Delineating the set of faces perceived as natural in the Basel Face Model

Veronica Bossio Botero (vb2516@cumc.columbia.edu)
Columbia University, 3227 Broadway
New York, NY 10027 United States

Wenxuan Guo (wg2361@columbia.edu)
Columbia University, 3227 Broadway
New York, NY 10027 United States

Jasper JF van den Bosch (J.F.VanDenBosch@leeds.ac.uk)
University of Leeds,
LS2 9JT Woodhouse, Leeds, United Kingdom

Nikolaus Kriegeskorte (nk2765@columbia.edu)
Columbia University, 3227 Broadway
New York, NY 10027 United States

Abstract

The Basel Face Model (BFM) provides an important tool for human face perception research, enabling the generation of realistic face images from a latent space of 3D shape and texture defining variables. The latent space is designed as an isotropic normal distribution reflecting the distribution of the 200 human faces whose 3D scans formed the basis of the BFM. However, this distribution does not reflect which of the faces look like natural human faces to people. We collected binary judgements of the naturalness of BFM faces and offer a model that predicts the probability that any BFM face will be judged as natural. This model contributes to our understanding of human face perception and will be useful to face perception researchers looking to sample natural-looking BFM faces.

Keywords: face perception, generative models, psychophysics

Introduction

Principal Component Analysis (PCA)-based 3-dimensional morphable models (3DMMs) have been widely applied in face perception research as a method to generate stimuli that capture the distribution of real human faces (Gerig et al., 2018; Egger et al., 2020; Walker & Vetter, 2016; Jozwik et al., 2022). Although the latent space of BFM is modeled as an isotropic Gaussian distribution, the Gaussian density does not provide a reliable indication of the subset of faces that look natural to people. To experience this, see Fig.1, left. In order to understand face perception and also to be able to sample natural-looking faces, it is desirable to delineate the subset of natural-looking faces within BFM. We performed online behavioral experiments in which human subjects gave binary judgments of face naturalness. We offer a probabilistic model that assigns a probability between 0 and 1 to each location in the BFM latent space, predicting the probability that the corresponding face will be judged as natural-looking. This approach promises not only to enrich our understanding of how humans discern natural from unnatural faces but will help researchers sample natural-looking faces from BFM.

Behavioral Experiment

Participants were presented with one synthetic face at a time and answered "yes" or "no" to the question: "Could there be a human face that looks like this?". Each face in the stimulus set varied along a single principal component, with the rest of the dimensions set to zero. We sampled faces whose euclidean distance to the average face ranged from 0 to 40 units in both (positive and negative) directions. A subset of 200 faces was viewed by every participant twice, which allowed us to evaluate the inter- and intrarater test-retest reliability. Additionally, each subject viewed a unique subset of 600 faces. In total, each participant judged 1000 faces in an online experiment conducted using Meadows (meadows-research.com) and Prolific.

Mahalogistic models

We introduce a model class that uses a scaling factor σ_i for each PC-dimension i of BFM to predict the binary judgements. The unnaturalness of each face is defined as the Mahalanobis distance from the origin of face space using a diagonal covariance matrix defined by parameters σ_i. The Mahalanobis distance of each face from the origin provides the input to a logistic regression model. We define the probability that a face with latent representation $\boldsymbol{\alpha}$ will be judged as natural by a random observer, by:

$$p_{nat}(\boldsymbol{\alpha}) = \frac{1}{1+e^{D(\boldsymbol{\alpha})-D_0}} = \frac{1}{1+e^{\sqrt{D_s(\boldsymbol{\alpha}_s)^2+D_t(\boldsymbol{\alpha}_t)^2}-D_0}} \quad (1)$$

where $D_s(\boldsymbol{\alpha}_s)^2, D_t(\boldsymbol{\alpha}_t)^2$ are the Mahalanobis distances given by the covariance matrices $\Sigma_{shape}, \Sigma_{texture}$ respectively. We assume the covariance matrices for shape and texture are diagonal matrices such that

$$D(\boldsymbol{\alpha}) = \sqrt{\boldsymbol{\alpha}^\top \Sigma^{-1} \boldsymbol{\alpha}} = \sqrt{\sum_i (\alpha_i/\sigma_i)^2} \quad (2)$$

Additionally, we assume $\sigma_i = f(i)$ where i is the PC-index and f is a smooth function of i. We use the following set of candidate smooth functions for the computation of the Mahalanobis distance:

$$f(i) \in \{mi, \quad mi+b, \quad mi^k, \quad mi^k+b\} \quad (3)$$

Finally, we introduce a parameter to account for the probability of motor error (pressing the unintended button independent of stimulus):

$$p_{nat}^{err}(\boldsymbol{\alpha}) = p_{err} \cdot [1-p_{nat}(\boldsymbol{\alpha})] + [1-p_{err}] \cdot p_{nat}(\boldsymbol{\alpha}) \quad (4)$$

