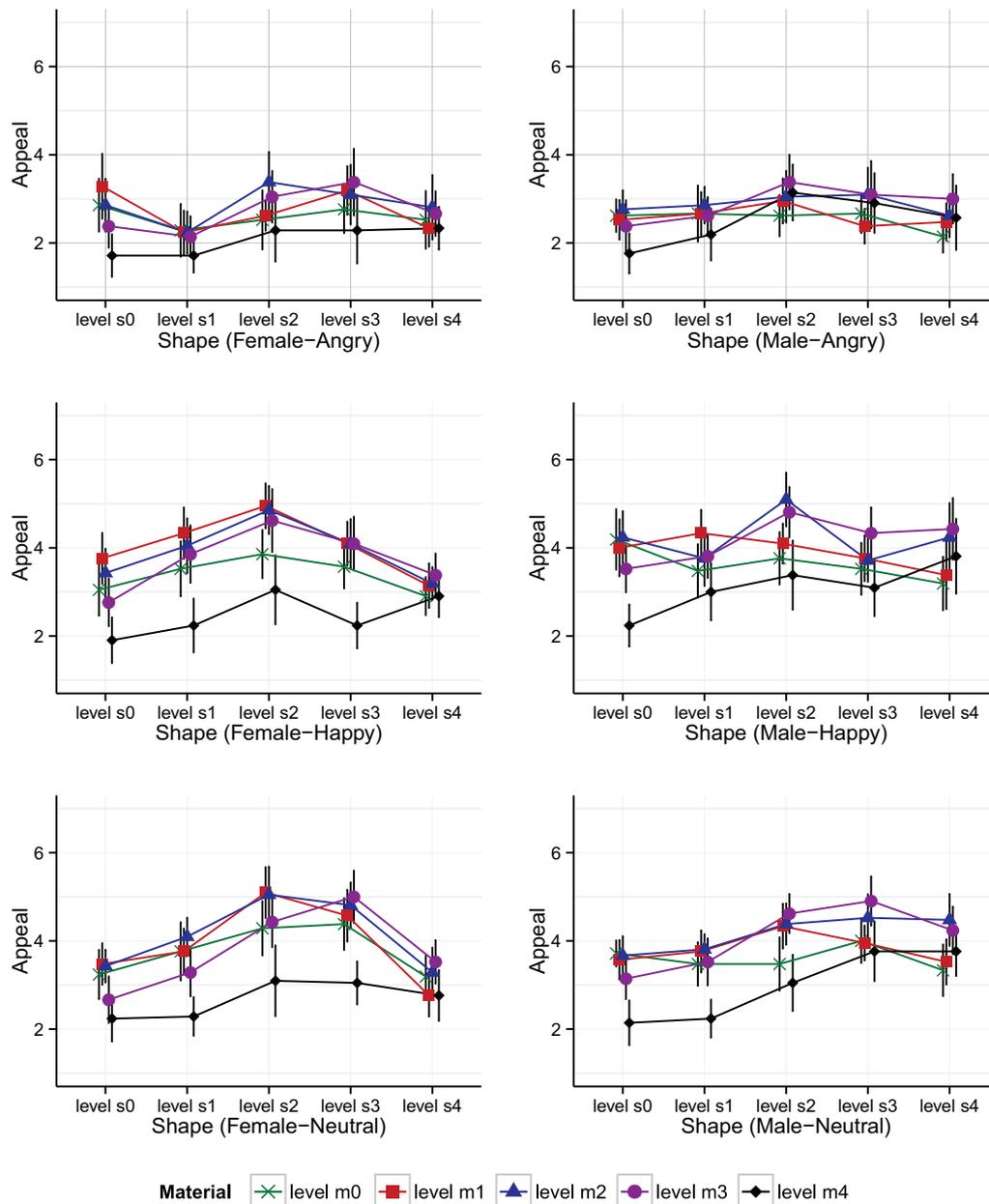
Figure A.5: Results for Experiment 2: Ratings for perceived appeal, separated by expression, male and female.
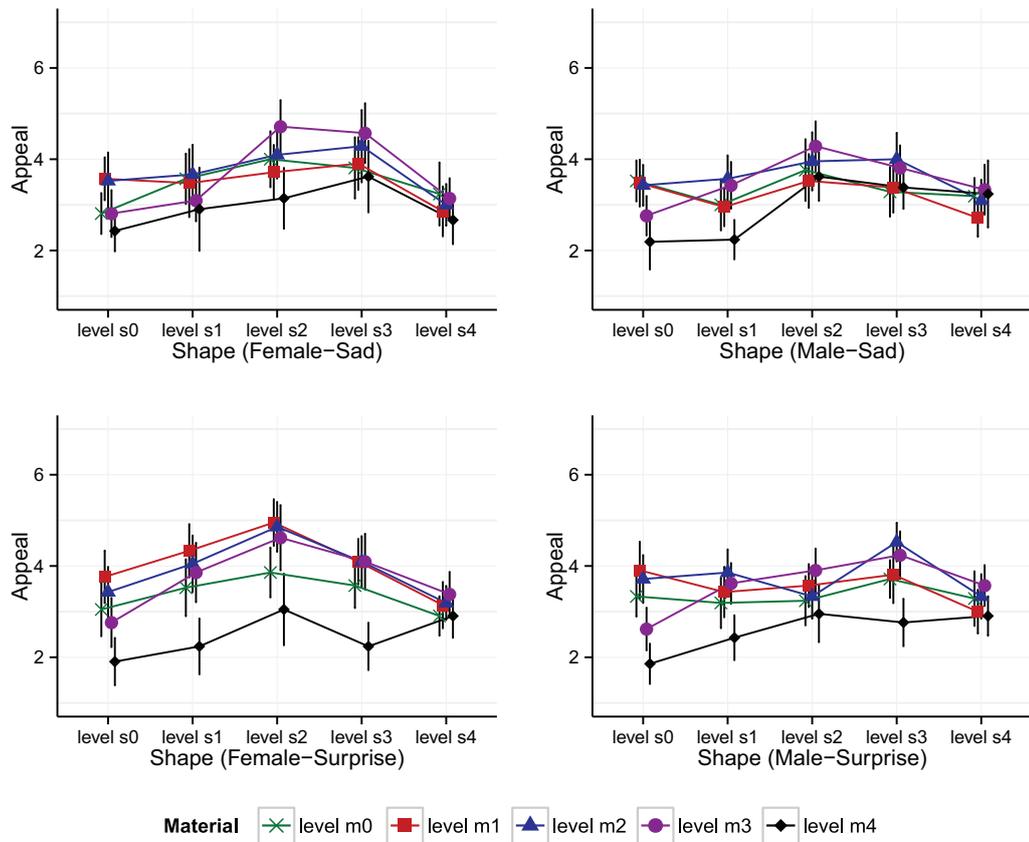
Figure A.6: Results for Experiment 2: Ratings for perceived appeal, separated by expression, male and female.

