



Co-Editors: Benjamin Balaz & Jonathan Flombaum

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/pvis20>

Exploring perceptual similarity and its relation to image-based spaces: an effect of familiarity

Rosyl S. Somai & Peter J.B. Hancock

To cite this article: Rosyl S. Somai & Peter J.B. Hancock (2022): Exploring perceptual similarity and its relation to image-based spaces: an effect of familiarity, Visual Cognition, DOI: [10.1080/13506285.2022.2089416](https://doi.org/10.1080/13506285.2022.2089416)

To link to this article: <https://doi.org/10.1080/13506285.2022.2089416>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



View supplementary material [↗](#)



Published online: 29 Jun 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Exploring perceptual similarity and its relation to image-based spaces: an effect of familiarity

Rosyl S. Somai  and Peter J.B. Hancock 

Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK

ABSTRACT

The lack of controlled stimuli transformations is an obstacle to the study of face identity recognition. Researchers are often limited to verbalizable transformations in the creation of a dataset. An alternative approach to verbalization for interpretability is finding image-based measures that allow us to quantify transformations. We explore whether PCA could be used to create controlled facial transformations by testing the effect of these transformations on human perceptual similarity and on computational differences in Gabor, Pixel and DNN spaces. We found that perceptual similarity and the three image-based spaces are linearly related, almost perfectly in the case of the DNN, with a correlation of 0.94. This provides a controlled way to alter the appearance of a face. In Experiment 2, the effect of familiarity on the perception of multidimensional transformations was explored. Our findings show that there is a significant relationship between the number of components transformed and both the perceptual similarity and the same three image-based spaces used in Experiment 1. Furthermore, we found that familiar faces are rated more similar overall than unfamiliar faces. The ability to quantify, and thus control, these transformations is a powerful tool in exploring the factors that mediate a change in perceived identity.

ARTICLE HISTORY

Received 15 July 2021
Accepted 4 June 2022

KEYWORDS

Perceptual space; similarity;
image-based spaces; PCA;
familiarity




Introduction

The lack of controlled stimuli transformations is an obstacle to the study of face identity recognition. The aim of this work is to explore a method for generating variations in the appearance of a face image, and methods for assessing the likely effect on human perceptions of such variations. We then use these methods to test the effect of familiarity on the perception of suitably controlled alterations to faces.

Most research that explores facial variation, uses the approach of reverse engineering: manipulating the visual input to assess whether this content is used in face identity recognition. For example, changing facial components such as the nose, mouth and eyes (Abudarham & Yovel, 2016) or displacing them (Goffaux & Rossion, 2007). This approach is useful but requires well-understood transformations of the visual input to be meaningful. For example, a change in eyebrow thickness is easily verbalized and can be varied systematically, but a change based on “morphing” – gradually changing from one identity

to another – is difficult to interpret in terms of dimensional change. Verbalizable transformations are insightful but are only a part of the complete story.

In order to successfully identify a face under a wide range of viewing conditions, our internal representations of a face need to capture variability in appearance, in order to “tell people together”, but also tell people apart (Burton et al., 2011; Jenkins et al., 2011). Burton et al. (2016) explored the way faces vary using PCA, and found that the variation is idiosyncratic. One identity may vary in ways that another does not. As an example, someone with a prominent brow ridge will show more variation in lighting on their eyes. Although the exact nature of between and within face variability has yet to be determined, it is clear that this variation goes beyond easy to verbalize features. An alternative approach to verbalization for interpretability is finding image-based measures that allow us to quantify image transformations. Such an approach would

CONTACT Rosyl S. Somai  rosylsoma@gmail.com  Psychology, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/13506285.2022.2089416>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

need image-based measures that preferably correlate with perceptual space, and well-controlled transformations.

Principal Component Analysis (PCA) was first proposed as a means of face recognition by Kirby and Sirovich (1990), and Turk and Pentland (1991). PCA produces a compact representation of a face but is very sensitive to image properties such as lighting. Nevertheless, using controlled images, PCA will reveal systematic differences between faces, such as gender (O'Toole et al., 1998). The analysis is greatly improved by first morphing all faces to the same average shape and performing PCA on the shape vectors and the "shape free" faces separately (Craw & Cameron, 1991). This separate analysis yields significant correlations with human perceptions of memorability and distinctiveness of faces (Hancock et al., 1996). It can also be used to synthesise novel face images (Hancock, 2000) and by extension, provides a controlled way to produce changes in the shape or surface appearance of face images. It is this ability to produce controlled transformations that is exploited here.

One aim of this work is to test the utility of PCA-induced changes in face images to probe human face perception. PCA can produce graded changes and it is to be expected that a large PCA change will produce a bigger perceptual difference than a smaller one. However, it would be helpful to have a separate computational measure of the likely perceptual effect of varying different components or combinations of them. If a purely image-based measure (without a biological premise) correlates with perceptual ratings, it can be used in the creation of stimuli to make a prediction of how a human observer will perceive the image. For example, one could predict whether a certain image transformation is strong enough for the altered face and the original face to be perceived as different from one another. Here, we test three different computational measures of image similarity.

The simplest measure of image similarity is pixel-wise Euclidean distance. For this to have any chance of working for face images, they need to be very carefully controlled. A simple misalignment between two copies of the same image will produce a large difference, while changes in lighting will also result in a large difference that may be barely noticed by a human observer. Because of this, the measure is little tested in face perception. However, since our images are generated by a PCA model that is based

on averages and we select the later components which result in filtering out variation such as lighting and angle (for details, see Methods: apparatus and stimuli), they are inherently well-controlled. Meytlis (2011) used carefully controlled images and found an almost perfect correlation between distinctiveness ratings and Pixel Euclidean distance.

A more biologically plausible measure commonly used in computer vision (Lades et al., 1993; Wiskott et al., 1997) and neurocomputational models of face recognition (Dailey & Cottrell, 1999; O'Reilly & Munka, 2000) is Gabor filtering. Gabor filters are found to be a good representation of how cells in the primary cortex respond to a certain scale and orientation (De Valois & De Valois, 1990; Jones & Palmer, 1987; Ringach, 2002). As Gabor filtering reflects very early processing and not any higher order processing, it would seem unlikely that Gabor distances would relate to perceptual similarity of faces. However, this is exactly what Yue et al. (2012) found: Gabor-based image measures strongly correlate with human performance in a discrimination task with complex figures. For this reason, we included Gabor space to quantify image transformations.