We fit each of the four models to the "unique" face trial data using likelihood maximization (parameters: m, k, b, D_0, p_{err} for the full model). We included a 4-fold cross-validation procedure on the set of subjects. For statistical comparisons on the performance of the different models, we bootstrap resampled the data 1500 times. Additionally, we included a Gaussian Kernel Density Estimator as a reference model. This model's non-parametric approach, allows us to estimate the data's joint distribution, capturing its inherent structure and variability. This delineates a noise ceiling, benchmarking the performance of the rest of the models.

Results

In Fig. 2, we present heatmap visualizations of model predictions for idealized latent vectors. Specifically, these predictions are for the set of faces that are zero everywhere except for one PC at a time. The isoprobability lines delineate the regions in the latent space where faces are more or less likely to be perceived as natural. For both shape and texture trials, the Mahalogistic model defined by the variance function $f(i) = mi^k + b$ closely aligns with empirical observations. Our statistical analyses confirm that the $mi^k + b$ Mahalogistic model predicts the naturalness judgement data significantly better than every other Mahalogistic model for both shape and texture-varying faces. Moreover, the performance of this model exceeds that of the noise ceiling, notably outperforming the Gaussian KDE model in explaining the data (Fig. 3A, B, right). Finally, to give a clearer picture of what "natural" faces

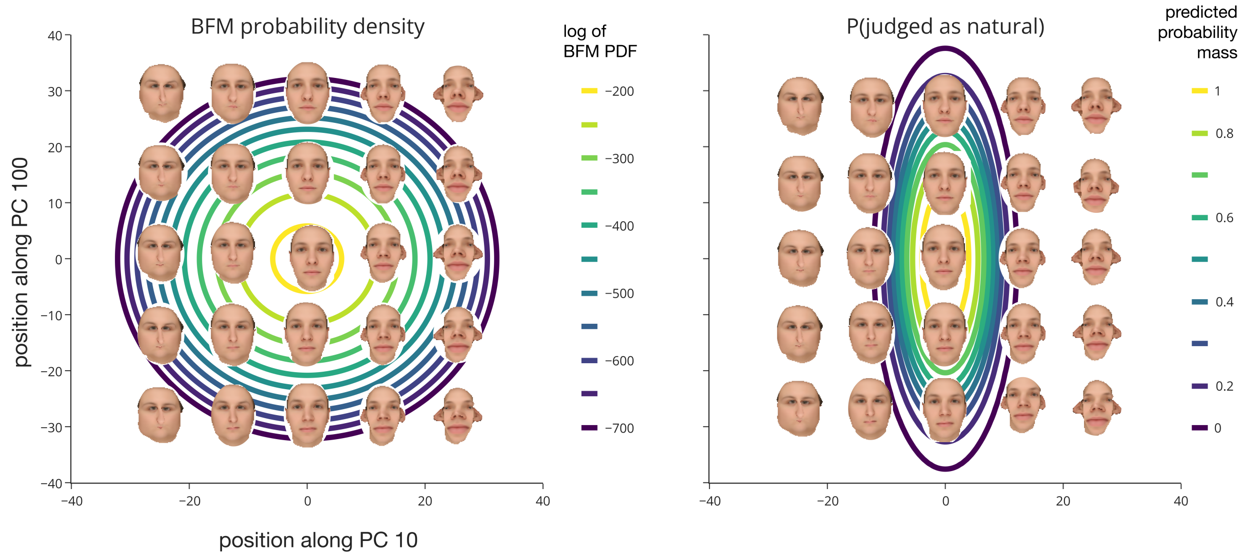

Figure 1: The left panel shows a set of example faces and their probability densities according to the Gaussian PDF defined by the BFM. Each face corresponds to a latent representation that is zero for every PC except for PC10 and PC100. The color-coded circles denote the the iso-probability-density contours. Faces that lie on the same circle are assumed to be equally likely by the model. The right panel shows the same set of example faces, this time with the predicted probability according to the $mi^k + b$ Mahalogistic model that the face will be judged to be natural. Iso-probability contours for the distribution learned on the basis of human judgements are shown. Faces that lie on the same line are predicted by the model to be equally likely to be judged as natural by a random observer.

look like according to the $mi^k + b$ model, we overlay a selection of faces—those varying only along PC10 and PC100—on top of the model's iso-probability contours (Fig. 1B).