A.2 Renderings of all Stimuli

The following figures show all shading and lighting combinations of Experiment 1b (Section 6.3.2) and all shape and material combinations of Experiment 2 (Section 6.4). In all cases, the x-axis represents the facial expressions. The y-axis represents the shading and lighting setup or the material stylization levels respectively.
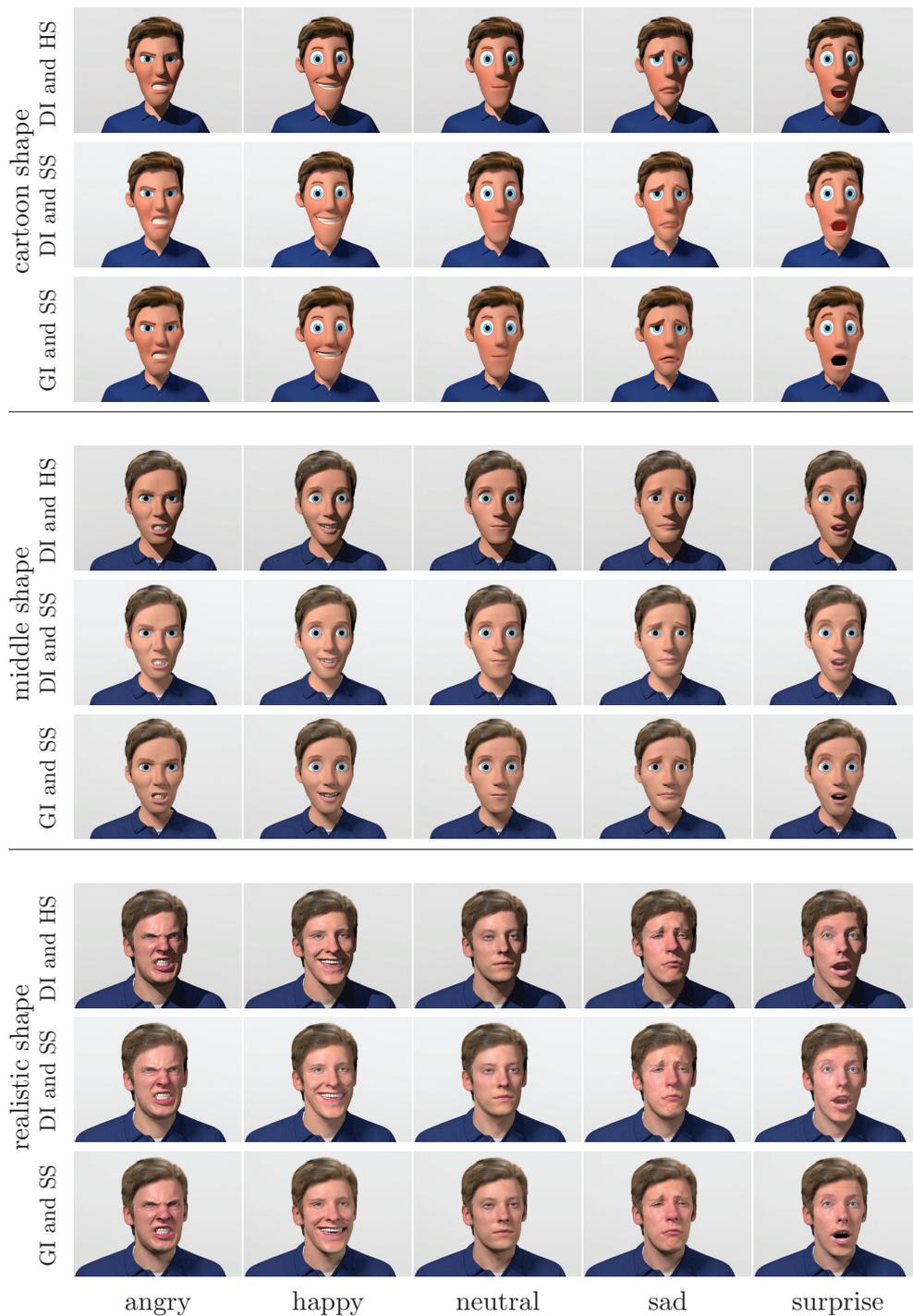
Figure A.7: Experiment 1b, shading and lighting. The five expressions (*x-axis*) of the cartoon shape with Phong shader for skin, cloth etc., under three different lighting setups (*y-axis*) – *From top to bottom row*: direct illumination with hard shadows (*DI and HS*), direct illumination with soft shadows (*DI and SS*), and global illumination with soft shadows (*GI and SS*).
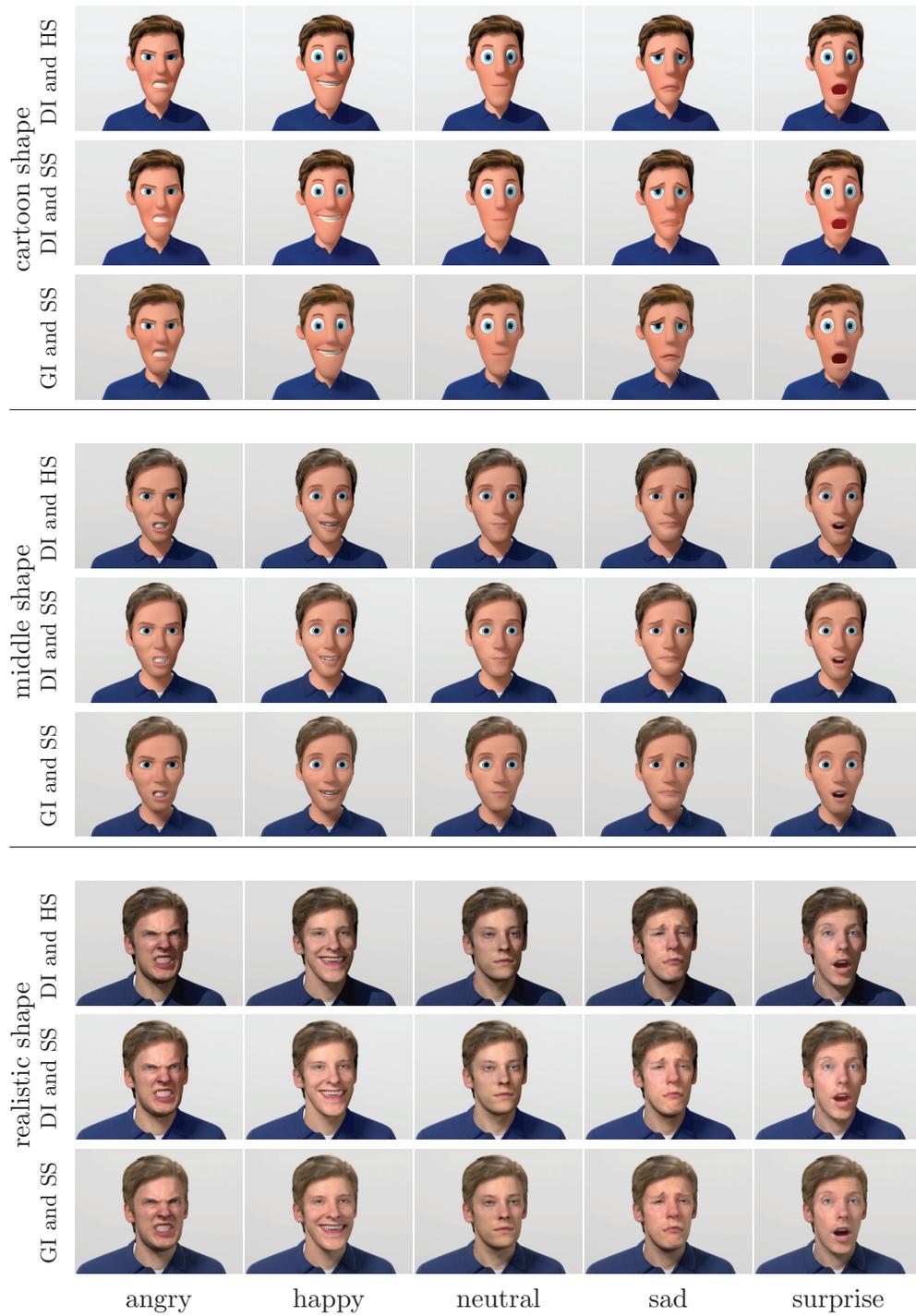
Figure A.8: Experiment 1b, shading and lighting. The five expressions (*x-axis*) of the cartoon shape with sophisticated materials for skin, cloth etc., under three different lighting setups (*y-axis*) – *From top to bottom row*: direct illumination with hard shadows (*DI and HS*), direct illumination with soft shadows (*DI and SS*), and global illumination with soft shadows (*GI and SS*).