Current artificial face recognition deploys deep neural networks (DNNs) and their face matching ability easily surpasses human performance in many situations. Hancock et al. (2020) compared six state-of-the-art DNNs with human performance on four face matching tasks designed to be challenging for human participants. The humans averaged 73.3%, while the best DNN achieved an accuracy of 98.6%. Despite the difference in accuracy, there were significant correlations between the six DNN similarity scores for a pair of faces and the human accuracy on that pair. This was especially true for matching pairs, where the two images do show the same person. Humans and the DNNs seem to agree on what makes faces look similar. However, the correlation for mismatch face pairs was lower, indicating that humans and the DNNs show less agreement on what makes faces look different. Indeed, the DNNs worked in distinctly non-human like ways, often reporting that faces which differed in apparent race or sex were the same identity. Given these discrepancies, it is valuable to include DNN similarity measures in the present analysis of controlled image changes.

Hancock et al. (1998) looked at correlations between human similarity ratings and PCA on existing

face images. Here, we invert that process and ask humans to rate the similarity of face images that have been altered in a controlled way by varying principal components. We compare the human data with similarity scores obtained from Gabor, Pixel and DNN analysis of the same images. In Experiment 1, we test the effects of single component transformations of both colour and shape, using unfamiliar faces, with the aim of providing better methods for choosing stimuli. In Experiment 2, we test the effect of multiple component transformations on both familiar and unfamiliar faces. This allows us to test the effect of familiarity on perceptual similarity.

Experiment 1 – exploring the effects of single component transformations

Background

One major issue in exploring face identity recognition is finding an image transformation that reflects a likely dimension in face space which can also be interpreted in some meaningful way. PCA transformations can be difficult to interpret, as they reflect variation in pixel values that commonly occur together. For example, a PCA component that changes eye colour will also change hair colour, because people with darker eyes often have darker hair than people with lighter eye colours. Thus, verbalizing the change of PCA components can be difficult. On some occasions, a component can be classified after visual review, such as a component that varies apparent masculinity. However, for most components the variation is difficult to classify. In this experiment, we will explore the effect of single component transformations on Gabor, Pixel and DNN space, and (human) perceptual space in order to assess their usefulness in quantifying image transformations.

Methods

Participants

We tested 25 participants (11 female, 13 male, 1 other, mean age = 39.7 years, 2 missing values, SD = 16.3 age range = 12–74 years). Participants were naive to the purpose of the experiment and took part on a completely voluntary basis without any monetary compensation. The experiment was approved by the

local General University Ethics Panel of the University of Stirling.

Apparatus and stimuli

The experiment was programmed through the online platform Qualtrics and participation was allowed via PC, laptop, phones and tablets. Face image stimuli were generated from a PCA-based face model, written in Matlab (MATLAB, 2019). The PCA space was calculated from 72 eye-aligned images, each an average of 10 images of a famous (mostly Caucasian) female. Psychomorph (Tiddeman et al., 2001) was used to create the averages. Use of average images helps to reduce variations due to lighting and viewpoint. Even so, the first few components encode image properties such as lighting and head movements, so we selected components 10–16 to alter. Three different types of transformations were used: shape-only, colour-only and both. To explore the effect of the size of the transformation, two transformation sizes were used: 3 and 6 SD. The SD for each component was scaled according to the variance in the (72- eye-aligned) original images. This gave 42 pairs in total: 7 shape, 7 colour and 7 both, two transformation sizes each. For each pair a different face image was generated by setting the first 20 shape and colour components to a random normal value, mean zero, SD 1, though truncated at $\pm 2SD$ to prevent faces looking too strange. An example of the transformation is shown in Figure 1. In addition to the PCA transformations, three single feature transformations derived from Independent Component Analysis (ICA) were added. However, these were purely exploratory; a discussion is given in supplementary materials. The images produced were all 420 by 595 pixels, size in the experiment was browser and screen dependent.

Procedure

Participants were shown an information sheet and then gave consent online. The face matching task consisted of 45 trials (14 alterations for three conditions and the 3 exploratory ICA transformations) that were presented in a random order. In each trial, two faces were presented along with the question “Could these be two pictures of the same person? 1 – certain no; 2 – think no; 3 – guess no; 4 – guess yes; 5 – think yes; 6 – certain yes”. An example of a trial is shown in Figure 2.

