A Morphing-Based Analysis of the Perceptual Distance Metric of Human Faces

Nadine Gummersbach*
University of Siegen

Volker Blanz†
University of Siegen

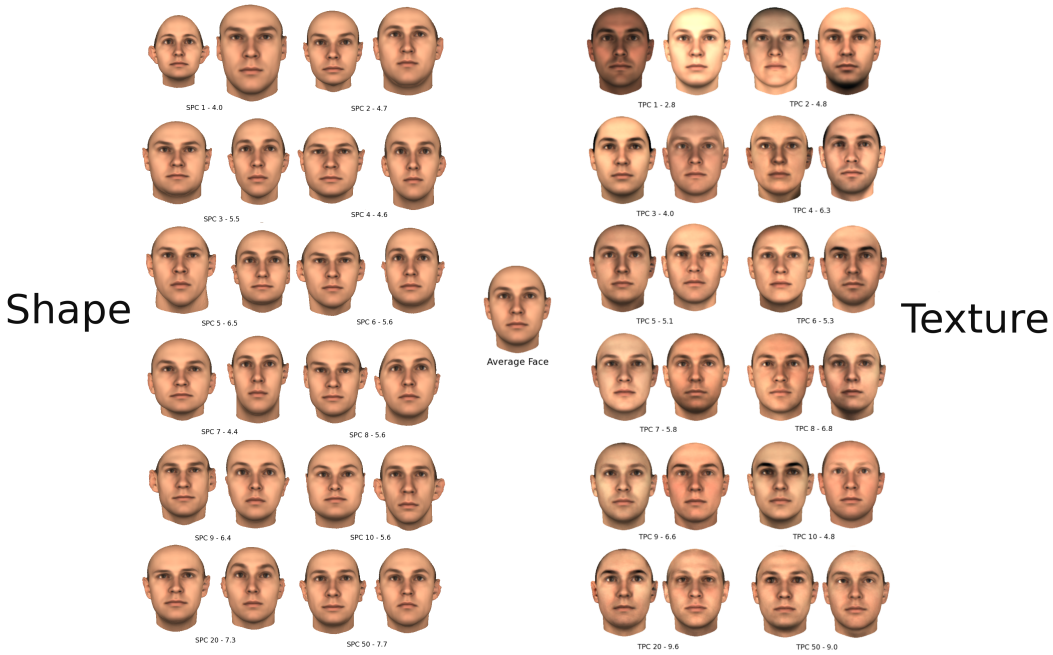

Figure 1: *Visualization of the principal components (PCs) for shape (left) and texture (right): Each pair of faces shows the effect of adding and subtracting a multiple λ of a PC (Equations (1) and (2)). The numbers below each pair indicate the number i of the PC (1 - 50) and λ.*

Abstract

This paper presents a psychophysical analysis of the discrimination thresholds of human faces that are varied along different directions in Face Space. Generated by a 3D Morphable Model, the stimuli are frontal views of original laser scans that are modified in shape, texture or both. Discrimination thresholds are then measured in a four Alternative Forced Choice (4AFC) design and a staircase method.

In Experiment 1, texture and shape are varied separately along a set of principal component directions. For different components, the results show a consistent pattern of high and low thresholds across individual participants. We compare Mahalanobis distance, Euclidean distance in face space, and 2D image differences as potential predictors for human discrimination thresholds.

The goal of Experiment 2 is to investigate the interaction between shape and texture. The stimuli include combined and separate variations of shape and texture in the 4AFC staircase setup, which are analyzed in a within-subject paradigm. The results indicate that participants rely on both shape and texture for their decision. The experiments help to assess different potential models of the fusion of shape and texture cues, and provide useful information for computer graphics applications such as character design interfaces.

CR Categories: H.1.2 [Information Systems]: MODELS AND PRINCIPLES—User/Machine Systems: Human information processing; I.4.8 [Computing Methodologies]: IMAGE PROCESSING AND COMPUTER VISION—Scene Analysis: Object recognition, Shape; J.4 [Computer Applications]: SOCIAL AND BEHAVIORAL SCIENCES—Psychology

Keywords: Faces, Face Space, Distance Measures, Metric, Morphing, Morphable Model, Shape Perception

1 Introduction

For several decades, researchers in computer graphics, computer vision, psychology and neuroscience have investigated how human faces can be represented in a computer or in the human visual system (HVS), and how this information can be processed. An issue that is closely related to the internal representation and the processing mechanisms is the similarity measure or distance metric between faces. In computer vision, it defines whether two faces are found to be different individuals or whether they show the same person in different imaging conditions. In graphics, the distance measure is important for character design interfaces, and it may serve as an objective quality criterion for automated or manual 3D repro-

*e-mail: nadineg@informatik.uni-siegen.de
†e-mail: blanz@informatik.uni-siegen.de

ductions of real actors' heads in special effects. In psychology and neuroscience, the similarity measure between faces may provide useful insights into the neural representations and mechanisms in the HVS. The goal of this paper is to use computer graphics stimuli in psychophysical experiments to investigate the similarity measure applied by human participants, and to compare these findings to different potential candidates of computational metrics.

The most straight-forward criterion for the similarity of faces would be pixel-by-pixel image difference. However, this measure would not distinguish between changes that are due to imaging conditions, such as lighting or pose, and intrinsic differences in facial shape or texture.