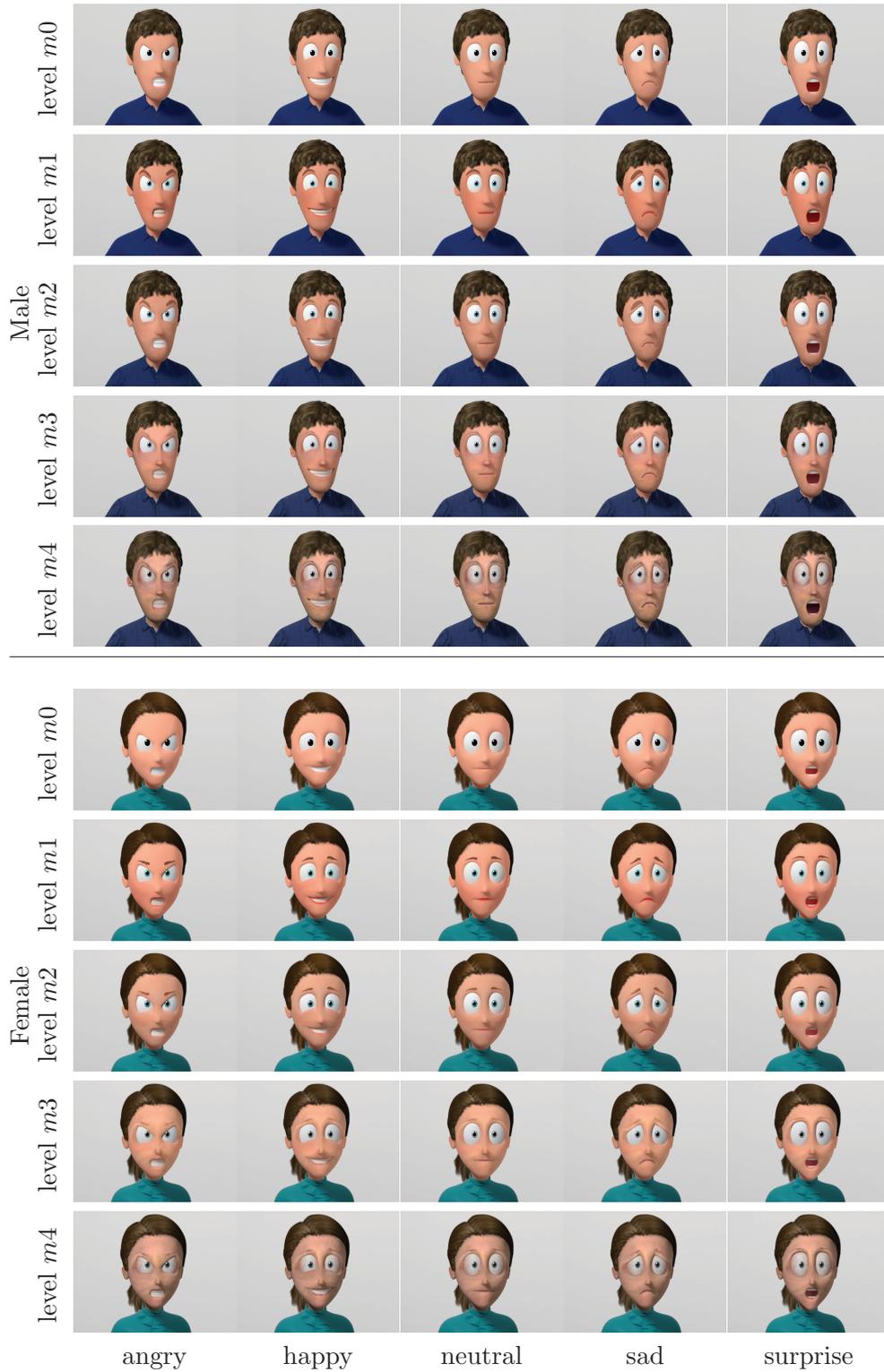
Figure A.9: Experiment 2. Five expressions (*x-axis*) of the level *s*0 shape of both characters, combined with the five levels of material stylization (*y-axis*).
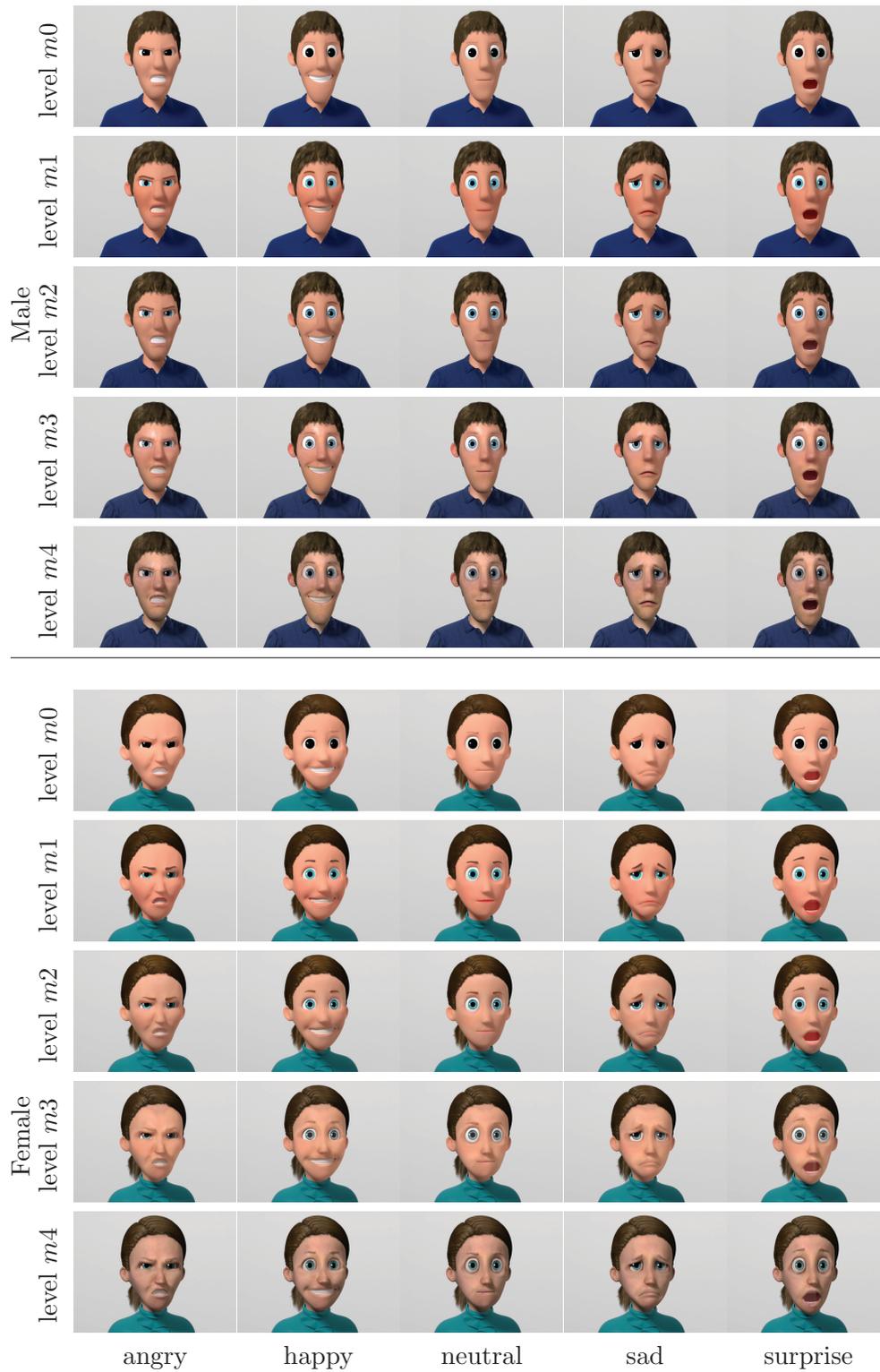
Figure A.10: Experiment 2. Five expressions (*x-axis*) of the level $s1$ shape of both characters, combined with the five levels of material stylization (*y-axis*).
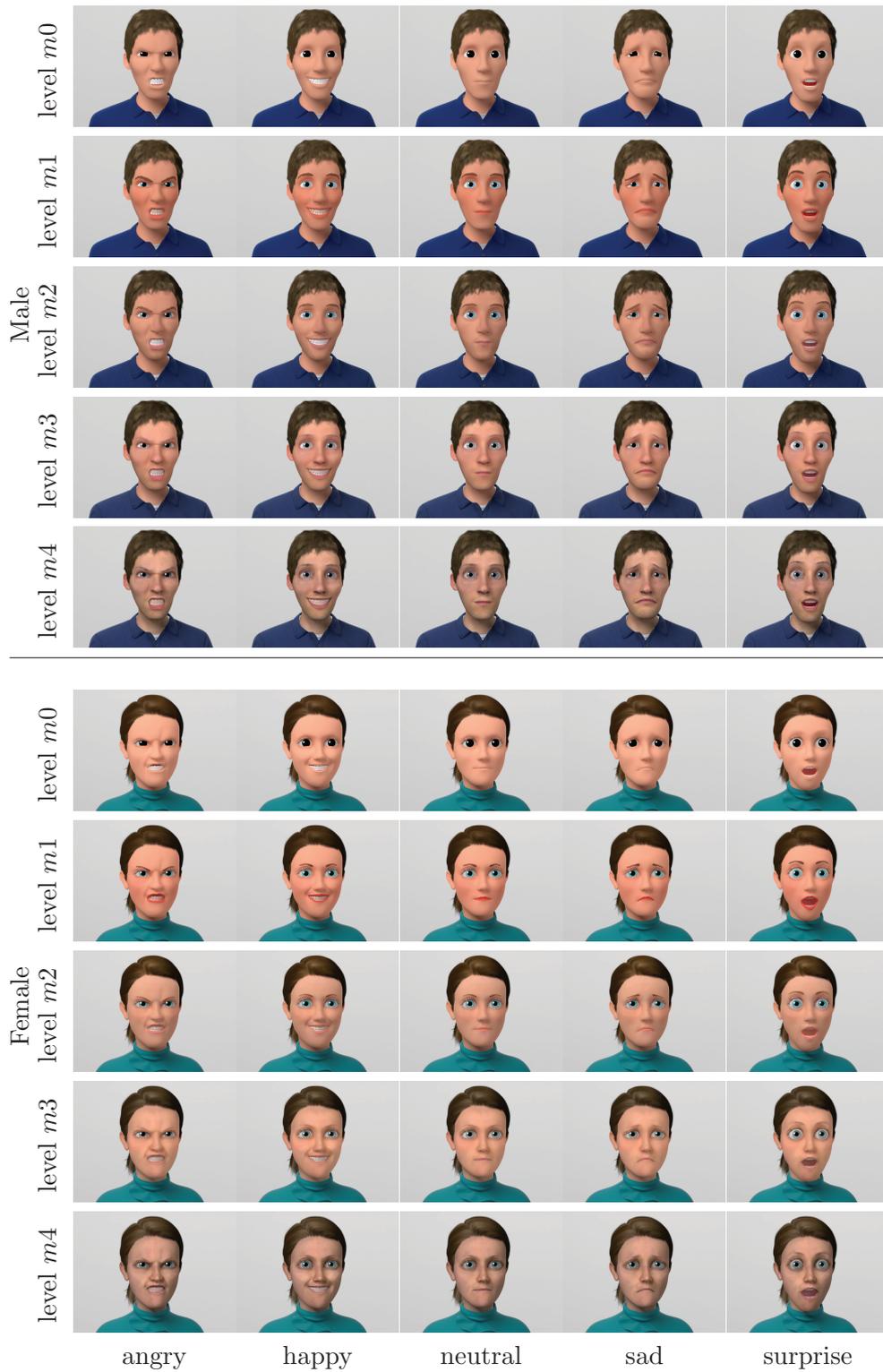
Figure A.11: Experiment 2. Five expressions (*x-axis*) of the level *s2* shape of both characters, combined with the five levels of material stylization (*y-axis*).
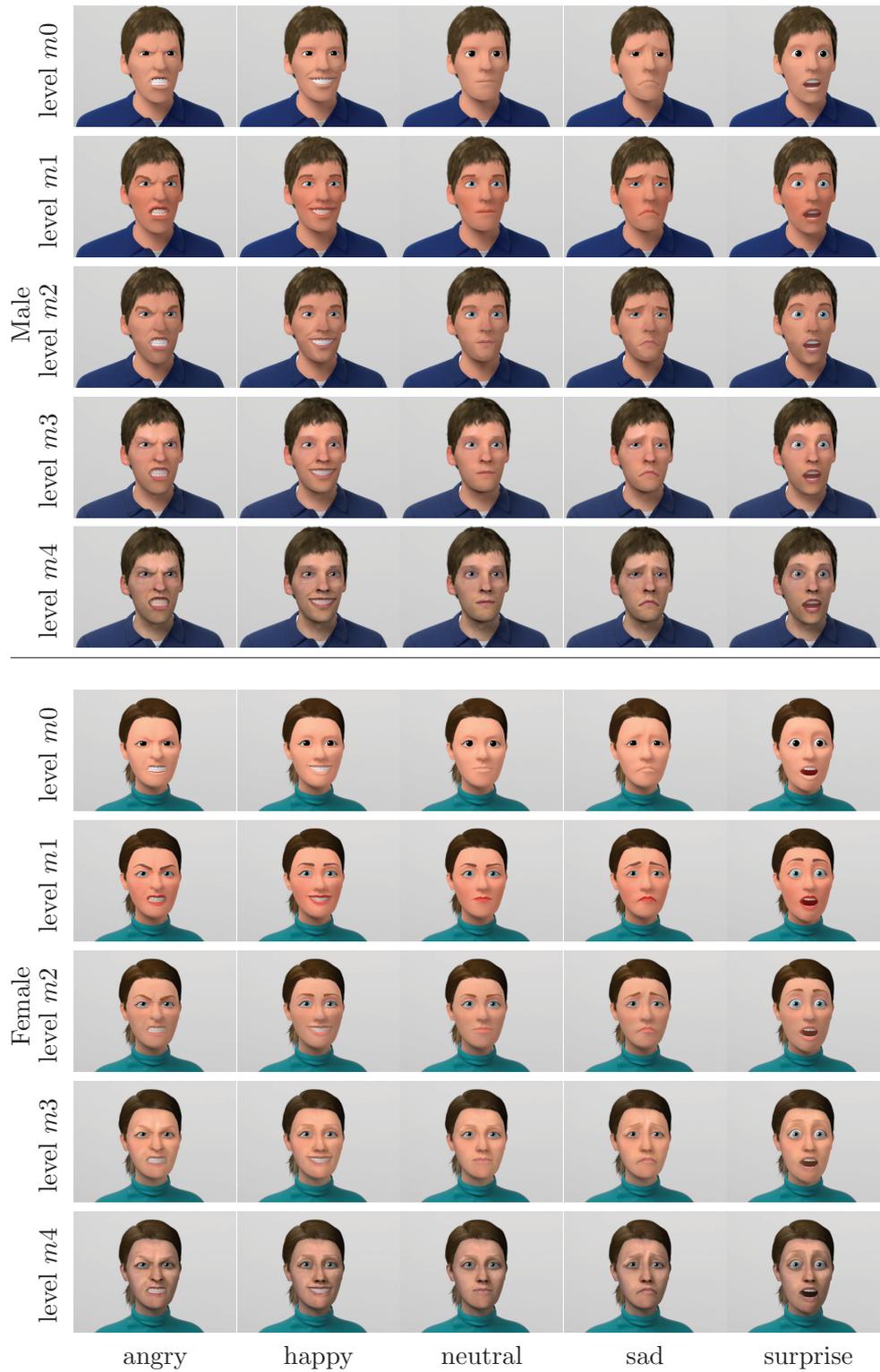
Figure A.12: Experiment 2. Five expressions (*x-axis*) of the level *s3* shape of both characters, combined with the five levels of material stylization (*y-axis*).
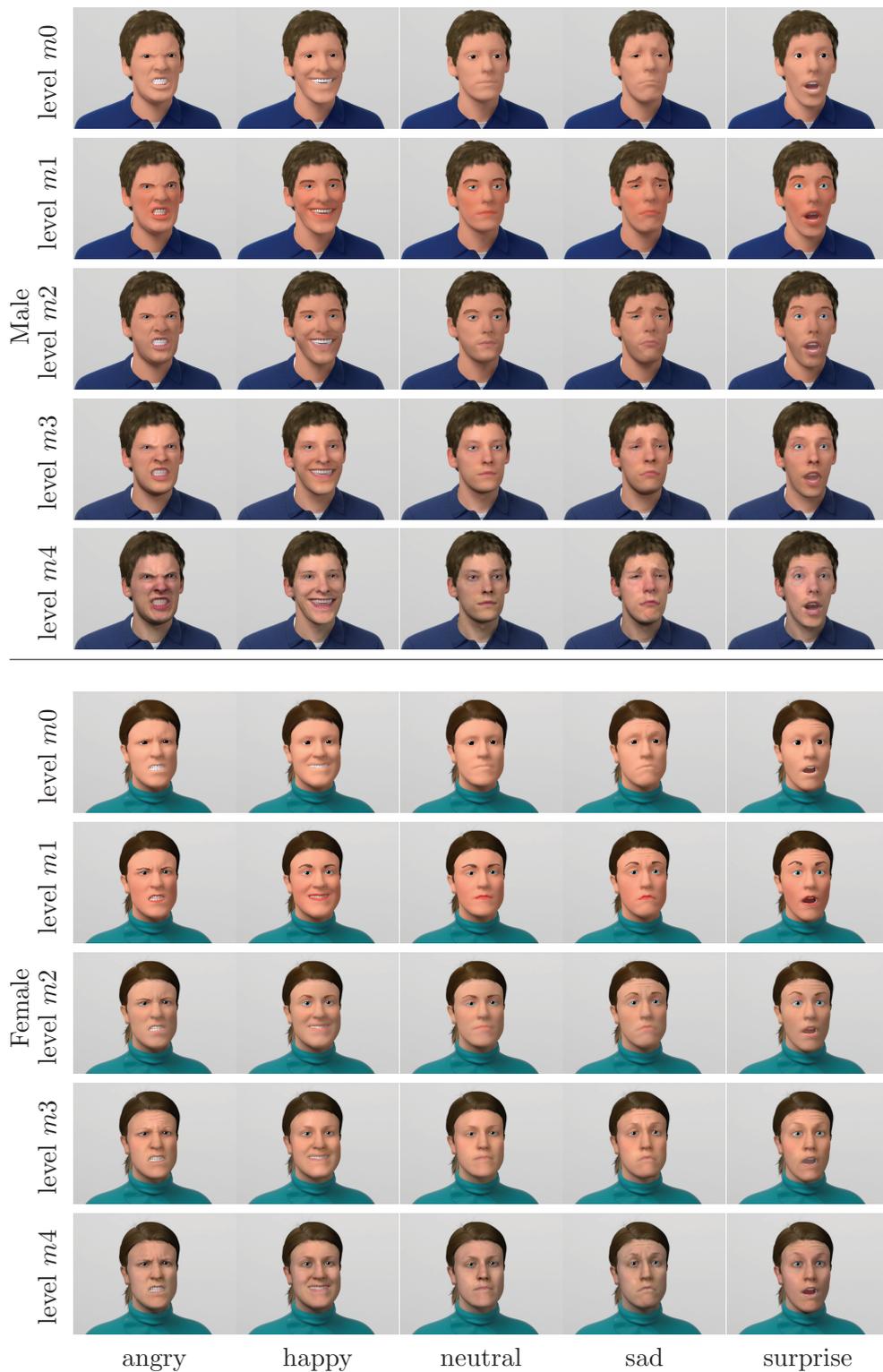
Figure A.13: Experiment 2. Five expressions (*x-axis*) of the most realistic shape (level *s4*) of both characters, combined with the five levels of material stylization (*y-axis*).