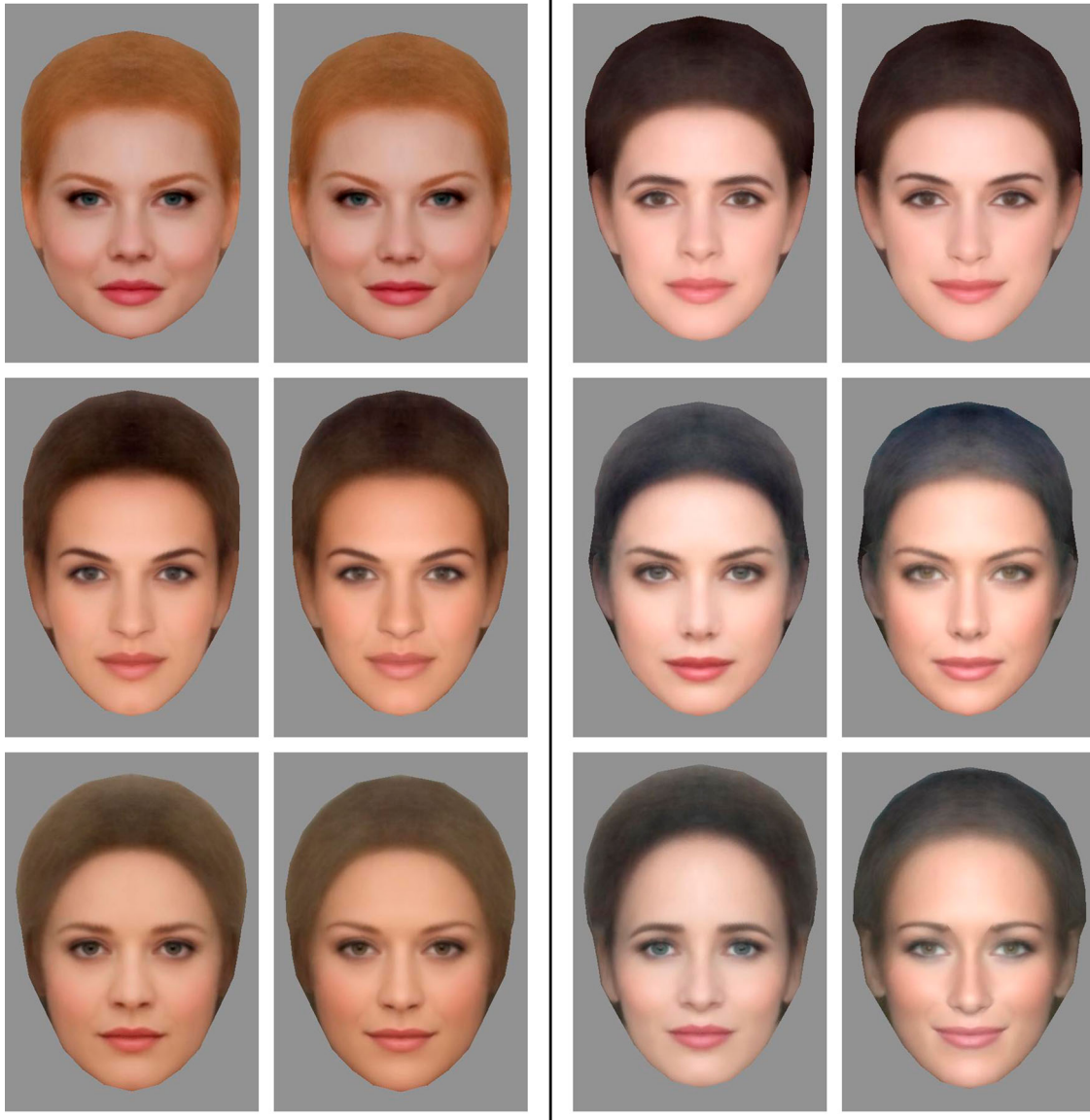


Figure 1. Example of the transformations used. On the left, 3SD transformations and on the right 6SD transformations. For each size of change, the left column shows one end of the scale which is -1.5 and -3 SD from the original respectively and the right column is the other end of the scale which is $+1.5$ and $+3$ SD. The top row shows shape transformations only, the middle row shows colour transformations only and the last row shows shape and colour transformations. The base image for each pair was generated at random, as described in the text.

The three computational spaces

Gabor

We used the Gabor Features in Signal and Image Processing Toolbox (Kämäräinen et al., 2002b, 2002a), written in Matlab (2019). This requires a 256 square input image, so the original images were cropped to just below the chin, resized and converted to monochrome, Figure 3. We used the default settings of the toolbox, which computes phase and magnitude for 8 orientations and 5 scales across a 10×10 grid, giving a vector of 8000 values for each image. The

difference between two images is given by the Euclidean distance between the two vectors, calculated using the norm function in Matlab (MATLAB, 2019).

Pixel

A simple way to compare two images is Euclidean distance in pixel space. Since our images are all aligned by the eyes, pixel distance might plausibly capture the perceptual changes caused by variations purely in the colour space. However, shape transformations will result in different face features being compared,

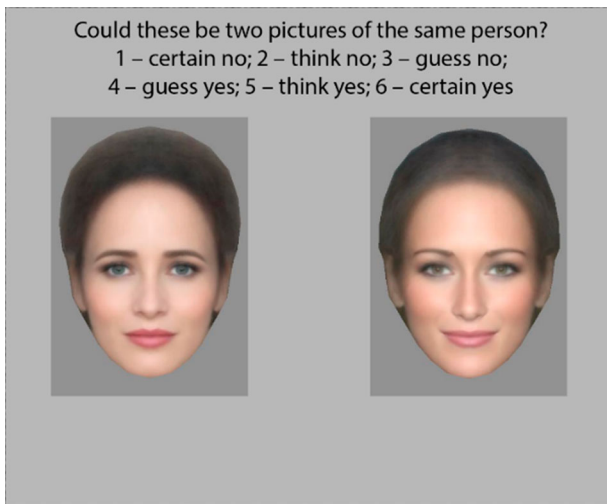


Figure 2. Example of a trial.

or face pixels in one image compared with background in another. Meytlis (2011) normalized each image to the same average grey level and cropped them to a tight oval. Since we are comparing pairs of images that differ by adding a principal component that is inherently zero mean we did not normalize the colour values, but did crop tightly around the face with an oval 378×277 pixels, see Figure 3. The difference between images was computed by concatenating the RGB vectors and calculating the Euclidean distance between them using the norm function in Matlab (MATLAB, 2019).

DNN

The DNN used was an experimental system from the University of Surrey. It produces an output vector of length 512. The similarity of the two images was given by the cosine of the angle between the two output vectors. Identical images give a similarity of 1, with smaller values being less similar. This is in contrast to the Euclidean distance of Gabor and Pixel spaces where larger numbers mean less similar.

Results

We treat our participant responses, about whether the two faces look like different people, as a similarity scale, with 6 being the most similar. On a 6-point scale, 3 was “guess different” and 4 was “guess same”. Thus, the boundary for same or different identity was at 3.5. We found a significant negative correlation between the human similarity ratings and the Gabor distances ($r = -.87$, $p = <.001$, shown in Figure 4), the Pixel distances ($r = -.85$, $p = <.001$, Figure 5), and the DNN distances ($r = .94$, $p = <.001$, Figure 6).

Discussion experiment 1

The linear relationship between the DNN similarity space and human similarity is remarkable. The human response scale is, strictly speaking, ordinal: there is no reason to think that the differences between “guess”, “think” and “certain” should be the same number, and the same as the difference between “guess same” and “guess different”. The DNN similarity space is given by the cosine of the angle between the feature vectors. Cosine is clearly non-linear, although most of the non-linearity occurs in the region above 0.8 and most of our points are below that, so within the relatively linear region. Yet these two non-linear spaces, averaged over a number of observations, contrive to give an almost perfect straight-line relationship.