A more sophisticated metric would be based on a Face Space representation, for example in terms of a 3D Morphable Model [Blanz and Vetter 1999], which compares shapes and textures of corresponding surface points of faces, such as the tips of the noses. The simple Euclidean norm between shape and texture vectors would be a robust distance measure that separates imaging parameters from intrinsic facial characteristics.

Based on such a vector space representation of faces, a criterion that is based on Principal Component Analysis (PCA) would take into account the statistical distribution of faces. A PCA of a set of samples (vectors) defines an orthogonal basis of eigenvectors which are usually sorted in descending order according to the variance observed along these directions, so the first principal components describe vector space directions along which the data have the largest variance. Sirovich and Kirby [Sirovich and Kirby 1987] showed that PCA is an efficient representation of high-dimensional pictures of faces in a lower dimensional subspace. Based on this method, Turk and Pentland developed a system for face recognition [Turk and Pentland 1991]. Since then PCA has been used in many face recognition algorithms. Additionally, numerous psychophysical studies investigated the psychological plausibility of PCA (e. g. [Hancock et al. 1996]).

In PCA, it is assumed that the sample vectors of the training set are drawn from a probability density function that takes the form of a multidimensional normal distribution, so PCA essentially fits a normal distribution to the data. Vectors that have equal probability are located on ellipsoids centered around the arithmetic mean, and the geometric shape of these ellipsoids is defined by the eigenvectors and the variances. The normal distribution defines a metric known as Mahalanobis distance [Duda et al. 2001], which compensates for the different variances along different axes. All vectors on the equal-probability ellipsoids have equal Mahalanobis distance from the mean. We investigate Mahalanobis distance in this paper as a similarity measure between faces.

In smooth surface data, where adjacent surface points are more correlated than distant points, these first principal components tend to capture large scale variations, such as overall size or overall brightness changes in shape or texture data, while the higher order principal components with small variances describe variations at a finer scale (i.e. higher spatial frequency). It is interesting to find out how the different principal components contribute to the distance measure applied by human observers. O'Toole et al. found that faces are identified more reliably if principal components with smaller variances are used for reconstruction [O'Toole et al. 1993]. In contrast, for the decision whether a face is male or female, global information is sufficient: O'Toole et al. [O'Toole et al. 1993] discovered that features related to the sex of a face are coded in the first principal components (i.e. high variances). Similar results were found for classification of age or race [Buchala et al. 2005].

Unlike these identification or classification experiments, we measure the perceived distance in a discrimination task and compare the thresholds to Mahalanobis distance, Euclidean norm and image difference. In particular, we are interested in the question whether there are relationships between the empirical measurement of perception thresholds and the statistically described Face Space, determined by a PCA. The discrimination task measures facial similarity on a rather small scale, because participants are very sensitive to small changes in faces. We reduced the sensitivity by using presentation times that are too short to scrutinize the images too thoroughly.

Perceived similarity on a larger scale, i.e. for more distant pairs of faces, can be obtained by direct similarity ratings by participants. Such ratings were investigated in a number of experiments and compared to a PCA based measure in Face Space [Scheuchenpflug 1998], [Tredoux 2002] [Hancock et al. 1997]. Results of these studies are basically that similarity of persons is determined to a certain extent by their distance in Face Space, but experimental data of different participants vary considerably, so that it is difficult to obtain reliable results. In our own pilot studies with ratings on substantially different stimuli (original scans of individual faces), participants reported that they found it very hard to judge and rate similarities. This may be due to the fact that their response may be based on a variety of criteria, such as overall shape, facial details, shape or texture, and that humans can consciously choose their priorities, which makes the response more arbitrary than unconscious responses would be.

The second experiment in our paper investigates how shape and texture contribute to the discriminability. This is done by measuring thresholds for stimuli in which both shape and texture are changed at the same time, and stimuli with modification of shape only or texture only. We compare our results to different hypotheses how shape and texture cues could be integrated in the HVS. Hancock et al. [Hancock et al. 1996] asked how easy it would be to recognize a certain person at a station to analyze which faces are easier to remember than others. In this work, they used stimuli with the original shape as well as so called shape-free-faces. The shape-free faces keep the original texture, but their shape is replaced by the average. One result is that variations of shape do not correlate with the ability to recognize persons (more precisely: the false positives). Only the texture seems to be important. The authors conclude that shape and texture information are processed separately by the HVS.

O'Toole et al. [O'Toole et al. 1999] analyzed to what extent the 3D-shape-information and the 2D-texture-information contribute to the recognition of faces. The main question of that work was how these different contributions change under varying viewpoints. Results show that both components are equally important for good performance in face recognition tasks. Interestingly, there are differences between male and female faces. While the shape and texture information are equally useful for female faces, there are great differences for male faces. Males with average textures are recognized much better than females. Hill et al. [Hill et al. 1995] found that the combination of shape and texture is also important for the classification of sex and race.

2 Stimulus creation

The stimuli were created using a database of 200 laser scans of faces represented in a 3D Morphable Face Model [Blanz and Vetter 1999]. Each face consists of a shape vector S_{orig} that contains x, y and z coordinates of 75000 vertices on the facial surface, and a texture vector T_{orig} that contains the red, green and blue color of each vertex.