A.3 Task Description for Participants (in German)

A.3.1 Task Description for Experiments 1-2

In dem folgenden Experiment bitten wir Sie, unterschiedliche Charaktere zu beurteilen. Als Charaktere bezeichnen wir realistische oder abstrakte Figuren, die in einem Film oder Animationsfilm auftauchen. Jeder Charakter wird Ihnen für jeweils drei Sekunden gezeigt. Im Anschluss bitten wir Sie, die gezeigte Figur nach den folgenden Kriterien zu beurteilen. Beim Beurteilen können Sie jederzeit auf diese Kriterienbeschreibung zurückgreifen.

Kriterienbeschreibung

- *abstoßend - ansprechend*: Abstoßend bedeutet, dass Ihnen der Charakter nicht gefällt und Sie ihn nur ungern anschauen. Einen ansprechenden Charakter empfinden Sie hingegen als angenehm und schauen ihn gerne an.

- *unheimlich - beruhigend*: Ein unheimlicher Charakter erzeugt bei Ihnen Schauder, Angst oder Grusel. Beruhigend bedeutet hingegen, dass der Charakter in Ihnen Gefühle von Sicherheit, Ruhe und Vertrauen auslöst.

- *abstrakt - realistisch*: Beschreiben Sie, ob der gezeigte Charakter eher eine abstrakte Darstellung eines Menschen oder eine realistische Darstellung ist, die mit einem Foto vergleichbar ist.

- *ungewohnt - vertraut*: Ungewohnt bedeutet, dass Sie noch nie einen vergleichbaren Charakter gesehen haben. Vertraut bedeutet hingegen, dass Sie ähnliche Charaktere schon öfter gesehen haben.

- *unattraktiv - attraktiv*: Beschreiben Sie, ob der gezeigte Charakter unattraktiv oder hässlich ist, oder ob Sie ihn als schön und attraktiv bezeichnen würden.

A.3.2 Task Description for Experiment 3

In dem folgenden Experiment bitten wir Sie, den Gesichtsausdruck unterschiedlicher Charaktere zu beurteilen. Als Charaktere bezeichnen wir realistische oder abstrakte Figuren, die in einem Film oder Animationsfilm auftauchen. Jeder Charakter wird Ihnen für jeweils vier Sekunden gezeigt. Danach sollen Sie den Gesichtsausdruck mit dem aus Ihrer Sicht passendsten Adjektiv beschreiben. Ihnen stehen dazu die folgenden Adjektive zur Verfügung:

- wütend

- fröhlich

- neutral

- traurig

- überrascht

Im Anschluss sollen Sie bewerten, wie stark dieser Gesichtsausdruck ausgeprägt ist. Die zwei folgenden Beispiele sollen die Aufgabe nochmal verdeutlichen.

Bsp 1: Sie empfanden, die Figur wirkte sehr überrascht. Also wählen Sie als erste Antwort, dass sie überrascht ist und als zweite Antwort, dass der Gesichtsausdruck sehr stark ist.

Bsp 2: Sie empfanden, die Figur wirkte hauptsächlich neutral und ein klein bisschen fröhlich. Also wählen Sie als erste Antwort neutral aus und als zweite Antwort, dass der Gesichtsausdruck mäßig stark ausgeprägt ist.

List of Figures

Name	Icon	Copyright
Billy		©Jana Bergevin
Dave		©TurboSquid
Face Scan 1		©Beeler et al. (2010)
Face Scan 2		©Infinite Realities
Face Template		©Chen et al. (2014)
Loki		©Mark Pauly
Mery		©http://meryproject.com
Mono		free character from Freebie 3D
Osipa		©Jason Osipa
Slimer		©Weise et al. (2011)
Viktor		©Faceware Technologies
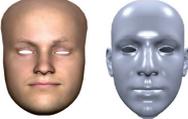
Table A.1: List of all third party characters and their copyright owners.

List of Tables

Bibliography

Abdel Rahman, R. (2011), 'Facing good and evil: Early brain signatures of affective biographical knowledge in face recognition', *Emotion* **11**(6), 1397–1405.

Abdel Rahman, R. & Sommer, W. (2012), 'Knowledge scale effects in face recognition: An electrophysiological investigation', *Cognitive, Affective and Behavioral Neuroscience* **12**(1), 161–174.

Achenbach, J., Zell, E. & Botsch, M. (2015), Accurate face reconstruction through anisotropic fitting and eye correction, *in* 'Proc. of Vision, Modeling and Visualization', pp. 1–8.

Aigerman, N., Poranne, R. & Lipman, Y. (2014), 'Lifted bijections for low distortion surface mappings', *ACM Transactions on Graphics* **33**(4), 69:1–69:12.

Aigerman, N., Poranne, R. & Lipman, Y. (2015), 'Seamless surface mappings', *ACM Transactions on Graphics* **34**(4), 72:1–72:13.

Alexa, M. (2002), 'Recent advances in mesh morphing', *Computer Graphics Forum* **21**(2), 173–198.

Alexa, M. (2003), 'Differential coordinates for mesh morphing and deformation', *The Visual Computer* **19**(2), 105–114.

Alexander, O., Rogers, M., Lambeth, W., Chiang, J.-Y., Ma, W.-C., Wang, C.-C. & Debevec, P. (2010), 'The Digital Emily project: Achieving a photorealistic digital actor', *IEEE Computer Graphics and Applications* **30**(4), 20–31.

Allen, B., Curless, B. & Popović, Z. (2003), 'The space of human body shapes: Reconstruction and parametrization from range scans', *ACM Transactions on Graphics* **22**(3), 587–594.

Alley, T. R. & Cunningham, M. R. (1991), 'Averaged faces are attractive, but very attractive faces are not average', *Psychological Science* **2**(2), 123–125.

Amberg, B., Romdhani, S. & Vetter, T. (2007), Optimal step nonrigid ICP algorithms for surface registration, *in* 'Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)', pp. 1–8.

Anjyo, K., Todo, H. & Lewis, J. P. (2012), 'A practical approach to direct manipulation blendshapes', *Journal of Graphics Tools* **16**(3), 160–176.

Asthana, A., Zafeiriou, S., Cheng, S. & Pantic, M. (2013), Robust discriminative response map fitting with constrained local models, *in* 'Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)'.

Balas, B. & Pacella, J. (2015), 'Artificial faces are harder to remember', *Computers in Human Behavior* **52**, 331–337.

Barla, P., Thollot, J. & Markosian, L. (2006), X-toon: An extended toon shader, *in* 'Proc. of Symp. on Non-photorealistic animation and rendering (NPAR)', pp. 127–132.

Barrielle, V., Stoiber, N. & Cagniart, C. (2016), 'BlendForces: A dynamic framework for facial animation', *Computer Graphics Forum* **35**(2), 341–352.

Beckmann, P. & Spizzichino, S. (1963), *The Scattering of Electromagnetic Waves from Rough Surfaces*, MacMillan.

Beeler, T., Bickel, B., Beardsley, P., Sumner, B. & Gross, M. (2010), 'High-quality single-shot capture of facial geometry', *ACM Transactions on Graphics* **29**(4), 40:1–40:9.

Beeler, T., Bickel, B., Noris, G., Beardsley, P., Marschner, S., Sumner, R. W. & Gross, M. (2012), 'Coupled 3D reconstruction of sparse facial hair and skin', *ACM Transactions on Graphics* **31**(4), 117:1–117:10.

Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R. W. & Gross, M. (2011), 'High-quality passive facial performance capture using anchor frames', *ACM Transactions on Graphics* **30**(4), 75:1–75:10.

Belkin, M. & Niyogi, P. (2005), Towards a theoretical foundation for Laplacian-based manifold methods, *in* 'Proc. Conf. on Learning Theory', pp. 486–500.

Benson, P. & Perrett, D. (1992), 'Face to face with the perfect image', *New Scientist* **133**, 1809.

Bentin, S., Allison, T., Puce, A., Perez, E. & McCarthy, G. (1996), 'Electrophysiological studies of face perception in humans', *Journal of Cognitive Neuroscience* **8**(6), 551–565.

Bérard, P., Bradley, D., Gross, M. & Beeler, T. (2016), 'Lightweight eye capture using a parametric model', *ACM Transactions on Graphics* **35**(4), 117:1–117:12.

Bérard, P., Bradley, D., Nitti, M., Beeler, T. & Gross, M. (2014), 'High-quality capture of eyes', *ACM Transactions on Graphics* **33**(6), 223:1–223:12.

Berger, H. (1929), 'Über das Elektrenkephalogramm des Menschen', *Archiv für Psychiatrie und Nervenkrankheiten* **85**, 527–570.

Bermano, A., Beeler, T., Kozlov, Y., Bradley, D., Bickel, B. & Gross, M. (2015), 'Detailed spatio-temporal reconstruction of eyelids', *ACM Transactions on Graphics* **34**(4), 44:1–44:11.

Besl, P. J. & McKay, N. D. (1992), 'A method for registration of 3D shapes', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 239–256.

Bickel, B., Botsch, M., Angst, R., Matusik, W., Otaduy, M., Pfister, H. & Gross, M. (2007), 'Multi-scale capture of facial geometry and motion', *ACM Transactions on Graphics* **26**(3), 33:1–33:10.

Bishop, S. J. (2008), 'Neural mechanisms underlying selective attention to threat', *Annals of the New York Academy of Sciences* **1129**(1), 141–152.

Blanz, V. & Vetter, T. (1999), A morphable model for the synthesis of 3D faces, *in* 'Proc. of ACM SIGGRAPH', pp. 187–194.

Blausen (2014), 'Medical gallery of Blausen Medical', *WikiJournal of Medicine* **1**(2).

Blinn, J. F. (1977), 'Models of light reflection for computer synthesized pictures', *SIGGRAPH Computer Graphics* **11**(2), 192–198.

Bonarrigo, F., Signoroni, A. & Botsch, M. (2014), 'Deformable registration using patch-wise shape matching', *Graphical Models* **76**(5), 554–565.

Botsch, M. & Sorkine, O. (2008), 'On linear variational surface deformation methods', *IEEE Transactions on Visualization and Computer Graphics* **14**(1), 213–230.

Bouaziz, S. & Pauly, M. (2014), Semi-supervised facial animation retargeting, Technical Report 202143, EPFL.

Bouaziz, S., Tagliasacchi, A. & Pauly, M. (2014), Dynamic 2D/3D registration, *in* 'Eurographics Tutorial'.

Bouaziz, S., Wang, Y. & Pauly, M. (2013), 'Online modeling for realtime facial animation', *ACM Transactions on Graphics* **32**(4), 40:1–40:10.

Bradley, D., Heidrich, W., Popa, T. & Sheffer, A. (2010), 'High resolution passive facial performance capture', *ACM Transactions on Graphics* **29**(4), 41:1–41:10.

Bregler, C., Loeb, L., Chuang, E. & Deshpande, H. (2002), 'Turning to the masters: Motion capturing cartoons', *ACM Transactions on Graphics* **21**(3), 399–407.

Bronstein, A. M., Bronstein, M. M. & Kimmel, R. (2006), 'Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching', *Proc. of the National Academy of Sciences (PNAS)* **103**(5), 1168–1172.

Bronstein, A. M., Bronstein, M. M. & Kimmel, R. (2008), *Numerical Geometry of Non-Rigid Shapes*, Springer.

Brown, B. J. & Rusinkiewicz, S. (2007), 'Global non-rigid alignment of 3D scans', *ACM Transactions on Graphics* **26**(3).

Bruce, V. & Young, A. (1986), 'Understanding face recognition', *British Journal of Psychology* **77**(3), 305–327.

Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M. & Alpers, G. W. (2014), 'Social and emotional relevance in face processing: Happy faces of future interaction partners enhance the late positive potential', *Frontiers Human Neuroscience* **8**, 493.

Buck, I., Finkelstein, A., Jacobs, C., Klein, A., Salesin, D. H., Seims, J., Szeliski, R. & Toyama, K. (2000), Performance-driven hand-drawn animation, *in* 'Proc. Symp. on Non-Photorealistic Animation and Rendering (NPAR)', pp. 101–108.

Bui, T. D., Poel, M., Heylen, D. & Nijholt, A. (2003), Automatic face morphing for transferring facial animation, *in* 'Proc. of IASTED Int. Conf. on Computers, Graphics, and Imaging', pp. 19–24.