Previous work indicates that humans and DNNs seem to agree on which faces look similar but show less agreement on what makes a face look different (Hancock et al., 2020). One noteworthy aspect of that study was the necessity to use rank correlations, because most of the DNNs appeared to have a marked non-linearity towards the top end of the scale. However, similarity scores of the commercial systems used were not based on the raw similarity between the representational vectors for each pair

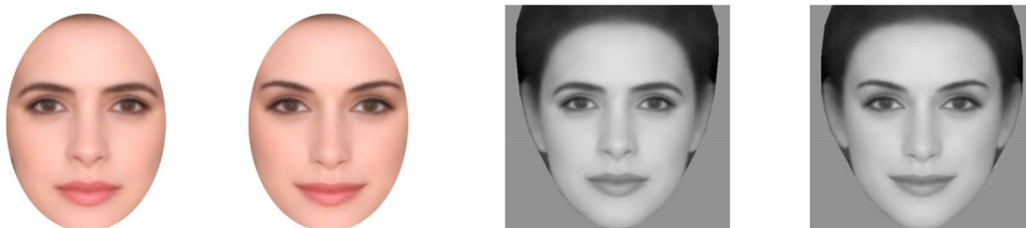


Figure 3. Example of cropped images, differing in shape component 10. Left, for Pixel comparison, right for Gabor.

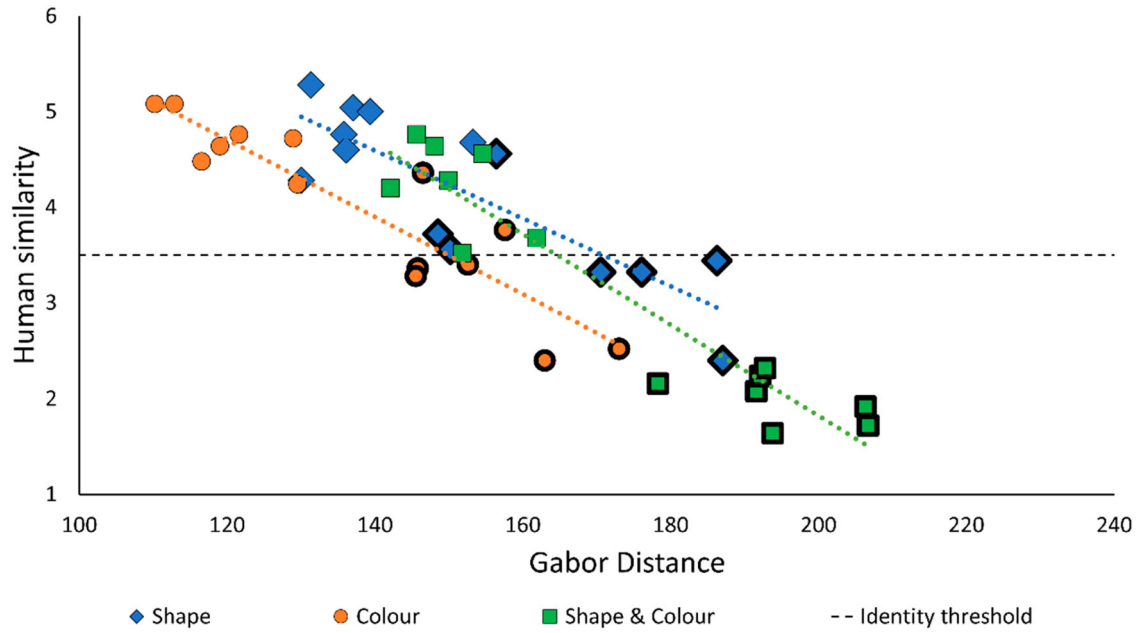


Figure 4. Correlation between perceptual similarity and Gabor distance. The 6 SD transformations are indicated by a bold outline.

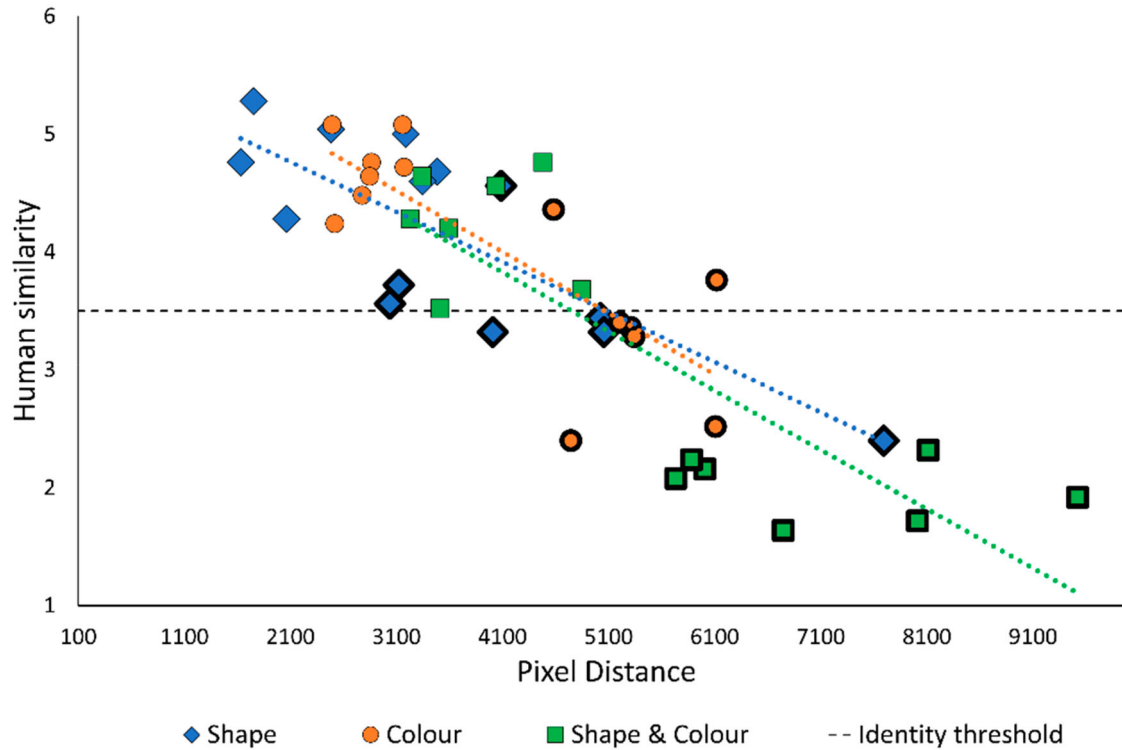


Figure 5. Correlation between perceptual similarity and Pixel distance. The 6 SD transformations are indicated by a bold outline.