For every trial, one of the original faces is picked randomly and displayed along with a modified version. This modification is done

by adding shape or texture changes along principal component directions. For shape, let $\mathbf{s_i}$ be the principal component eigenvector number i with a standard deviation $\sigma_{s,i}$ (where $\sigma_{s,i} \geq \sigma_{s,k}$ if $i < k$). For texture, let the eigenvectors be $\mathbf{t_i}$ with standard deviations $\sigma_{t,i}$. Modifications in shape are then achieved by

$$S_{modif} = S_{orig} + \lambda_{s,i}\, \sigma_{s,i}\, \mathbf{s_i} \tag{1}$$

and modifications in texture by

$$T_{modif} = T_{orig} + \lambda_{t,i}\, \sigma_{t,i}\, \mathbf{t_i}. \tag{2}$$

In this notation, if we add multiples of the unit length eigenvectors to a shape vector S_{orig},

$$S_{modif} = S_{orig} + \sum_i \lambda_{s,i}\, \sigma_{s,i}\, \mathbf{s_i}$$

the Mahalanobis distance (i.e. the variance-corrected distance) is

$$\|S_{modif} - S_{orig}\|^2_{Maha} = \sum_i \lambda^2_{s,i}$$

For texture, we obtain a similar result. In our stimuli, we modify only one principal component at a time, so

$$\|S_{modif} - S_{orig}\|_{Maha} = \lambda_{s,i}, \quad \|T_{modif} - T_{orig}\|_{Maha} = \lambda_{t,i}.$$

In the experiments, we apply modifications to shape, texture or both, and vary the principal component number i and the scaling factors λ.

The 3D faces are rendered in a front view pose at a frontal illumination with additional ambient light. The skin reflection is mostly diffuse, with only a mild specular component that does not produce distinctive specular highlights. To avoid aliasing artefacts along the silhouette, which would be giveaway cues of shape changes, we apply anti-aliasing by downsampling the images from a higher initial resolution.

The stimuli are presented on a standard 20 inch color monitor. Each face is about 5.5 cm wide and 7.5 cm high, and the distance from the observer is about 60 cm. The rendering parameters are kept constant throughout the experiment.

3 Experiment 1

The goal of this experiment is to measure thresholds for detecting changes in human faces along different principal components in shape and texture, and to find out how they relate to the standard deviations along these directions.

3.1 Procedure

In each trial, four faces are shown simultaneously on the screen at fixed positions (two rows and two columns). Three faces are identical, and the fourth is modified as described in Section 2. The identity of the original face and the position of the slightly modified face are decided randomly. After a presentation time of three seconds, grey rectangles replace the four faces to indicate the original positions. In a four Alternative Forced Choice (4AFC) task, participants select by mouse click which of the four positions showed the face that was different from the others. After the click, the next four faces are shown (Figure 1).

Presentation time is limited in order to prevent participants from thorougly scrutinizing the images. Instead, we are interested in discriminability on a greater scale based on more salient differences

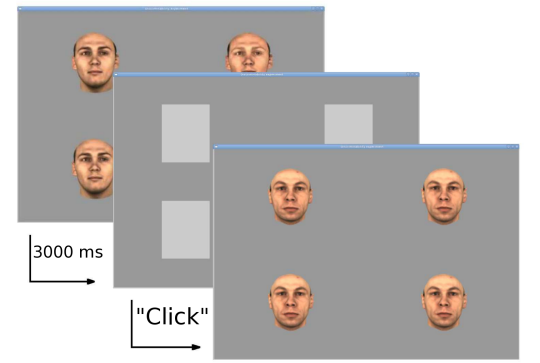

Figure 2: *Timeline of the experiment - After 3000 ms the four faces are replaced by grey rectancles. The next trial starts after participants click on one of the rectangles.*

and an overall impression. Due to the simultaneous presentation, participants can look at the faces in any order.

In 24 experimental conditions, we investigate separate variations along the following principal component directions:

$$\mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3}, \mathbf{s_4}, \mathbf{s_5}, \mathbf{s_6}, \mathbf{s_7}, \mathbf{s_8}, \mathbf{s_9}, \mathbf{s_{10}}, \mathbf{s_{20}}, \mathbf{s_{50}}, \tag{3}$$

$$\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3}, \mathbf{t_4}, \mathbf{t_5}, \mathbf{t_6}, \mathbf{t_7}, \mathbf{t_8}, \mathbf{t_9}, \mathbf{t_{10}}, \mathbf{t_{20}}, \mathbf{t_{50}}. \tag{4}$$

For each condition, we run a staircase method [Gescheider 1997] to determine the thresholds λ (Equation 1 and 2). In each trial, one of the 24 conditions is selected randomly, and the next step of the staircase method is performed for this condition. Due to the random order, participants cannot adjust their strategy of inspection on shape or texture changes or on specific types of modifications. They would be able to do that if we ran the staircase experiments for each condition separately.

Based on the response of the participant, the staircase method determines the next stimulus value such that the threshold value is measured precisely in a minimum number of trials. After a pilot study, we chose a starting value of $\lambda_0 = 2.0$ for all conditions. This is approximately the mean of all thresholds of all principal components. Starting with that value, λ is decreased if the participant selects the correct face, and increased otherwise. For each condition, we perform 20 trials. The threshold is then computed as the arithmetic mean of the turning points of λ.