Burleigh, T. J., Schoenherr, J. R. & 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), 759771.

Burley, B. (2012), Physically based shading at Disney, *in* 'ACM SIGGRAPH Course: Practical Physically Based Shading in Film and Game Production'.

Burley, B. (2015), Extending the Disney BRDF to a BSDF with integrated subsurface scattering, *in* 'ACM SIGGRAPH Course: Physically Based Shading in Theory and Practice'.

Calder, A. J. (1996), 'Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear', *Cognitive Neuropsychology* **13**(5), 699–745.

Cao, C., Bradley, D., Zhou, K. & Beeler, T. (2015), 'Real-time high-fidelity facial performance capture', *ACM Transactions on Graphics* **34**(4), 46:1–46:9.

Cao, C., Hou, Q. & Zhou, K. (2014), 'Displaced dynamic expression regression for real-time facial tracking and animation', *ACM Transactions on Graphics* **33**(4), 43:1–43:10.

Chai, M., Luo, L., Sunkavalli, K., Carr, N., Hadap, S. & Zhou, K. (2015), 'High-quality hair modeling from a single portrait photo', *ACM Transactions on Graphics* **34**(6), 204:1–204:10.

Chai, M., Shao, T., Wu, H., Weng, Y. & Zhou, K. (2016), 'Autohair: Fully automatic hair modeling from a single image', *ACM Transactions on Graphics* **35**(4), 116:1–116:12.

Chai, M., Wang, L., Weng, Y., Yu, Y., Guo, B. & Zhou, K. (2012), 'Single-view hair modeling for portrait manipulation', *ACM Transactions on Graphics* **31**(4), 116:1–116:8.

Chang, W., Li, H., Mitra, N., Pauly, M. & Wand, M. (2010), Geometric registration for deformable shapes, *in* 'Eurographics Course'.

Cheetham, M., Suter, P. & Jäncke, L. (2011), 'The human likeness dimension of the Uncanny Valley hypothesis: Behavioral and functional MRI findings', *Frontiers in Human Neuroscience* **5**, 126.

Chen, C., Weng, Y., Zhou, S., Tong, Y. & Zhou, K. (2014), 'FaceWarehouse: a 3D facial expression database for visual computing', *IEEE Transactions on Visualization and Computer Graphics* **20**(3), 413–425.

Chen, Y., Davis, T. A., Hager, W. W. & Rajamanickam, S. (2008), 'Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate', *ACM Transactions on Mathematical Software* **35**(3), 1–14.

Chen, Y. & Medioni, G. (1992), 'Object modeling by registration of multiple range images', *Image and Vision Computing* **10**(3), 145–155.

Chuang, E. & Bregler, C. (2002), Performance driven animation using blendshape interpolation, Technical Report CS-TR-2002-02, Stanford University.

Codispoti, M., Vera, F., Junghfer, M. & Schupp, H. T. (2006), 'The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs', *NeuroImage* **32**(2), 583–591.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*, Lawrence Erlbaum Associates, Inc.

Coleman, P., Bibliowicz, J., Singh, K. & Gleicher, M. (2008), Staggered poses: A character motion representation for detail-preserving editing of pose and coordinated timing, *in* 'Proc. Symp. on Computer Animation (SCA)', pp. 137–146.

Conde, M. H., Shahlaei, D., Blanz, V. & Loffeld, O. (2015), Efficient and robust inverse lighting of a single face image using compressive sensing, *in* 'Computer Vision Workshop (ICCVW), IEEE Int. Conf. on Inverse Rendering Workshop'.

Crookes, K., Ewing, L., dith Gildenhuys, J., Kloth, N., Hayward, W. G., Oxner, M., Pond, S. & Rhodes, G. (2015), 'How well do computer-generated faces tap face expertise?', *PloS One* **10**(11), 1–18.

Cui, Z., Shan, S., Zhang, H., Lao, S. & Chen, X. (2012), Image sets alignment for video-based face recognition, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 2626–2633.

Cunningham, D. & Wallraven, C. (2011), *Experimental Design: From User Studies to Psychophysics*, AK Peters / CRC Press.

Cunningham, M. R. (1986), 'Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty', *Journal of Personality and Social Psychology* **50**(5), 925–935.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N. & Lang, P. J. (2000), 'Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report', *Biological Psychology* **52**(2), 95–111.

de Borst, A. W. & de Gelder, B. (2015), 'Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective', *Frontiers in Psychology* **6**, 576.

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L. & Lueschow, A. (2007), 'MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus', *NeuroImage* **35**(4), 1495–1501.

Deng, Z., Chiang, P.-Y., Fox, P. & Neumann, U. (2006), Animating blendshape faces by cross-mapping motion capture data, *in* 'Proc. Symp. on Interactive 3D Graphics and Games (i3D)', pp. 43–48.

d'Eon, E., François, G., Hill, M., Letteri, J. & Aubry, J.-M. (2011), An energy-conserving hair reflectance model, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 1181–1187.

d'Eon, E., Luebke, D. & Enderton, E. (2007), Efficient rendering of human skin, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 147–158.

Desbrun, M., Meyer, M., Schröder, P. & Barr, A. H. (1999), Implicit fairing of irregular meshes using diffusion and curvature flow, *in* 'Proc. of ACM SIGGRAPH', pp. 317–324.

Dill, V., Flach, L. M., Hocevar, R., Lykawka, C., Musse, S. R. & Pinho, M. S. (2012), Evaluation of the Uncanny Valley in CG characters, *in* 'Proc. of Intelligent Virtual Agents', pp. 511–513.

Donner, C. & Jensen, H. W. (2005), 'Light diffusion in multi-layered translucent materials', *ACM Transactions on Graphics* **24**(3), 1032–1039.

Donner, C. & Jensen, H. W. (2006), A spectral BSSRDF for shading human skin, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 409–417.

Donner, C., Weyrich, T., d'Eon, E., Ramamoorthi, R. & Rusinkiewicz, S. (2008), 'A layered, heterogeneous reflectance model for acquiring and rendering human skin', *ACM Transactions on Graphics* **27**(5), 140:1–140:12.

Dubal, S., Foucher, A., Jouvent, R. & Nadel, J. (2011), 'Human brain spots emotion in non–humanoid robots', *Social Cognitive and Affective Neuroscience* **6**(1), 90–97.

Eimer, M. (2000), 'Event-related brain potentials distinguish processing stages involved in face perception and recognition', *Clinical Neurophysiology: Official Journal of the Int. Federation of Clinical Neurophysiology* **111**(4), 694–705.

Eimer, M., Holmes, A. & McGlone, F. P. (2003), 'The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions', *Cognitive, Affective and Behavioral Neuroscience* **3**(2), 97–110.

Ekman, P. (1972), 'Universal and cultural differences in facial expression of emotion', *Proc. of Nebraska Symp. on Motivation* **19**, 207–284.

Ekman, P. & Friesen, W. V. (1978), *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press.

Fan, K., Mian, A., Liu, W. & Li, L. (2016), 'Unsupervised manifold alignment using soft-assign technique', *Machine Vision and Applications* **27**(6), 929–942.

Ferrey, A. E., Burleigh, T. J. & Fenske, M. J. (2015), 'Stimulus-category competition, inhibition, and affective devaluation: a novel account of the Uncanny Valley', *Frontiers in Psychology* **6**, 249.

Fink, B., Bunse, L., Matts, P. J. & D'Emiliano, D. (2012), 'Visible skin colouration predicts perception of male facial age, health and attractiveness', *Int. Journal of Cosmetic Science* **34**(4), 307–310.

Fink, B., Grammer, K. & 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), 845–857.

Fink, B. & 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), 493–498.

Flaisch, T., Häcker, F., Renner, B. & Schupp, H. T. (2011), 'Emotion and the processing of symbolic gestures: An event-related brain potential study', *Social Cognitive and Affective Neuroscience* **6**(1), 109–118.

Fleming, R., Mohler, B. J., Romero, J., Black, M. J. & Breidt, M. (2016), Appealing female avatars from 3D body scans: Perceptual effects of stylization, *in* 'Int. Conf. on Computer Graphics Theory and Applications (GRAPP)'.

Fleming, R. W. & Bülthoff, H. H. (2005), 'Low-level image cues in the perception of translucent materials', *ACM Transactions on Applied Perception* **2**(3), 346–382.

Fleming, R. W., Dror, R. O. & Adelson, E. H. (2003), 'Real-world illumination and the perception of surface reflectance properties', *Journal of Vision* **3**(6), 347–368.

François, G., Gautron, P., Breton, G. & Bouatouch, K. (2009), 'Image-based modeling of the human eye', *IEEE Transactions on Visualization and Computer Graphics* **15**(5), 815–827.

Fyffe, G., Graham, P., Tunwattanapong, B., Ghosh, A. & Debevec, P. (2016), 'Near-instant capture of high-resolution facial geometry and reflectance', *Computer Graphics Forum* **35**(2), 353–363.

Gajewski, P. D. & Stoerig, P. (2011), 'N170 – an index of categorical face perception? An ERP study of human, nonhuman primate, and dog faces', *Journal of Psychophysiology* **25**(4), 174179.

Garrido, P., Zollhöfer, M., Casas, D., Valgaerts, L., Varanasi, K., Pérez, P. & Theobalt, C. (2016), 'Reconstruction of personalized 3D face rigs from monocular video', *ACM Transactions on Graphics* **35**(3), 28:1–28:15.

Geller, T. (2008), 'Overcoming the Uncanny Valley', *IEEE Computer Graphics and Applications* **28**(4), 11–17.

Ghosh, A., Fyffe, G., Tunwattanapong, B., Busch, J., Yu, X. & Debevec, P. (2011), 'Multiview face capture using polarized spherical gradient illumination', *ACM Transactions on Graphics* **30**(6), 129:1–129:10.

Goldman, D. B. (1997), Fake fur rendering, *in* 'Proc. of ACM SIGGRAPH', pp. 127–134.

González-Roldan, A. M., Martínez-Jauand, M., Muñoz-García, M. A., Sitges, C., Cifre, I. & Montoya, P. (2011), 'Temporal dissociation in the brain processing of pain and anger faces with different intensities of emotional expression', *PAIN* **152**(4), 853–859.

Graham, P., Tunwattanapong, B., Busch, J., Yu, X., Jones, A., Debevec, P. & Ghosh, A. (2013), 'Measurement-based synthesis of facial microgeometry', *Computer Graphics Forum* **32**(2), 335–344.

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

Guennebaud, G., Jacob, B. et al. (2016), 'Eigen v3.3'.
URL: *http://eigen.tuxfamily.org*

Habel, R., Christensen, P. H. & Jarosz, W. (2013*a*), Classical and improved diffusion theory for subsurface scattering, Technical report, Disney Research Zürich.

Habel, R., Christensen, P. H. & Jarosz, W. (2013*b*), 'Photon beam diffusion: A monte carlo method for subsurface scattering', *Computer Graphics Forum* **32**(4), 27–37.

Hagmann, P., Cammoun, L., Gigandet, X., Meul, R., Honey, C. J., Wedeen, V. J. & Sporns, O. (2008), 'Mapping the structural core of human cerebral cortex', *PLoS Biology* **6**(7), e159.

Hanson, D. (2005), Expanding the aesthetics possibilities for humanlike robots, *in* 'Proc. of IEEE Humanoid Robotics Conf., special session on the Uncanny Valley'.

Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. (2000), 'The distributed human neural system for face perception', *Trends in Cognitive Sciences* **4**(6), 223–233.

Herrera, T. L., Zinke, A. & Weber, A. (2011), Towards hairstyle reconstruction using thermal imaging, *in* 'SIGGRAPH Asia Sketches', pp. 13:1–13:2.

Herrera, T. L., Zinke, A., Weber, A. & Vetter, T. (2010), Toward image-based facial hair modeling, *in* 'Proc. of Spring Conf. on Computer Graphics (SCCG)', pp. 93–100.