of images, but on a black-box verification method. It appears these verification methods have the effect of increasing the reported confidence of a match between sufficiently similar images. We tried some commercial DNNs on the data reported in Figure 6

and they produced a clear “hockey stick” shape (see supplementary analyses). The commercial systems overlook minor differences between face images, to which human observers are sensitive. It therefore seems that the linear relationship observed in the

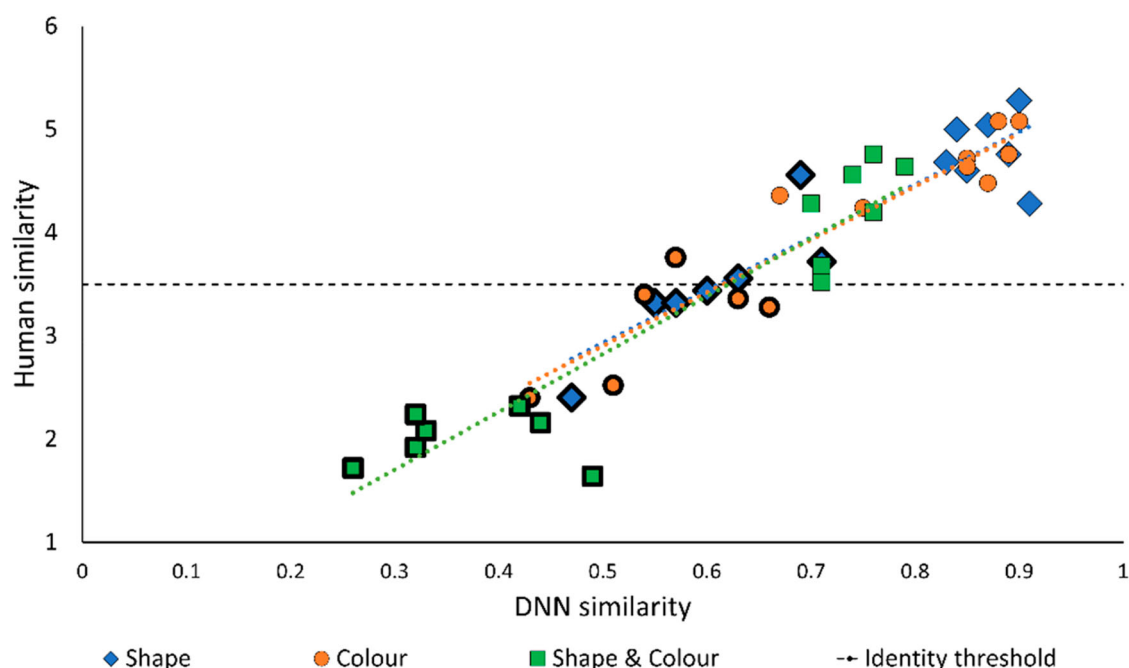


Figure 6. Correlation between perceptual similarity and DNN similarity. The 6 SD transformations are indicated by a bold outline.

current study results from using the relatively raw similarity between the representational vectors for each pair of images. Thus, understanding the computation of the similarity scores of systems is important when using DNN similarity scores to predict human similarity scores, because the computation affects the form of the relationship.

Similar to the findings of Meytlis (2011) on Euclidean distance, our results show that all three image-based distance values correlate significantly with perceptual similarity. Although the results are straightforward, it remains unclear if these linear relationships are also present when multidimensional transformations are used. In Experiment 2, multidimensional transformations are explored by using transformations with a varying number of principal components changed.

Experiment 2: the relationship between the number of principal components transformed, perceptual similarity and familiarity

Background

This experiment is similar to Experiment 1, but with multidimensional instead of single-dimensional transformations. Specifically, we explore the effect of changing different numbers of PCs on perceptual, Gabor,

Pixel and DNN space. Based on the findings of Experiment 1, we hypothesize that the relationship between the number of PCs changed and any of the four image spaces is linear. Since it is well known within face identity research that people are better at processing familiar than unfamiliar faces (e.g., Johnston & Edmonds, 2009) and representations of familiar faces are more robust (Burton et al., 2005), we included familiarity to explore the effects of familiarity on perceptual similarity. Therefore, transformations in this experiment were performed on famous faces, as opposed to the unfamiliar, randomly generated faces in Experiment 1. A familiarity check was added in the beginning of the experiment.

Although there is an abundance of research on the effect of familiarity, it remains unclear what underlies this enhanced performance to recognize familiar over unfamiliar faces. A common proposal is that the representation and processing of familiar faces is different from that of unfamiliar faces (Johnston & Edmonds, 2009; Ramon & Gobbini, 2018). Recent work from Ritchie and Burton (2017) shows that variability is important for learning a new face. Additionally, the variability within images of a specific identity appears to be idiosyncratic information that is key for learning new faces (Burton et al., 2016).

That two images of a person look more similar when familiar is almost the definition of familiarity.

Someone familiar with the person can identify them both as being the same individual, while someone unfamiliar has to rely on what they can infer from the appearance (Jenkins et al., 2011). Here, we are applying fixed changes to faces, some of which will be familiar to a given participant and others not. It is plausible that in this situation, viewers familiar with a face will be better able to identify a change: e.g., “that’s X, but the nose is wrong”. They may be able to use their existing memory of the face to identify the changes.

The results from Beale and Keil (1995) make a more subtle prediction. They used morph sequences between different individuals’ faces to test the effect of familiarity on the decision boundary. They found that familiar face pairs had a sharper decision boundary. Moving along the morph continuum, the face looked like the first identity until nearly halfway, when it rapidly shifted to looking like the other identity. For unfamiliar face pairs, the shift was more linear, with a gradual change from one identity to the other. The prediction from this for our imposed changes in appearance is that, for a familiar face, the changed images should look like the original up to some critical point, where it will rather suddenly no longer look like them. If unfamiliar with the original face, then the changes should produce a more linear change in similarity. Therefore, the prediction is that changed faces will initially look more similar when familiar than unfamiliar but that, if the change becomes large enough, this will reverse, with changed familiar faces looking more different.

In these experiments, we are asking participants to decide about probable identity, rather than about similarity directly. So, while those familiar with the face may be able to say, “the nose is wrong”, we expect participants to say “same identity” with more confidence if they are familiar with the face. Whether we see the reversal predicted by the results of Beal and Keil will depend on whether the changes are sufficiently large.