The values for λ are increased and decreased on a logarithmic scale. New values are determined by the function $\lambda = \lambda_0 * \alpha^l$, with $\alpha = 1.5$. The exponent l defines the current level of the staircase and is incremented or decremented in steps of 1. The stimuli for shape-conditions are

$$S_{modif} = S_{orig} + \lambda_{s,i}\sigma_{s,i}\mathbf{s_i} = S_{orig} + \lambda_0 \alpha^{l_{s,i}} \sigma_{s,i}\mathbf{s_i} \qquad l_{s,i} \in \mathcal{Z} \tag{5}$$

and for texture

$$T_{modif} = T_{orig} + \lambda_{t,i}\sigma_{t,i}\mathbf{t_i} = T_{orig} + \lambda_0 \alpha^{l_{t,i}} \sigma_{t,i}\mathbf{t_i} \qquad l_{t,i} \in \mathcal{Z}. \tag{6}$$

Initially, we tested only modifications along the positive directions $\mathbf{s_i}$ or $\mathbf{t_i}$. As we verify later in Experiment 1b, the results for negative directions (i.e. $-\mathbf{s_i}$ or $-\mathbf{t_i}$) are the same.

The experiment consists of 480 trials per participant (12 principal components for shape and texture each, 20 staircase trials per condition). The experiment takes about 40 min per person. Participants were allowed to take a break once every 50 trials.

3.2 Participants

Our participants were 21 students and members of staff of our university (7 females and 14 males). They were not paid or compensated otherwise. Two participants already took part in a pilot study, the others had no experience with psychophysical experiments.

3.3 Results

In this section, we analyze the thresholds, which are obtained by averaging the values where the staircase of each individual condition changes from increment to decrement or vice versa.

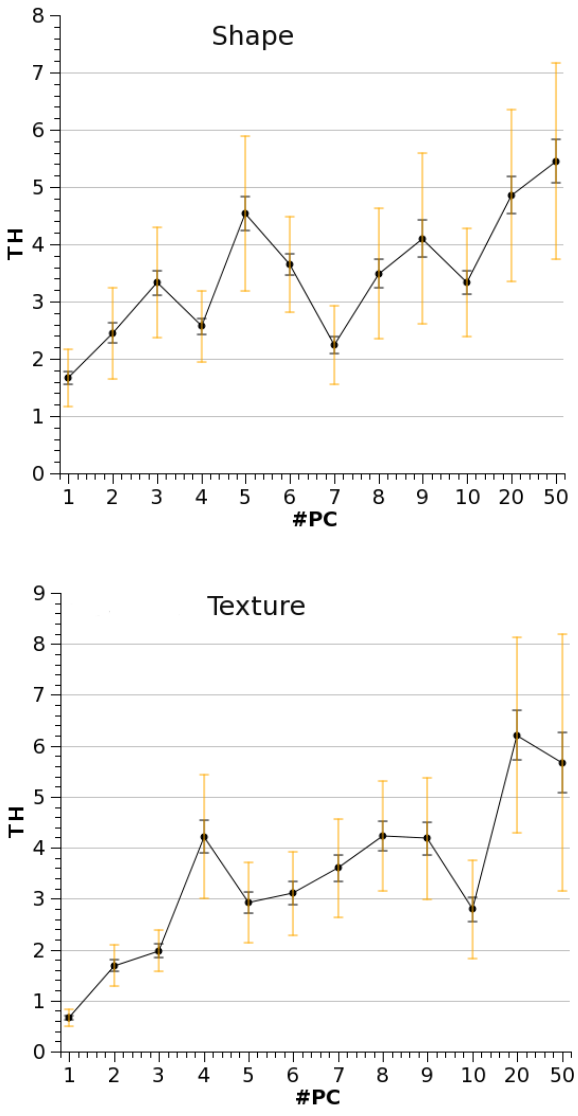

Figure 3: *Mean thresholds (TH = λ) of the shape and texture components (PC = i). The black error bars indicate standard errors, the orange bars indicate standard deviations. Both for shape and for texture, there is a slight positive trend.*

Figure 3 shows the average thresholds pooled over all participants. The diagrams show the thresholds $\lambda_{s,i}$ and $\lambda_{t,i}$ for principal components i (horizontal axis), measured in units of standard deviations. If Mahalanobis distance would explain human responses perfectly, we would expect a constant line in the diagrams.

Instead, there are significant differences in thresholds for different i, and these are consistent across individual participants: For example, the fifth principal in shape and the fourth principal component in texture seem to be difficult to detect. We would have expected a more smooth function, because the directions of the principal components (unlike the standard deviations) do not need to be relevant perceptually: in the extreme case where two directions have the same standard deviations (degenerate eigenspace), the directions of the two principal components that span this space are not even mathematically well defined.

For shape, Mahalanobis distance seems to predict the general trend of human thresholds over the entire range $i = 1$ through 50 quite well, despite the slight increase in Figure 3. By linear regression, we found a slope of 0.051. In contrast, Figure 4 shows the thresholds measured in terms of the Euclidean distance (L_2 norm) $\lambda_{s,i} \cdot \sigma_{s,i}$ (Note that all eigenvectors $\mathbf{s_i}$ have unit length.) In the morphable model, Euclidean distance is the square root of the sum of all squared vertex displacements in 3-dimensional space. For shape, this curve falls off substantially (slope: $-3.3 \cdot 10^4$). Our results indicate that human shape thresholds are well adapted to the statistics of faces.