Hill, S., McAuley, S., Burley, B., Chan, D., Fascione, L., Iwanicki, M., Hoffman, N., Jakob, W., Neubelt, D., Pesce, A. & Pettineo, M. (2015), Physically based shading in theory and practice, *in* 'ACM SIGGRAPH Courses', pp. 22:1–22:8.

Hinojosa, J. A., Mercado, F. & Carreti, L. (2015), 'N170 sensitivity to facial expression: A meta-analysis', *Neuroscience and Biobehavioral Reviews* **55**, 498–509.

Ho, C.-C. & MacDorman, K. F. (2010), 'Revisiting the Uncanny Valley theory: Developing and validating an alternative to the godspeed indices', *Computers in Human Behavior* **26**(6), 1508–1518.

Ho, Y.-X., Landy, M. S. & Maloney, L. T. (2006), 'How direction of illumination affects visually perceived surface roughness', *Journal of Vision* **6**(8), 634–648.

Horn, B. K. P. (1987), 'Closed-form solution of absolute orientation using unit quaternions', *Journal of the Optical Society of America* **4**(4), 629–642.

Hu, L., Ma, C., Luo, L. & Li, H. (2015), 'Single-view hair modeling using a hairstyle database', *ACM Transactions on Graphics* **34**(4), 125:1–125:9.

Huang, Q.-X., Adams, B., Wicke, M. & Guibas, L. J. (2008), 'Non-rigid registration under isometric deformations', *Computer Graphics Forum* **27**(5), 1449–1457.

Ichim, A.-E., Bouaziz, S. & Pauly, M. (2015), 'Dynamic 3D avatar creation from hand-held video input', *ACM Transactions on Graphics* **34**(4), 45:1–45:14.

Ichim, A.-E., Kavan, L., Nimier-David, M. & Pauly, M. (2016), Building and animating user-specific volumetric face rigs, *in* 'Proc. Symp. on Computer Animation (SCA)', pp. 107–117.

Iglesias-Guitian, J. A., Aliaga, C., Jarabo, A. & Gutierrez, D. (2015), 'A biophysically-based model of the optical properties of skin aging', *Computer Graphics Forum* **34**(2), 45–55.

Ille, N., Berg, P. & Scherg, M. (2002), 'Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies', *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society* **19**(2), 113–124.

Itz, M. L., Schweinberger, S. R., Schulz, C. & Kaufmann, J. M. (2014), 'Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance', *NeuroImage* **105**, 736–747.

Jachnik, J., Goldman, D. B., Luo, L. & Davison, A. J. (2015), 'Interactive 3D face stylization using sculptural abstraction', *arXiv:1502.01954 [cs.GR]* .
URL: *arxiv.org/abs/1502.01954*

Jack, R. E., Garrod, M. O. & Schyns, P. G. (2014), 'Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time', *Current Biology* **24**(2), 187–192.

Jack, R. E. & Schyns, P. G. (2015), 'The human face as a dynamic tool for social communication', *Current Biology* **25**(14), R621–R634.

Jakob, W., Moon, J. T. & Marschner, S. (2009), 'Capturing hair assemblies fiber by fiber', *ACM Transactions on Graphics* **28**(5), 164:1–164:9.

Jarabo, A., Eyck, T. V., Sundstedt, V., Bala, K., Gutierrez, D. & O'Sullivan, C. (2012), 'Crowd light: Evaluating the perceived fidelity of illuminated dynamic scenes', **31**(2), 565–574.

Jensen, H. W., Marschner, S. R., Levoy, M. & Hanrahan, P. (2001), A practical model for subsurface light transport, *in* 'Proc. of SIGGRAPH', pp. 511–518.

Jimenez, J., Danvoye, E. & von der Pahlen, J. (2013), 'Next generation character rendering', *Game Developer Conf. 2013* .

Jimenez, J., Scully, T., Barbosa, N., Donner, C., Alvarez, X., Vieira, T., Matts, P., Orvalho, V., Gutierrez, D. & Weyrich, T. (2010), 'A practical appearance model for dynamic facial color', *ACM Transactions on Graphics* **29**(6), 141:1–141:10.

Jimenez, J., Sundstedt, V. & Gutierrez, D. (2009), 'Screen-space perceptual rendering of human skin', *ACM Transactions on Applied Perception* **6**(4), 23:1–23:15.

Jimenez, J., Zsolnai, K., Jarabo, A., Freude, C., Auzinger, T., Wu, X.-C., von der Pahlen, J., Wimmer, M. & Gutierrez, D. (2015), 'Separable subsurface scattering', *Computer Graphics Forum* **34**(5), 188–197.

Johnston, V. S. & Oliver-Rodriguez, J. C. (1997), 'Facial beauty and the late positive component of event-related potentials', *Journal of Sex Research* **34**(2), 188–198.

Jones, E., Oliphant, T., Peterson, P. et al. (2001), 'SciPy: Open source scientific tools for Python'.
URL: *http://www.scipy.org/*

Kajiya, J. T. & Kay, T. L. (1989), 'Rendering fur with three dimensional textures', *SIGGRAPH Computer Graphics* **23**(3), 271–280.

Kanwisher, N., McDermott, J. & Chun, M. M. (1997), 'The fusiform face area: A module in human extrastriate cortex specialized for face perception', *The Journal of Neuroscience* **17**, 4302–4311.

Karis, B. (2013), 'Real shading in Unreal Engine 4', *SIGGRAPH Course: Physically Based Shading in Theory and Practice* .

Karis, B. (2016), 'Physically based hair shading in Unreal', *SIGGRAPH Course: Physically Based Shading in Theory and Practice* .

Kätsyri, J., Förger, K., Mäkäräinen, M. & Takala, T. (2015), 'A review of empirical evidence on different Uncanny Valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness', *Frontiers in Psychology* **6**, 390.

Kaufmann, J. M. & Schweinberger, S. R. (2012), 'The faces you remember: Caricaturing shape facilitates brain processes reflecting the acquisition of new face representations', *Biol. Psychol.* **89**, 21–33.

Kazhdan, M., Solomon, J. & Ben-Chen, M. (2012), 'Can mean-curvature flow be modified to be non-singular?', *Computer Graphics Forum* **31**(5), 1745–1754.

Kelemen, C. & Szirmay-Kalos, L. (2001), 'A microfacet based coupled specular-matte BRDF model with importance sampling', *Eurographics Short Presentations* .

Keller, A. (1997), Instant radiosity, *in* 'Proc. of ACM SIGGRAPH', pp. 49–56.

Kholgade, N., Matthews, I. & Sheikh, Y. (2011), Content retargeting using parameter-parallel facial layers, *in* 'Proc. Symp. on Computer Animation (SCA)', pp. 195–204.

Kim, V. G., Lipman, Y. & Funkhouser, T. (2011), 'Blended intrinsic maps', *ACM Transactions on Graphics* **30**(4), 79:1–79:12.

Klehm, O., Rousselle, F., Papas, M., Bradley, D., Hery, C., Bickel, B., Jarosz, W. & Beeler, T. (2015), 'Recent advances in facial appearance capture', *Computer Graphics Forum* **34**(2), 709–733.

Kolluri, R., Shewchuk, J. R. & O'Brien, J. F. (2004), Spectral surface reconstruction from noisy point clouds, *in* 'Proc. Symp. of Geometry Processing', pp. 11–21.

Kraevoy, V. & Sheffer, A. (2004), 'Cross-parametrization and compatible remeshing of 3D models', *ACM Transactions on Graphics* **23**(3), 861–869.

Kraevoy, V., Sheffer, A. & Gotsman, C. (2003), 'Matchmaker: Constructing constrained texture maps', *ACM Transactions on Graphics* **22**(3), 326–333.

Křivánek, J., Ferwerda, J. A. & Bala, K. (2010), 'Effects of global illumination approximations on material appearance', *ACM Transactions on Graphics* **29**(4), 112:1–112:10.

Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., De Luca, E. W. & Albayrak, S. (2010), Spectral analysis of signed graphs for clustering, prediction and visualization, *in* 'Proc. Int. Conf. on Data Mining'.

Kyprianidis, J. E., Collomosse, J., Wang, T. & Isenberg, T. (2013), 'State of the art: A taxonomy of artistic stylization techniques for images and video', *IEEE Transactions on Visualization and Computer Graphics* **19**(5), 866–885.

Lam, M. W. & Baranoski, G. V. (2006), 'A predictive light transport model for the human iris', *Computer Graphics Forum* **25**(3), 359–368.

Lau, M., Chai, J., Xu, Y.-Q. & Shum, H.-Y. (2009), 'Face Poser: Interactive modeling of 3D facial expressions using facial priors', *ACM Transactions on Graphics* **29**(1), 3:1–3:17.

Lee, W.-S. & Magnenat-Thalmann, N. (2000), 'Fast head modeling for animation', *Journal of Image and Vision Computing* **18**(4), 355–364.

Lefohn, A., Budge, B., Shirley, P., Caruso, R. & Reinhard, E. (2003), 'An ocularist's approach to human iris synthesis', *IEEE Computer Graphics and Application* **23**(6), 70–75.

Lévy, B. (2001), Constrained texture mapping for polygonal meshes, *in* 'Proc. of ACM SIGGRAPH', pp. 417–424.

Lewis, J. P., Anjyo, K., Rhee, T., Zhang, M., Pighin, F. & Deng, Z. (2014), Practice and theory of blendshape facial models, *in* 'Eurographics State of the Art Reports'.

Lewis, J. P., Mooser, J., Deng, Z. & Neumann, U. (2005), Reducing blendshape interference by selected motion attenuation, *in* 'Proc. Symp. on Interactive 3D Graphics and Games (i3D)', pp. 25–29.

Li, H., Adams, B., Guibas, L. J. & Pauly, M. (2009), 'Robust single-view geometry and motion reconstruction', *ACM Transactions on Graphics* **28**(5), 175:1–175:10.

Li, H., Sumner, R. W. & Pauly, M. (2008), 'Global correspondence optimization for non-rigid registration of depth scans', *Computer Graphics Forum* **27**(5), 1421–1430.

Li, H., Weise, T. & Pauly, M. (2010), 'Example-based facial rigging', *ACM Transactions on Graphics* **29**(4), 32:1–32:6.

Li, H., Yu, J., Ye, Y. & Bregler, C. (2013), 'Realtime facial animation with on-the-fly correctives', *ACM Transactions on Graphics* **32**(4), 42:1–42:10.

Lipman, Y. & Funkhouser, T. (2009), 'Möbius voting for surface correspondence', *ACM Transactions on Graphics* **28**(3), 72:1–72:12.

Little, A. C. & Hancock, P. J. (2002), 'The role of masculinity and distinctiveness in judgements of human male facial attractiveness', *British Journal of Psychology* **93**(4), 451–464.

Litvak, V. & Friston, K. (2008), 'Electromagnetic source reconstruction for group studies', *Neuroimage* **42**, 1490–1498.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W. & Friston, K. (2011), 'EEG and MEG data analysis in SPM8', *Computational Intelligence and Neuroscience* **2011**(861705), 1–32.

Looser, C. E. & Wheatley, T. (2010), 'The tipping point of animacy how, when, and where we perceive life in a face', *Psychological Science* **21**(12), 1854–1862.

Lu, Y., Wang, J., Wang, L., Wang, J. & Qin, J. (2014), 'Neural responses to cartoon facial attractiveness: An event-related potential study', *Neuroscience Bulletin* **30**(3), 441–450.

Luck, S. J. (2005), *An Introduction to the Event-Related Potential Technique*, MIT Press.

Luo, L., Li, H. & Rusinkiewicz, S. (2013), 'Structure-aware hair capture', *ACM Transactions on Graphics* **32**(4), 76:1–76:12.

Ma, W.-C., Hawkins, T., Peers, P., Chabert, C.-F., Weiss, M. & Debevec, P. (2007), Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 183–194.

MacDorman, K. F. (2006), Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the Uncanny Valley, *in* 'Proc. of ICCS/CogSci Long Symp.: Towards Social Mechanisms of Android Science'.