Methods

Participants

We tested 61 participants (43 females, 16 males, 1 other, 1 missing value, mean age = 29.9 years, SD = 9, age range = 20–68 years). Testing was done

through an online platform, Testable. Participants were naive to the purpose of the experiment. All participants agreed to an informed consent form after a written explanation of the procedure. Participation was on a completely voluntary basis without any monetary compensation. The experiment was approved by the local General University Ethics Panel of the University of Stirling.

Apparatus and stimuli

The experiment was programmed through the online platform Testable and participation was only allowed via a PC or laptop. A calibration to control stimulus size was performed by matching the size of an on-screen bar to a credit card, to ensure proper visualisation throughout the experiment. Stimuli preparation and creation were done using Matlab (2019), using the same PCA space as described in Experiment 1. Eight of the 72 average images that went into the PCA were chosen for use in this experiment. The stimuli consisted of these 8 average images, regenerated exactly from the model by using all the components, and variations that had either 2, 3, 5 or 7 colour components (components 10 and 11; 10–12; 10–14; 10–16, respectively) in the PCA space changed. The size of the change was 4 SD. To avoid any directional effects from the PCA transformation, for half the trials the negative transformation and the other half the positive transformation was used. As shown in Experiment 1, any of the three spaces can be used for insight into the size of the perceptual change. We applied this principle in this experiment to ensure that the effect of the transformation was similar for all identities. Note that we would expect some differences between identities. Suppose a component affects the darkness of the eyebrows. At the image level, a face with thin eyebrows will change less than one with thick eyebrows. Figure 7 shows the Gabor distance between each original image and its four variations. It can be seen that the pattern of differences is very similar for each of the identities, suggesting that we should not expect any marked differences between items in our set. An example of the transformation can be found in Figure 8.

Procedure

The experiment consisted of a familiarity check and a face matching task. In the first 8 trials, participants

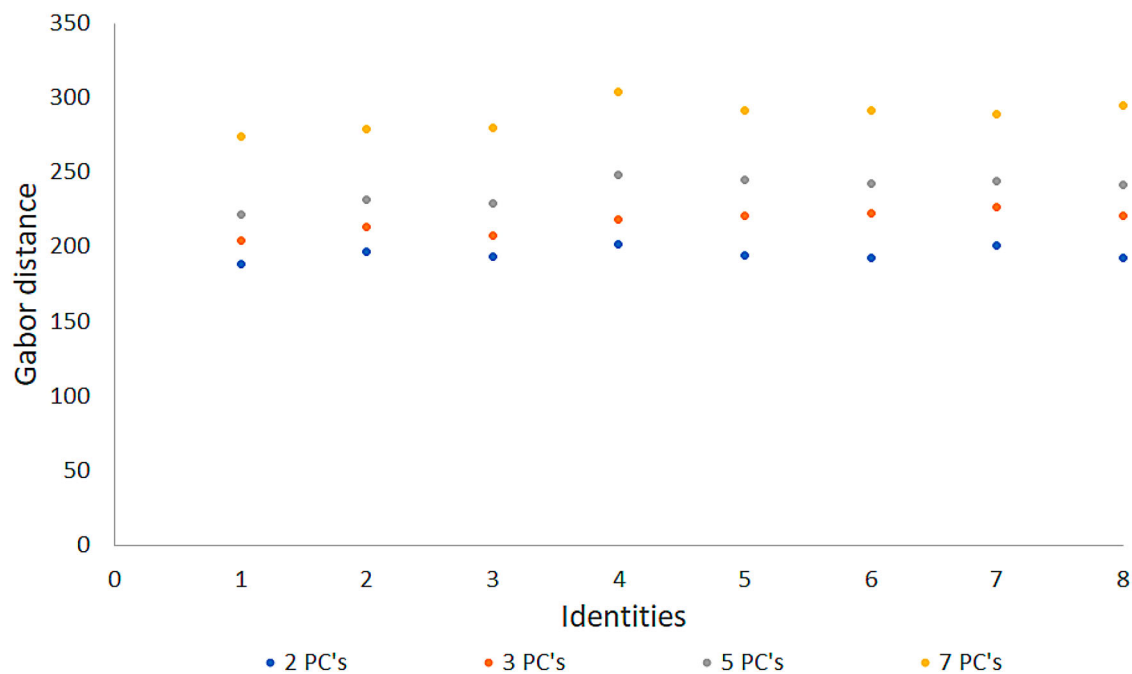


Figure 7. Gabor distance between the original and altered images to check if the transformations had a similar effect in Gabor space for all the identities.

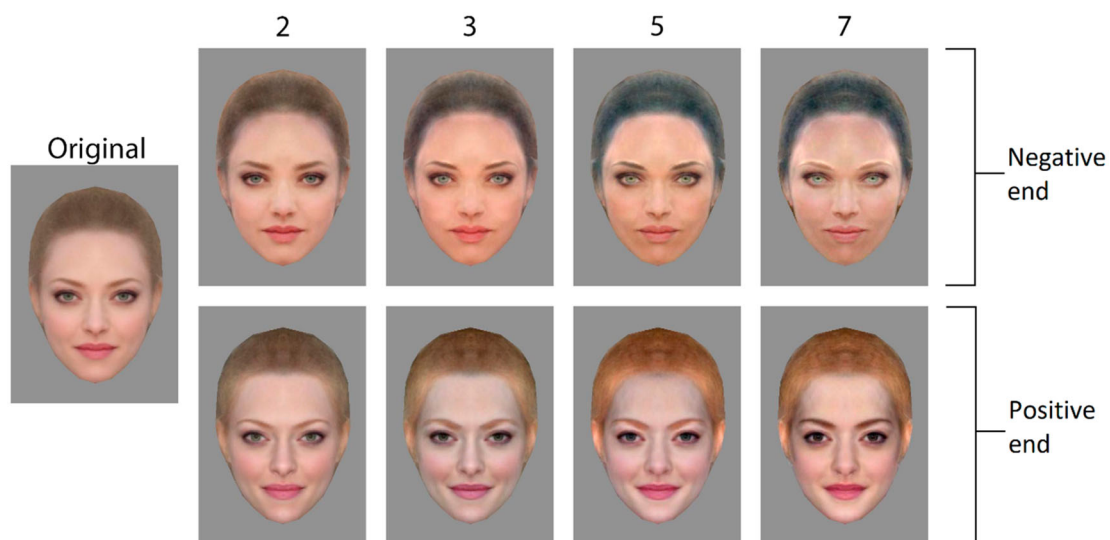


Figure 8. An example of the transformation: the negative end images are -4 SD and the positive end images are $+4$ SD from the original. The number above the images indicate how many components are changed. In the experiment the original was paired with transformations of either the positive or the negative end.

were asked “Would you recognise this person on the street? yes(y)/ no(n)” for each identity (presented in a random order) to establish prior familiarity. Afterwards, participants were presented with the face matching task instructions and had 2 initial practice trial with an identity not included in the actual experiment. This was followed by 32 trials in a random

order. In each trial, the original was shown, followed by a mask, the altered image and another mask. Finally, the participants were presented with the question “Could these be two pictures of the same person? 1 – certain no; 2 – think no; 3 – guess no; 4 – guess yes; 5 – think yes; 6 – certain yes”. An overview of a trial is shown in Figure 9.