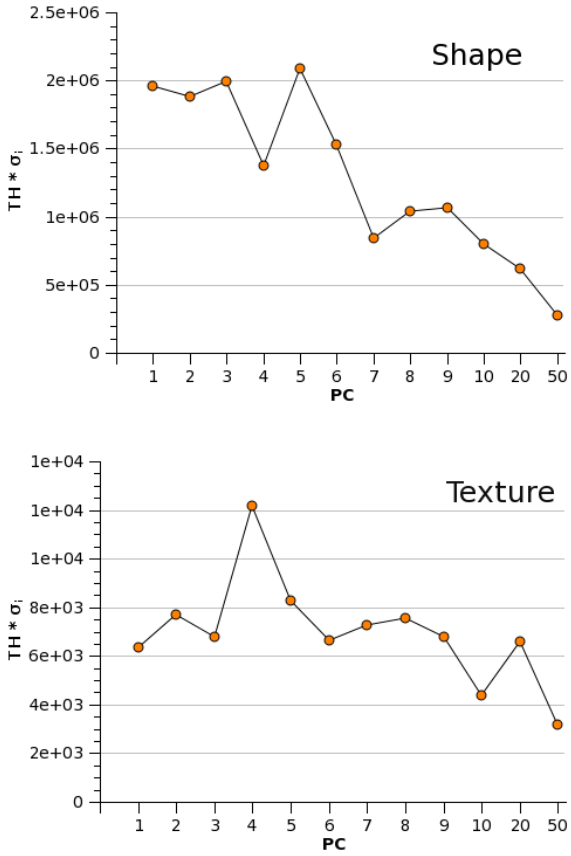

Figure 4: *Mean thresholds along shape and texture PCs in terms of Euclidean distance ($\lambda \cdot \sigma_i$). There is an overall negative trend for shape, but less so for texture.*

For texture, Figures 3 and 4 show that Mahalanobis distance does not explain the threshold measurements better than Euclidean dis-

tance: Figure 3 shows a general increase in thresholds (slope: 0.056), so participants were more sensitive to the first principal components (large scale changes) than to the fine scale changes that occur with higher order principal components. In terms of Euclidean distance, however, the thresholds are relatively constant (slope: -97.5 on a significantly larger scale of distance values), with some outliers. This finding is surprising to us not only because we would have expected all thresholds to be adapted to the statistics of faces. The first principal components describe coarse, low spatial frequency changes in faces (Figure 1), some of which could also be due to imaging and lighting effects. Therefore, we would have expected that humans neglect these factors and focus more on fine, invariant details. It may be due to the simultaneous presentation in our experimental design that participants are highly sensitive to these principal components.

3.4 Experiment 1b – Relevance of the sign of λ

In the previous experiment, we tested only positive stimulus values λ, based on the assumption that modifications of faces along a certain axis in Face Space affect the same features of a face, so the perceived difference should be independent of the direction. In fact, if the modified faces look as natural as the original faces, a change of sign is equivalent to changing the roles of original and modified faces.

To verify if this assumption is true, we repeated the experiment in a slightly different version. We tested the principal components 1, 3 and 20 for shape and texture, but with eigenvectors $\mathbf{s_i}$ and $-\mathbf{s_i}$, $\mathbf{t_i}$ and $-\mathbf{t_i}$. The results indicate clearly that it is not important for the value of a threshold in which direction we modify a face: The differences between both directions lie within one standard deviation of the thresholds found in Experiment 1 (The full set of measurements is available on our website *http://mi.informatik.uni-siegen.de*).

4 Experiment 2

After testing separate modifications of shape and texture in the first experiment, we investigated in the second experiment how the thresholds change when both features are modified at the same time. We considered three hypotheses:

1. Texture distance $f(T)$ and shape distance $f(S)$ are added to a single value of perceived distance.
$$f(S) + f(T) > t \tag{7}$$
2. Textures and shapes are processed separately, and faces are perceived as different if one of the distance measures is above its threshold:
$$f(S) > t_s \cup f(T) > t_t \tag{8}$$
3. Both components are processed together in a joint distance function.
$$f(S,T) > t \tag{9}$$

4.1 Procedure

In a within-subject design, we measure for each subject the separate thresholds for shape and texture, as in Experiment 1, and in combination. This allows us to compare the thresholds in the combined conditions with those of the separate conditions without any between-subject variation.

To reduce the number of conditions, the combinations are tested only with a reduced number of principal components ($i = 1, 2, 3, 7, 20$) and only in combinations of $\mathbf{s_i}$ and $\mathbf{t_i}$ for the same number i. The choice of these combinations is arbitrary, because there is no fundamental reason why the first principal component of shape should interact with the first principal component of texture more than with the second. Mathematically, there is no connection between $\mathbf{s_i}$ and $\mathbf{t_i}$, because both PCAs are computed separately. We still found these combinations reasonable to make sure that the face modifications are on similar spatial frequency domains (both coarse for $\mathbf{s_1}, \mathbf{t_1}$, and both fine for $\mathbf{s_{20}}, \mathbf{t_{20}}$). However, it would still be interesting to investigate more combinations, such as $\mathbf{s_1}$ with $\mathbf{t_{20}}$.