MacDorman, K. F., Green, R. D., Ho, C.-C. & Koch, C. T. (2009), 'Too real for comfort? Uncanny responses to computer generated faces', *Computers in Human Behavior* **25**(3), 695–710.

Marschner, S. R., Jensen, H. W., Cammarano, M., Worley, S. & Hanrahan, P. (2003), 'Light scattering from human hair fibers', *ACM Transactions on Graphics* **22**(3), 780–791.

Marzi, T. & Viggiano, M. P. (2010), 'When memory meets beauty: Insights from event-related potentials', *Biological Psychology* **84**(2), 192–205.

McAuley, S. & Hill, S. (2016), Physically based shading in theory and practice, *in* 'ACM SIGGRAPH Courses'.

McCloud, S. (1993), *Understanding Comics*, William Morrow Paperbacks.

McDonnell, R., Breidt, M. & 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), 91:1–91:11.

Meyer, M., Desbrun, M., Schröder, P. & Barr, A. H. (2003), Discrete differential-geometry operators for triangulated 2-manifolds, *in* 'Visualization and Mathematics III', pp. 35–57.

Mitchell, J. L., Francke, M. & Eng, D. (2007), Illustrative rendering in Team Fortress 2, *in* 'ACM SIGGRAPH Courses'.

Moon, J. T. & Marschner, S. R. (2006), 'Simulating multiple scattering in hair using a photon mapping approach', *ACM Transactions on Graphics* **25**(3), 1067–1074.

Moon, J. T., Walter, B. & Marschner, S. (2008), 'Efficient multiple scattering in hair using spherical harmonics', *ACM Transactions on Graphics* **27**(3), 31:1–31:7.

Mori, M., MacDorman, K. F. & Kageki, N. (2012), 'The Uncanny Valley [From the field]', *IEEE Robotics and Automation Magazine* **19**(2), 98–100.

Mühlberger, A., Wieser, M. J., Herrmann, M. J., Weyers, P., Tröger, C. & Pauli, P. (2009), 'Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons', *Journal of Neural Transmission* **116**(6), 735–746.

Nagano, K., Fyffe, G., Alexander, O., Barbiç, J., Li, H., Ghosh, A. & Debevec, P. (2015), 'Skin microstructure deformation with displacement map convolution', *ACM Transactions on Graphics* **34**(4), 109:1–109:10.

Ngan, A., Durand, F. & Matusik, W. (2005), Experimental analysis of BRDF models, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 117–126.

Nguyen, H. & Donnelly, W. (2006), Hair animation and rendering in the Nalu demo, *in* 'GPU Gems 2', pp. 361–380.

Nitzsche, B. & Rose, K. (2011), *Bodyshop: The Photoshop Retouching Guide for the Face and Body*, Wiley.

Noh, J.-y. & Neumann, U. (2001), Expression cloning, *in* 'Proc. of ACM SIGGRAPH', pp. 277–288.

Olkkonen, M. & Brainard, D. H. (2010), 'Perceived glossiness and lightness under real-world illumination', *Journal of Vision* **10**(9).

Olkkonen, M. & Brainard, D. H. (2011), 'Joint effects of illumination geometry and object shape in the perception of surface reflectance', *i-Perception* **2**, 1014–1034.

Olofsson, J. K., Nordin, S., Sequeira, H. & Polich, J. (2008), 'Affective picture processing: An integrative review of ERP findings', *Biological Psychology* **77**(3), 247–265.

Orvalho, V., Bastos, P., Parke, F., Oliveira, B. & Alvarez, X. (2012), A facial rigging survey, *in* 'Eurographics State of the Art Reports'.

Orvalho, V. C., Zacur, E. & Susin, A. (2008), 'Transferring the rig and animations from a character to different face models', *Computer Graphics Forum* **27**(8), 1997–2012.

Ovsjanikov, M., Merigot, Q., Memoli, F. & Guibas, L. (2010), 'One point isometric matching with the heat kernel', *Computer Graphics Forum (Proc. SGP)* **29**(5), 1555–1564.

Pan, S. J. & Yang, Q. (2010), 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering* **22**(10), 1345–1359.

Panozzo, D., Baran, I., Diamanti, O. & Sorkine-Hornung, O. (2013), 'Weighted averages on surfaces', *ACM Transactions on Graphics* **32**(4), 60:1–60:12.

Paris, S., Briceño, H. M. & Sillion, F. X. (2004), 'Capture of hair geometry from multiple images', *ACM Transactions on Graphics* **23**(3), 712–719.

Paris, S., Chang, W., Kozhushnyan, O. I., Jarosz, W., Matusik, W., Zwicker, M. & Durand, F. (2008), 'Hair photobooth: Geometric and photometric acquisition of real hairstyles', *ACM Transactions on Graphics* **27**(3), 30:1–30:9.

Parke, F. I. & Waters, K. (2008), *Computer Facial Animation*, AK Peters Ltd.

Pei, Y., Huang, F., Shi, F. & Zha, H. (2012), 'Unsupervised image matching based on manifold alignment', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(8), 1658–1664.

Pérez, P., Gangnet, M. & Blake, A. (2003), 'Poisson image editing', *ACM Transactions on Graphics* **22**(3), 313–318.

Peyk, P., De Cesarei, A. & Junghöfer, M. (2011), 'Electro magneto encephalograhy software: Overview and integration with other EEG/MEG toolboxes', *Computational Intelligence and Neuroscience* **2011**, 861705:1–11.

Pharr, M. & Humphreys, G. (2010), *Physically Based Rendering: From Theory to Implementation*, Morgan Kaufmann.

Phong, B. T. (1975), 'Illumination for computer generated pictures', *Communications of ACM* **18**(6), 311–317.

Pinkall, U. & Polthier, K. (1993), 'Computing discrete minimal surfaces and their conjugates', *Experimental Mathematics* **2**(1), 15–36.

Pitcher, D., Walsh, V. & Duchaine, B. (2011), 'The role of the occipital face area in the cortical face perception network', *Experimental Brain Research* **209**(4), 481–493.

Pitcher, D., Walsh, V., Yovel, G. & Duchaine, B. (2007), 'TMS evidence for the involvement of the right occipital face area in early face processing', *Current Biology* **17**(18), 1568–1573.

Polich, J. & Kok, A. (1995), 'Cognitive and biological determinants of P300: An integrative review', *Biological Psychology* **41**, 103–146.

Pont, S. C. & te Pas, S. F. (2006), 'Material – illumination ambiguities and the perception of solid objects', *Perception* **35**, 1331–1350.

Potter, T., Corneille, O., Ruys, K. I. & Rhodes, G. (2007), 'Just another pretty face: A multidimensional scaling approach to face attractiveness and variability', *Psychonomic Bulletin and Review* **14**(2), 368372.

Proverbio, A. M., Riva, F., Martin, E. & Zani, A. (2010), 'Face coding is bilateral in the female brain', *PLoS ONE* **5**(6), e11242.

Rademacher, P. M. (2002), *Measuring the Perceived Visual Realism of Images*, PhD Thesis.

Rademacher, P. M., Lengyel, J., Cutrell, E. & Whitted, T. (2001), Measuring the perception of visual realism in images, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 235–247.

Ren, Z., Zhou, K., Li, T., Hua, W. & Guo, B. (2010), 'Interactive hair rendering under environment lighting', *ACM Transactions on Graphics* **29**(4), 55:1–55:8.

Rusinkiewicz, S. & Levoy, M. (2001), Efficient variants of the ICP algorithm, *in* 'Proc. of Int. Conf. on 3D Digital Imaging and Modeling'.

Rustamov, R. M. (2010), 'Barycentric coordinates on surfaces', *Computer Graphics Forum* **29**(5), 1507–1516.

Sadeghi, I., Pritchett, H., Jensen, H. W. & Tamstorf, R. (2010), 'An artist friendly hair shading system', *ACM Transactions on Graphics* **29**(4), 56:1–56:10.

Sagiv, N. & Bentin, S. (2001), 'Structural encoding of human and schematic faces: Holistic and part-based processes', *Journal of Cognitive Neuroscience* **13**(7), 937–951.

Sahillioğlu, Y. & Kavan, L. (2016), 'Detail-preserving mesh unfolding for non-rigid shape retrieval', *ACM Transactions on Graphics* **35**(3), 27:1–27:11.

Saito, J. (2013), Smooth contact-aware facial blendshapes transfer, *in* 'Proc. Symp. on Digital Production (DigiPro)', pp. 7–12.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J. & 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), 413–422.

Scheuermann, T. (2004), Practical real-time hair rendering and shading, *in* 'SIGGRAPH Talks'.

Schindler, S. & Kissler, J. (2016*a*), 'People matter: Perceived sender identity modulates cerebral processing of socio-emotional language feedback', *NeuroImage* **134**, 160–169.

Schindler, S. & Kissler, J. (2016*b*), 'Selective visual attention to emotional words: Early parallel frontal and visual activations followed by interactive effects in visual cortex', *Human Brain Mapping* **37**, 3575–3587.

Schindler, S., Wegrzyn, M., Steppacher, I. & Kissler, J. (2015), 'Perceived communicative context and emotional content amplify visual word processing in the fusiform gyrus', *The Journal of Neuroscience* **35**(15), 6010–6019.

Schlick, C. (1994), 'An inexpensive BRDF model for physically-based rendering', *Computer Graphics Forum* **13**(3), 149–162.

Schneider, E., Wang, Y. & Yang, S. (2007), Exploring the Uncanny Valley with Japanese video game characters, *in* 'Situated Play: Proc. of Digital Games Research Association Conference'.

Schreiner, J., Asirvatham, A., Praun, E. & Hoppe, H. (2004), 'Inter-surface mapping', *ACM Transactions on Graphics* **23**(3), 870–877.

Schulz, C., Kaufmann, J. M., Kurt, A. & Schweinberger, S. R. (2012), 'Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces', *NeuroImage* **63**, 491–500.

Schupp, H. T., Flaisch, T., Stockburger, J. & Junghöfer, M. (2006), 'Emotion and attention: Event-related brain potential studies', *Progress in Brain Research* **156**, 31–51.

Schupp, H. T., Junghöfer, M., Weike, A. I. & Hamm, A. O. (2004), 'The selective processing of briefly presented affective pictures: An ERP analysis', *Psychophysiology* **41**, 441–449.

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J. & Hamm, A. O. (2004), 'The facilitated processing of threatening faces: An ERP analysis', *Emotion* **4**(2), 189–200.

Schupp, H. T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A. I. & Hamm, A. O. (2007), 'Selective visual attention to emotion', *The Journal of Neuroscience* **27**, 1082–1089.

Schyns, P. G., Petro, L. S. & Smith, M. L. (2007), 'Dynamics of visual information integration in the brain for categorizing facial expressions', *Current Biology* **17**(18), 1580–1585.

Seo, J., Irving, G., Lewis, J. P. & Noh, J. (2011), 'Compression and direct manipulation of complex blendshape models', *ACM Transactions on Graphics* **30**(6), 164:1–164:10.

Seol, Y., Lewis, J. P., Seo, J., Choi, B., Anjyo, K. & Noh, J. (2012), 'Spacetime expression cloning for blendshapes', *ACM Transactions on Graphics* **31**(2), 14:1–14:12.

Seol, Y., Ma, W.-C. & Lewis, J. P. (2016), Creating an actor-specific facial rig from performance capture, *in* 'Proc. Symp. on Digital Production (DigiPro)', pp. 13–17.

Seol, Y., Seo, J., Kim, P. H., Lewis, J. P. & Noh, J. (2011), 'Artist friendly facial animation retargeting', *ACM Transactions on Graphics* **30**(6), 162:1–162:10.

Seyama, J. & Nagayama, R. S. (2007), 'The Uncanny Valley: Effect of realism on the impression of artificial human faces', *Presence: Teleoperators and Virtual Environments* **16**(4), 337–351.