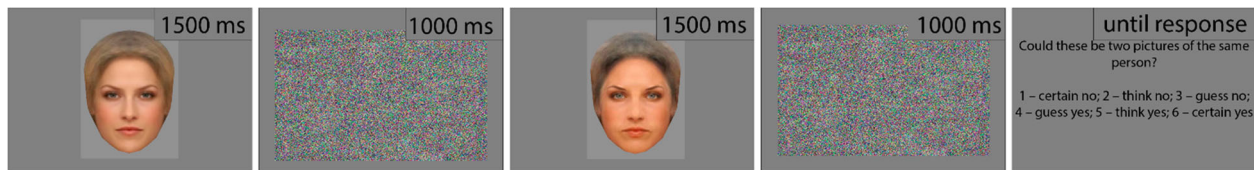


Figure 9. Overview of a trial.

Data analysis

To explore the effects of familiarity, we constructed a linear mixed effects model with the similarity rating as dependent variable. The model included Familiarity and Transformation level as categorical factors (Familiarity: familiar and unfamiliar rescaled to -0.5 and 0.5 , respectively, to allow for the analysis of the interaction; Transformation Level: 2, 3, 5, or 7), and the intercept, slope and interaction of familiarity and Transformation level in the performance per participant as random effects. Due to convergence issues, the main effects were excluded from the random effects. We report the parameter estimate (b), standard error (SE), t value and p value. The linear mixed effect analysis was performed in R (R Core Team, 2019) with *lmerTest* (Kuznetsova et al., 2017). The threshold for significance was set at $\alpha = 0.05$. Normality of the residuals on the model were inspected visually and showed no violation.

Results and discussion

In the current study, we explored the effect of multi-dimensional transformations on the three image-based spaces and perceptual similarity. Additionally, we investigated the effect of familiarity on similarity ratings.

We found a significant correlation between the human similarity ratings and the Gabor distances ($r = -0.77$, $p < 0.001$), the Pixel distances ($r = -0.76$, $p < .001$), and DNN distances ($r = .79$, $p < .001$), see Figure 10. As expected, the changes at each level affected the different identities to different extents. However, the overall effect of increasing the number of changes is consistent.

The effect of size of change and familiarity on perceptual similarity

The descriptive statistics of Experiment 2 are shown in Figure 10. Table 1 in the supplementary materials shows the distribution of familiarity of the 8 identities.

The results show that increasing the number of components transformed decreased the similarity rating for all levels compared to the base change (2 components changed, all p -values are $p < 0.001$, see Tables 2–4 in supplementary materials). Interestingly, the similarity rating between 5 and 7 components did not significantly differ (estimate = -0.14 , SE = 0.09 , $t = -1.62$, $p = 0.106$). Secondly, we found that participants rated familiar faces as more similar overall (estimate = 0.34 , SE = 0.13 , $t = 2.59$, $p = 0.010$; see Figure 10). Lastly, there was no interaction found between familiarity and level of change (see Tables 2–4 in supplementary materials).

The lack of a significant difference in similarity rating between 5 and 7 components (see Figure 10) is surprising because the difference in Gabor space between 5 and 7 components changed is the biggest in the whole stimuli set (see Figure 7). Therefore, one might expect the difference in perceptual distance would be similarly large, but this was not found. As one could see, the relationship is linear until 5 components. This could indicate that Gabor space and perceptual space do not relate linearly for bigger transformations. Thus, the lack of a significant difference could be an indication that more multidimensional transformations are perceived differently. Alternatively, it could be simply an effect of encoding rate and limited presentation duration: the presentation of all faces was 1500 ms so participants might not have had enough time to perceive all the changes.

As described above, the work of Beale and Keil (1995) predicts that changed faces will initially look more similar when familiar than unfamiliar but that, if the change becomes large enough, this will reverse, with changed familiar faces looking more different than when the same transformation is applied to an unfamiliar face. If this prediction became true, there would be a main effect of familiarity and an interaction effect with transformation level. As expected, we observed greater perceptual