Even though Experiment 1b demonstrated that the directions of modifications do not matter, signs do matter in the combined condition: For a fixed positive sign $+\mathbf{s_1}$, an additional change $\mathbf{t_1}$ is different from $-\mathbf{t_1}$: depending on the relative sign of shape and texture, their effects can either enhance each other or cancel each other out in some respects (brightness changes due to shading and texture changes in a given pixel). Therefore, we test both signs of texture changes for a fixed sign of shape change. In addition, we verify the findings of Experiment 1b by adding a condition with $-\mathbf{s_3}$. Therefore, we investigate the following combinations:

Shape	$\mathbf{s_1}$	$\mathbf{s_2}$	$\mathbf{s_3}$	$-\mathbf{s_3}$	$\mathbf{s_7}$	$\mathbf{s_{20}}$
Texture	$\pm\mathbf{t_1}$	$\pm\mathbf{t_2}$	$\pm\mathbf{t_3}$	$\pm\mathbf{t_3}$	$\pm\mathbf{t_7}$	$\pm\mathbf{t_{20}}$

The procedure is the same as in Experiment 1 (4AFC setup, randomly interleaved staircases with 20 trials per condition, presentation time is three seconds, followed by rectangles as replacements.) The staircases for separate shape and texture conditions and for the combined conditions are mixed, so for each trial, a random condition is selected and the staircase for this condition is continued by one step. In the combined conditions, the staircase levels $l_{i,\pm}$ for shape and texture are coupled (simultaneous increases and decreases), so

$$S_{modif} = S_{orig} + \lambda_{s,i} \cdot \alpha^{l_{i,\pm}}\, \sigma_{s,i}\, \mathbf{s_i} \tag{10}$$

$$T_{modif} = T_{orig} \pm \lambda_{t,i} \cdot \alpha^{l_{i,\pm}}\, \sigma_{t,i}\, \mathbf{t_i} \tag{11}$$

$$l_{i,\pm} \in Z, \alpha = 1.2 \tag{12}$$

We take the thresholds determined in Experiment 1 as starting values $\lambda_{s,i}$ and $\lambda_{t,i}$:

PC	1	2	3	7	20
λ_{s_0}	1.7	2.4	3.3	2.2	4.8
λ_{t_0}	0.7	1.7	2.0	3.6	6.2

In Equations (10), (11), these values serve not only as starting values, but also as constant scaling factors between shape and texture throughout the staircases (only the exponents $l_{i,\pm}$ are varied.) As a consequence, modifications in shape and texture should have equal perceived contributions to the final stimuli, and their ratios are the same for all trials and all participants.

4.2 Participants

In Experiment 2, 22 participants (5 females and 17 males) volunteered without payment or any other compensation. They were students and members of staff of our university. Six of them already participated in Experiment 1.

4.3 Results

The main focus of Experiment 2 is the interaction of shape and texture. After the scaling of the shape and texture components was already controlled, as described at the end of Section 4.1, we

rescale the thresholds again for the data analysis in order to eliminate between-subject variations: All thresholds for the combined conditions are scaled in a within-subject paradigm to the individual separate thresholds for each shape and texture component.

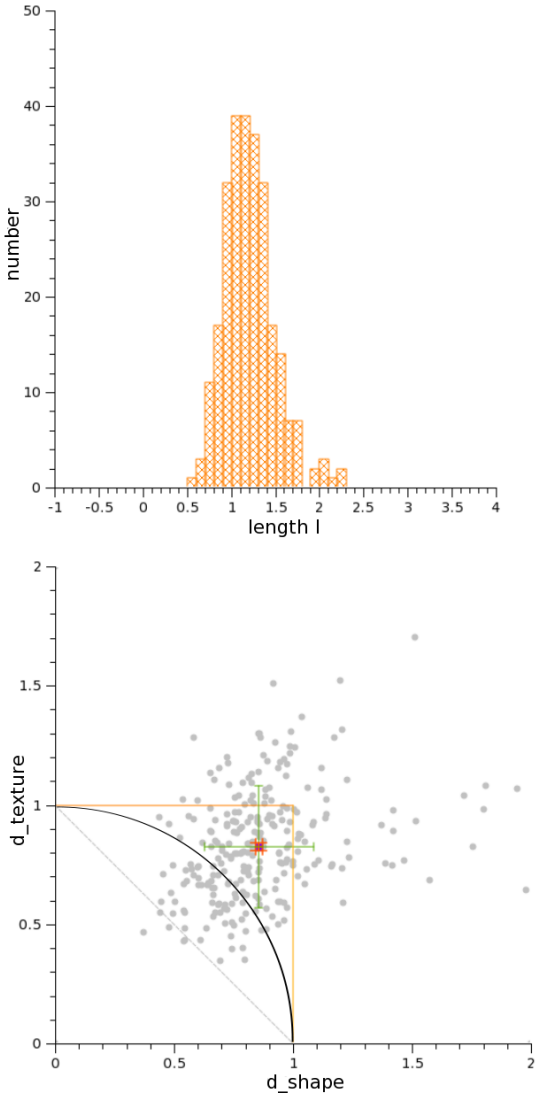

Figure 5: *Thresholds for combined modification of shape and texture. The top part shows the histogram for the variable length. The diagram on the bottom illustrates all measured, scaled thresholds. The highlighted point marks the mean of all observations, the green error bars indicate the standard deviations, the shorter orange bars the standard errors.*

All thresholds for all participants are shown in Figure 5 (bottom diagram). The position of each data point is determined by dividing the shape threshold of the combined condition $t_{s_{comb}}$ by the shape threshold with was measured for this dimension i separately $t_{s_{sep}}$ for this participant. This value is plotted as $d_{shape} = t_{s_{comb}}/t_{s_{sep}}$ on the horizontal axis. The value $d_{texture} = t_{t_{comb}}/t_{t_{sep}}$ for texture is calculated in the same way and is plotted on the vertical axis.