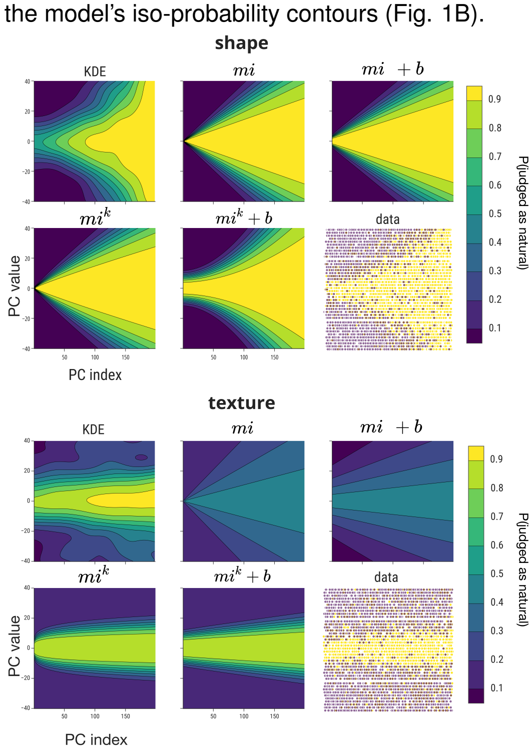

Figure 2: Each contour plot illustrates the probability that a face is perceived as natural by a random observer, as predicted by each of the candidate models. The x-axis represents the principal components (PCs), with the idealized latent faces characterized by zeroes everywhere except for a single PC. The y-axis indicates the value of the latent representation for the given PC. The last panel shows the behavioral data. Each dot represents one binary judgement by a participant to the corresponding face.

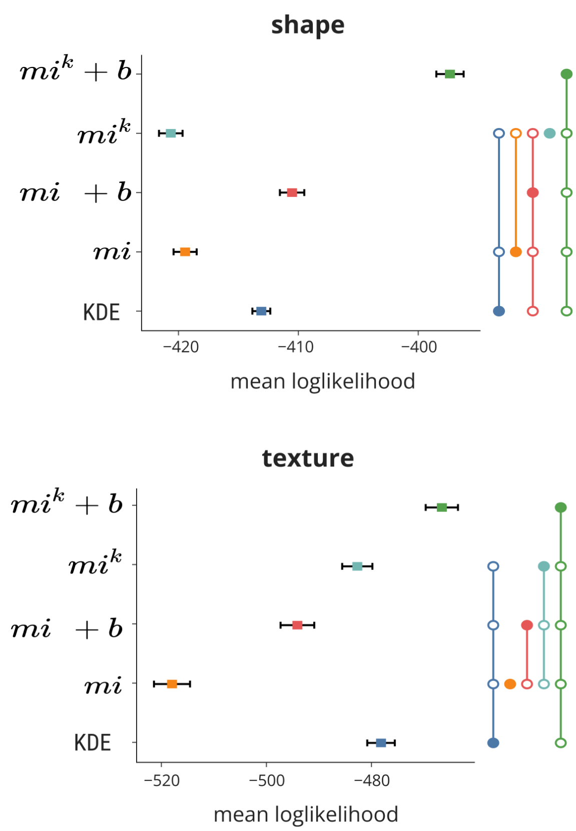

Figure 3: Each square on the plot represents the log-likelihood of the human perceptual judgment data given each candidate model (A, B; left). Error bars indicate the standard error of the mean (SEM) derived from 1,500 bootstrap resamples. Statistical significance is denoted as follows (A, B; right): A solid circle linked to open circles indicates that the model corresponding to the solid circle has a significantly higher log-likelihood compared to the models aligned with the open circles.

References

Egger, B., Smith, W. A. P., Tewari, A., Wuhrer, S., Zollhoefer, M., Beeler, T., ... Vetter, T. (2020, April). *3D Morphable Face Models – Past, Present and Future.* arXiv. (arXiv:1909.01815 [cs])

Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schoenborn, S., & Vetter, T. (2018, May). Morphable Face Models - An Open Framework. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 75–82). doi: 10.1109/FG.2018.00021

Jozwik, K. M., O'Keeffe, J., Storrs, K. R., Guo, W., Golan, T., & Kriegeskorte, N. (2022). Face dissimilarity judgments are predicted by representational distance in morphable and image-computable models. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(27), e2115047119. doi: 10.1073/pnas.2115047119

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, *110*(4), 609–624. doi: 10.1037/pspp0000064