Seymor, M. (2014), 'The many faces of Maleficent', *FXGuide* .
  URL: *https://www.fxguide.com/featured/maleficent/*

Shahlaei, D. & Blanz, V. (2015), Realistic inverse lighting from a single 2D image of a face, taken under unknown and complex lighting, *in* 'IEEE Int. Conf. on Automatic Face and Gesture Recognition'.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M. & Toga, A. W. (2008), 'Construction of a 3D probabilistic atlas of human cortical structures', *NeuroImage* **39**(3), 1064–1080.

Shoemake, K. (1985), 'Animating rotation with quaternion curves', *SIGGRAPH Comput. Graph.* **19**(3), 245–254.

Shultz, S. & McCarthy, G. (2014), 'Perceived animacy influences the processing of human-like surface features in the fusiform gyrus', *Neuropsychologia* **60**, 115–120.

Song, J., Choi, B., Seol, Y. & Noh, J. (2011), 'Characteristic facial retargeting', *Computer Animation and Virtual Worlds* **22**(2-3), 187–194.

Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C. & Seidel, H.-P. (2004), Laplacian surface editing, *in* 'Proc. of Symp. on Geometry Processing (SGP)', pp. 179–188.

Steevesa, J. K., Culhama, J. C., Duchaineb, B. C., Pratesia, C. C., Valyeara, K. F., Schindlerc, I., Humphreya, G. K., Milnerc, A. D. & Goodalea, M. A. (2006), 'The fusiform face area is not sufficient for face recognition: Evidence from a patient with dense prosopagnosia and no occipital face area', *Neuropsychologia* **44**(4), 594–609.

Steppacher, I., Schindler, S. & Kissler, J. (2015), 'Higher, faster, worse? An event-related potentials study of affective picture processing in migraine', *Cephalalgia* **36**(3), 249–257.

Sumner, R. W. & Popović, J. (2004), 'Deformation transfer for triangle meshes', *ACM Transactions on Graphics* **23**(3), 399–405.

Tam, G. K. L., Cheng, Z.-Q., Lai, Y.-K., Langbein, F. C., Liu, Y., Marshall, D., Martin, R. R., Sun, X.-F. & Rosin, P. L. (2013), 'Registration of 3D point clouds and meshes: A survey from rigid to non-rigid', *IEEE Transactions on Visualization and Computer Graphics* **19**(7), 1199–1217.

Taylor, J., Shehzad, Z. & McCarthy, G. (2016), 'Electrophysiological correlates of face-evoked person knowledge', *Biological Psychology* **118**, 136–146.

Tevs, A., Bokeloh, M., Wand, M., Schilling, A. & Seidel, H.-P. (2009), Isometric registration of ambiguous and partial data, *in* 'Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)', pp. 1185–1192.

Thies, J., Zollhöfer, M., Nießner, M., Valgaerts, L., Stamminger, M. & Theobalt, C. (2015), 'Real-time expression transfer for facial reenactment', *ACM Transactions on Graphics* **34**(6), 183:1–183:14.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. & Nießner, M. (2016), Face2Face: Real-time face capture and reenactment of RGB videos, *in* 'IEEE Conf. on Computer Vision and Pattern Recognition', pp. 2387–2395.

Thom, N., Knight, J., Dishman, R., Sabatinelli, D., Johnson, D. C. & Clementz, B. (2013), 'Emotional scenes elicit more pronounced self-reported emotional experience and greater EPN and LPP modulation when compared to emotional faces', *Cognitive, Affective and Behavioral Neuroscience* **14**(2), 849–860.

Tinwell, A. & Grimshaw, M. (2009), Bridging the uncanny: An impossible traverse?, *in* 'Proc. of Int. MindTrek Conf.: Everyday Life in the Ubiquitous Era'.

Tinwell, A., Grimshaw, M., Nabi, D. A. & Williams, A. (2011), 'Facial expression of emotion and perception of the Uncanny Valley in virtual characters', *Computers in Human Behavior* **27**(2), 741–749.

Tinwell, A., Grimshaw, M. & Williams, A. (2011), 'The Uncanny Wall', *Int. Journal of Arts and Technology* **4**(3).

Todorov, A., Said, C. P., Engell, A. D. & Oosterhof, N. N. (2008), 'Understanding evaluation of faces on social dimensions', *Trends in Cognitive Sciences* **12**(12), 455–460.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O. & Kanwisher, N. (2000), 'Response properties of the human fusiform face area', *Cognitive Neuropsychology* **17**(13), 257–280.

Torrance, K. E. & Sparrow, E. M. (1967), 'Theory for off-specular reflection from roughened surfaces', *Journal Optical Society America* **57**, 1105–1114.

Tsao, D. Y. & Livingstone, M. S. (2008), 'Mechanisms of face perception', *Annual Review of Neuroscience* **31**, 411–437.

Umeyama, S. (1991), 'Least-squares estimation of transformation parameters between two point patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(4), 376–380.

Valgaerts, L., Wu, C., Bruhn, A., Seidel, H.-P. & Theobalt, C. (2012), 'Lightweight binocular facial performance capture under uncontrolled lighting', *ACM Transactions on Graphics* **31**(6), 187:1–187:11.

van Kaick, O., Zhang, H., Hamarneh, G. & Cohen-Or, D. (2011), 'A survey on shape correspondence', *Computer Graphics Forum* **30**(6), 1681–1707.

Vanderah, T. & Gould, D. J. (2016), *Nolte's The Human Brain: An Introduction to its Functional Anatomy*, Elsevier.

Vangorp, P. (2009), *Human Visual Perception of Materials in Realistic Computer Graphics*, PhD Thesis, University Leuven.

Vangorp, P., Laurijssen, J. & Dutre, P. (2007), 'The influence of shape on the perception of material reflectance', *ACM Transaction on Graphics* **26**(3), 77:177:9.

von der Pahlen, J., Jimenez, J., Danvoye, E., Debevec, P., Fyffe, G. & Alexander, O. (2014), Digital Ira and beyond: Creating real-time photoreal digital actors, *in* 'ACM SIGGRAPH Courses', pp. 1:1–1:384.

Wallraven, C., Breidt, M., Cunningham, D. W. & Bülthoff, H. H. (2008), 'Evaluating the perceptual realism of animated facial expressions', *ACM Transactions on Applied Perception* **4**(4), 4:1–4:20.

Wallraven, C., Bülthoff, H. H., Cunningham, D. W., Fischer, J. & Bartz, D. (2007), 'Evaluation of real-world and computer-generated stylized facial expressions', *ACM Transactions on Applied Perception* **4**(3).

Walter, B., Marschner, S. R., Li, H. & Torrance, K. E. (2007), Microfacet models for refraction through rough surface, *in* 'Proc. of the Eurographics Symp. on Rendering (EGSR)', pp. 195–206.

Wang, C. & Mahadevan, S. (2009), Manifold alignment without correspondence, *in* 'Proc. Int. Joint Conf. on Artifical Intelligence', pp. 1273–1278.

Wang, C. & Mahadevan, S. (2011), Heterogeneous domain adaptation using manifold alignment, *in* 'Proc. Int. Joint Conf. on Artificial Intelligence', pp. 1541–1546.

Wang, C. & Mahadevan, S. (2013), Manifold alignment preserving global geometry, *in* 'Proc. Int. Joint Conf. on Artificial Intelligence', pp. 1743–1749.

Wang, Y., Gupta, M., Zhang, S., Wang, S., Gu, X., Samaras, D. & Huang, P. (2008), 'High resolution tracking of non-rigid motion of densely sampled 3D data using harmonic maps', *Int. Journal of Computer Vision* **76**(3), 283–300.

Wang, Y., Huang, X., Lee, C.-S., Zhang, S., Li, Z., Samaras, D., Metaxas, D., Elgammal, A. & Huang, P. (2004), 'High resolution acquisition, learning and transfer of dynamic 3D facial expressions', *Computer Graphics Forum* **23**(3), 677–686.

Ward, K., Bertails, F., Kim, T.-Y., Marschner, S. R., Cani, M.-P. & Lin, M. C. (2007), 'A survey on hair modeling: Styling, simulation, and rendering', *IEEE Transactions on Visualization and Computer Graphics* **13**(2), 213–234.

Wei, Y., Ofek, E., Quan, L. & Shum, H.-Y. (2005), 'Modeling hair from multiple views', *ACM Transactions on Graphics* **24**(3), 816–820.

Weise, T., Bouaziz, S., Li, H. & Pauly, M. (2011), 'Realtime performance-based facial animation', *ACM Transactions on Graphics* **30**(4), 77:1–77:10.

Weise, T., Li, H., Van Gool, L. & Pauly, M. (2009), Face/Off: Live facial puppetry, *in* 'Proc. of Symp. on Computer Animation (SCA)', pp. 7–16.

Werheid, K., Schacht, A. & Sommer, W. (2007), 'Facial attractiveness modulates early and late event-related brain potentials', *Biological Psychology* **76**(1-2), 100–108.

Weyrich, T., Matusik, W., Pfister, H., Bickel, B., Donner, C., Tu, C., McAndless, J., Lee, J., Ngan, A., Jensen, H. W., & Gross, M. (2006), 'Analysis of human faces using a measurement-based skin reflectance model', *ACM Transactions on Graphics* **25**(3), 1013–1024.

Wheatley, T., Weinberg, A., Looser, C., Moran, T. & Hajcak, G. (2011), 'Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP', *PLOS ONE* **6**(3), e17960.

Wieser, M. J. & Keil, A. (2013), 'Fearful faces heighten the cortical representation of contextual threat', *NeuroImage* **81**, 317–325.

Wilkerson, C. L., Syed, N. A., Fisher, M. R., Robinson, N. L., Wallow, H. L. & Albert, D. M. (1996), 'Melanocytes and iris colorlight microscopic findings', *Archives of Ophthalmology* **114**(4), 437–442.

Wisessing, P., Dingliana, J. & McDonnell, R. (2016), Perception of lighting and shading for animated virtual characters, *in* 'Proc. of ACM Symp. of Applied Perception (SAP)', pp. 25–29.

Woodman, G. F. (2010), 'A brief introduction to the use of event-related potentials (ERPs) in studies of perception and attention', *Attention, Perception & Psychophysics* **72**(8), 2031–2046.

Wu, C., Bradley, D., Garrido, P., Zollhöfer, M., Theobalt, C., Gross, M. & Beeler, T. (2016), 'Model-based teeth reconstruction', *ACM Transactions on Graphics* **35**(6), 220:1–220:13.

Xu, F., Chai, J., Liu, Y. & Tong, X. (2014), 'Controllable high-fidelity facial performance transfer', *ACM Transactions on Graphics* **33**(4), 42:1–42:11.

Xu, K., Ma, L.-Q., Ren, B., Wang, R. & Hu, S.-M. (2011), 'Interactive hair rendering and appearance editing under environment lighting', *ACM Transactions on Graphics* **30**(6), 173:1–173:10.

Yuksel, C., Schaefer, S. & Keyser, J. (2009), 'Hair meshes', *ACM Transactions on Graphics* **28**(6), 166:1–166:7.

Zhang, Y., Dong, W., Deussen, O., Huang, F., Li, K. & Hu, B.-G. (2014), Data-driven face cartoon stylization, *in* 'SIGGRAPH Asia Technical Briefs', pp. 14:1–14:4.

Zhu, J.-Y., Agarwala, A., Efros, A. A., Shechtman, E. & Wang, J. (2014), 'Mirror, mirror: Crowdsourcing better portraits', *ACM Transactions on Graphics* **33**(6), 234:1–234:12.

Zibrek, K. & McDonnell, R. (2014), Does render style affect perception of personality in virtual humans?, *in* 'Proc. of ACM Symp. on Applied Perception (SAP)', pp. 111–115.

Zinke, A. & Weber, A. (2007), 'Light scattering from filaments', *IEEE Transactions on Visualization and Computer Graphics* **13**(2), 342–356.

Zinke, A., Yuksel, C., Weber, A. & Keyser, J. (2008), 'Dual scattering approximation for fast multiple scattering in hair', *ACM Transactions on Graphics* **27**(3), 1–10.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320.