The top part of Figure 5 shows the histogram of the distances from the origin. This distance is measured by:

$$lenght = \sqrt{d^2_{shape} + d^2_{texture}}$$

Note that if we were to plot the thresholds for separate modifications of shape or texture on this scale, we would obtain $d_{shape} = 1$ and $d_{texture} = 1$ due to our normalization procedure (per participant and per component i.)

The mean distance from the original is 1.205 (σ = 0.286, σ_m = 0.0176). As we demonstrate below, this falsifies one of our hypotheses, while the decision between the two others remains unclear. The bottom part of Figure 5 additionally illustrates how the decision boundaries can be interpreted with respect to the hypotheses:

1. The dotted diagonal line between the points ($d_{shape} = 1, d_{texture} = 0$) and ($d_{shape} = 0, d_{texture} = 1$) represents the assumption that distance measures for texture and shape are added: $d_{combined}$ = d_{shape} + $d_{texture}$.

   Faces are perceived as different if $d_{combined} \geq 1$

   If both components contribute equally, we obtain $length = 0.7$.

2. The orange box would be the expected decision boundary for the decision rule

   Faces different if ($d_{shape} \geq 1$) **or** ($d_{texture} \geq 1$)

   The unit square shape is a result of scaling. If texture and shape thresholds contribute equally, the distance from the origin is $length \geq 1.4$. In principle it is possible that we concentrate only on one of the components, shape or texture. That would imply that only one condition is relevant in the equation above, while the other is ignored.

3. The circle illustrates the hypothesis that the shape vectors $\vec{s}$ and texture vectors $\vec{t}$ are concatenated to a single vector $\vec{x}$:

   $$\vec{s}, \vec{t} \rightarrow \vec{x} = \begin{pmatrix} \vec{s} \\ \vec{t} \end{pmatrix}, \qquad d_x = \sqrt{d^2_{shape} + d^2_{textur}} \qquad (13)$$

   Then, the distance of the vector of the original face to the vector of the modified face is compared to a threshold. Due to scaling, the radius of the circle is $length = 1.0$:

   Faces different if $d_x \geq 1.0$

From the observed thresholds in Figure 5, we can rule out the first hypothesis, which states that texture and shape information is simply added: Only 5% of the thresholds have a distance $length \leq 0.8$. Due to hypothesis 1, 50% should be below $\sqrt{0.5}$.

More problematic are the results concerning the second and third hypothesis, because the mean distance thresholds do not support one of them clearly. Our results would be explained by an L_p norm with $p > 2$. We can exclude the possibility that we only concentrate on one component and ignore the other, because only 12.5% of the points are above the shape-only or texture-only thresholds.

Even though our results cannot clearly support one of the two remaining models for cue integration, they give a quantitative empirical description that may be a starting point for further research and a useful basis for practical applications.

4.4 Analysis of image differences

In a post-hoc analysis, we computed the pixel-by-pixel image differences that occur when the observed thresholds for shape or texture are applied to the average face (Figure 7). We analyzed the conditions where shape and texture were modified separately.

For texture (left diagram), the fact that these values are almost constant indicates that simple image distance is quite consistent with the criterion applied by participants. In contrast, shape thresholds decrease as i increases, which demonstrates that a simple 2D image comparison mechanism does not explain the human responses: At equal image differences, participants would be more sensitive to detect the high-order principal components (low variance, high spatial frequency deformations) than the first principal components.

5 General discussion

This paper investigates the distance measure in human face perception in a discrimination task.

We have presented a new experimental setup that combines staircase methods with a 4AFC task. We exploit the fact that stimuli are computed on the fly during the experiment from a 3D Morphable Model. We believe that this paradigm may be extended to address a variety of other perceptual problems by varying other parameters in the model. For example, these can be attributes such as gender or body weight, which may be learned from labeled data [Blanz and Vetter 1999].

Our findings support the hypothesis that the discrimination thresholds applied by humans are, as far as shape modifications are concerned, adapted to the statistics of human faces. In other words, PCA can be used to predict the sensitivity of human observers for shape changes. For texture, however, we found our results to be more consistent with simpler distance measures such as Euclidean distance in texture space, or simple image difference. These two measures are closely related if the faces are displayed at the same geometrical parameters. If image conditions vary, we would expect that only Euclidean texture distance would explain our experimental data well.

We found that some principal components are consistently perceived with high sensitivity, while others are more difficult to see. This pattern generalizes across different participants and is unexpected because the directions of principal component directions may be determined by the specifics of the statistical distribution of training data, and it is very problematic to attach any perceptual meaning to those directions. Still, it seems that some of them contain facial attributes that we are very sensitive for. For example, the fourth principal component of texture captures some of the dark chin color due to stubbles in male faces. Note that all of the male individuals in the database were asked to shave before scanning, but still some had short stubbles. The shape thresholds are comparatively low for the fourth and seventh PC which both code the width of a face. It seems that the HVS is more sensitive to such specific attributes than to others.

We have also analyzed how shape and texture cues are integrated in the HVS. Our data show that participants rely on both cues, and that they do not simply add the distances of both to a combined criterion. However, our results did not clearly support either of two additional hypothesis about how the shape and texture information could be integrated. We assume that shape and texture information are integrated and analyzed in a joint distance metric, but this metric seems to be more complex than just the distance of concatenated vectors, which would be equivalent to a sum of squares of shape and texture differences.

Our results verify the assumption that the sign of the direction of modifications (adding or subtracting multiples of eigenvectors) does not affect the sensitivity. For combinations of shape and texture, we would have expected differences at least for some principal components. Especially for the first principal component, which represents information about the gender of a face, differences are likely. The first shape component varies in size and the first texture component represents skin complexion. A huge, dark face stands for male and small, bright faces for females. The combination can therefore be in the same or opposite sense according to gender, so that direction might matter. The analysis of our data showed that direction is not relevant for the combined conditions (Figure 6).

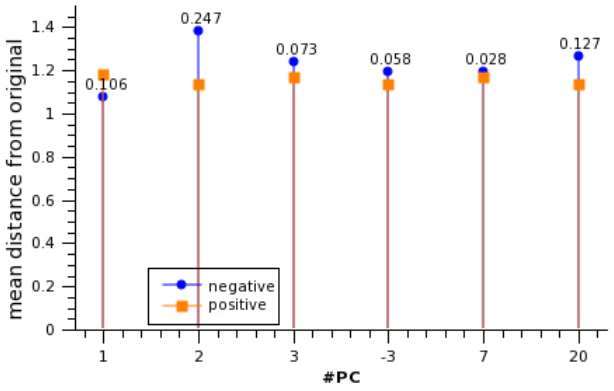

Figure 6: *Comparison of the distances in positive and negative direction of the texture.*

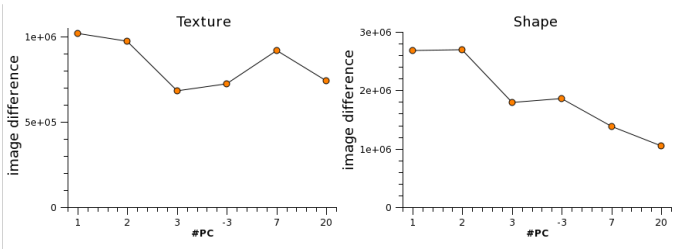

Figure 7: *Image difference when the average face is modified by a magnitude that corresponds to the thresholds from Experiment 2 along different PCs. The diagram shows mean square distances between the r,g,b- values of the original image and the modified image.*

In computer graphics, our results can be used in interactive tools for character design or for crowd generation to create a set of distinctive faces. In many fields in computer graphics, we are facing the problem to assess the quality of virtual reproductions of real faces, for example in face reconstruction from images, or in manual reproductions that are used in special effects. A good criterion for perceptual similarity may help to optimize the results of such projects. Getting a better understanding of the distance metric of human faces contributes both to basic research in psychology and neuroscience, and to practical applications in graphics and animation.

References

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Computer Graphics Proc. SIGGRAPH'99*.

BUCHALA, S., DAVEY, N., GALE, T., AND FRANK, R. 2005. Principal component analysis of gender, ethnicity, age and identity of face images. In *In IEEE ICMI International Conference on Multimodal Interfaces*.

DUDA, R., HART, P., AND STORK, D. 2001. *Pattern Classification*, 2nd ed. John Wiley & Sons, New York.

GESCHEIDER, G. 1997. *Psychophysics*, 3rd ed. Lawrence Erlbaum Associates, London.

HANCOCK, P. J. B., BURTON, M., AND BRUCE, V. 1996. Face processing: human perception and principal components analysis. *Memory and Cognition 26*, 26–40.

HANCOCK, P. J. B., BRUCE, V., AND BURTON, A. M. 1997. Testing principal component representations for faces. In *Proc. of 4th Neural Computation and Psychology Workshop*.

HILL, H., BRUCE, N., AND AKAMATSU, S. 1995. Perceiving the sex and race of faces: the role of shape and color. *Proceedings of the Royal Society 261*, 1362, 367–373.

O'TOOLE, A., ABDI, H., DEFFENBACHER, K., AND D.VALENTINE. 1993. Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America 10*, 3, 405–411.

O'TOOLE, A., DEFFENBACHER, K., AND D.VALENTINE. 1994. Structural aspects of face recognition and the other-race. *Memory and Cognition 22*, 208–224.

O'TOOLE, A., VETTER, T., AND BLANZ, V. 1999. Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vision Research 39*, 3145–3155.

SCHEUCHENPFLUG, R. 1998. Predicting face similarity judgements with a computational model of face space. *Acta psychologica 100*, 3, 229–242.

SIROVICH, L., AND KIRBY, M. 1987. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America 4*, 3, 519–524.

TREDOUX, C. 2002. A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology 8*, 3, 180–193.

TURK, M., AND PENTLAND, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3*, 1, 71–86.