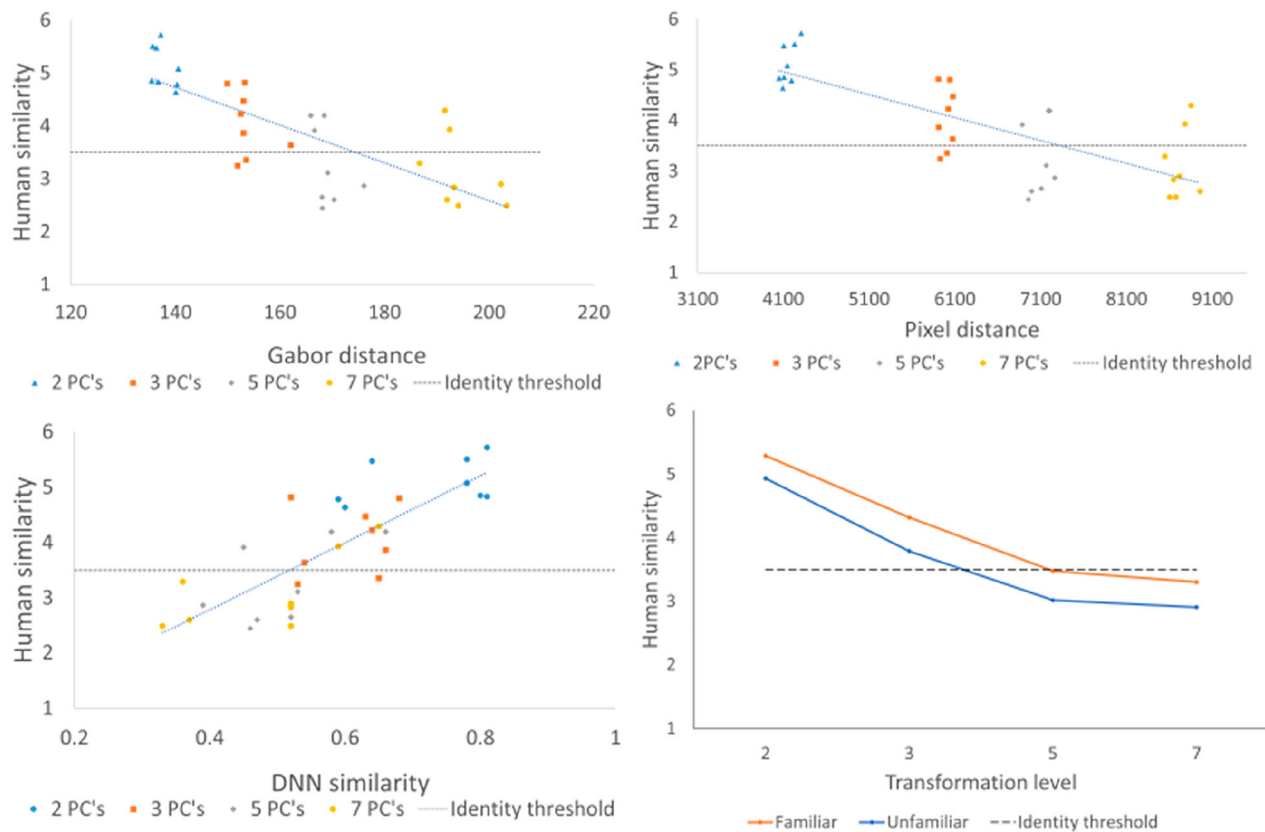


Figure 10. Top row and bottom left: Correlations between perceptual similarity and the 3 image-based space. The trendline is fitted to the data of all change size groups. Bottom right: Average similarity rating per level of change grouped by familiarity.

similarity for familiar over unfamiliar faces. However, the expected interaction with the transformation level was not found. Based on the current experiment alone, it is impossible to say if the prediction with regard to the increase in dissimilarity for familiar faces was wrong because the change could have simply been not great enough to change the identity to the extent needed to increase dissimilarity, i.e., cross and go beyond the border of a category. The current paradigm is an important first step to gaining more insight in how familiarity influences perceptual similarity and it offers a clear-cut way of comparing other transformations that will be used in future research.

General discussion

In this paper, we used transformations based on PCA to assess the relationship between perceptual similarity and Gabor, Pixel and DNN distance values. Experiment 1 explored the effects of shape and colour changes on the size of the perceptual change and the three image-based spaces. The results

showed that the PCA transformations used in this study result in a linear relationship between the three spaces and human similarity ratings. In addition to assessing the relationship between perceptual similarity and Gabor, Pixel and DNN distance values, Experiment 2 explored the effect of altering different numbers of principal components on the size of the perceptual change, and the effect of familiarity. Similar to Experiment 1, the relationship between each of the spaces and perceptual similarity was linear. Additionally, our findings show that there is a significant relationship between the number of components changed and both the perceptual similarity and the three image-based spaces and that participants rated familiar faces as more similar overall. These findings combined indicate that there is a clear relationship between perceptual and our image-based spaces.

Our findings support those of Yue et al. (2012) that Gabor-based image measures correlate highly with human performance in a discrimination task with complex figures and Meytlis (2011) that there is a relationship between distinctiveness ratings and

pixel-wise Euclidean distance. These studies and ours all had different tasks. In Yue et al. (2012), participants performed a match-to-sample task and in Meytlis (2011) participants were asked to rate single faces on distinctiveness and face pairs on perceptual distance (rating how different they looked). In both experiments the stimuli were simultaneously on-screen. Since it is known that availability of stimuli reduces the use of visual working memory and promotes the use of eye movements to gather visual input when needed (Somai et al., 2019), one could argue the observed relationship between perceptual and image-based spaces is due to the ongoing presence of the stimuli that serves as an external representation. Since the external representation is an image and the spaces are image-based, this could mean the findings do not reflect face space but simply the use of visual input. However, in this paper, Experiment 2 includes a memory component because there is a delay between the two faces rated and the findings are similar. This indicates that the internal representation of faces does indeed relate to image-based spaces.

We do not want to imply that the 3 chosen image-based spaces are superior to any others. This is simply unknown at the time of this paper, and these spaces were chosen based on personal experience with them, and the – albeit limited – existing literature (De Valois & De Valois, 1990; Hancock et al., 2020; Jones & Palmer, 1987; Meytlis, 2011; Ringach, 2002; Yue et al., 2012). Additionally, the Gabor and Pixel distance values are easily obtained and understood. The DNN used in this particular instance is not openly available, and other systems with different architecture and training might produce different results (see supplementary material; although the 6 DNNs tested in Hancock et al. (2020) showed similar, significant correlations). In the context of creating a human-like neural network, it will be interesting to compare the differences in DNN spaces and how they relate to perceptual measures.

The primary aim of this paper was to test the utility of PCA-induced changes in face images to probe human face perception. Our findings show that the graded changes produced by PCA translate to a graded change in perceptual space. In other words, a large PCA change indeed produced a bigger perceptual difference than a smaller one. Additionally, the bigger the number of principal components that

were altered, the bigger the perceptual change (up until 5 components). Another argument for the use of PCA-induced changes is the potential insights that can be gained from understanding why it describes perceptual similarity quite accurately in some cases and not in others. For example, in Experiment 2, 5 out of 8 faces with 5 principal components changed were rated by participants as dissimilar and 3 as similar. One could wonder why this variation exists: what makes some faces appear different with 5 principal components changed while other faces are still perceived the same? Did the specific components changed have a different effect on the different faces? For example, if someone has especially distinctive eyes, then a change to the eye region may have more perceptual effect than on a face with more typical eyes. In the current study, the faces had all the same components changed and no specific selection was made in identities.

To our knowledge, very little research has been done on the role of familiarity in the relationship between (systematic) changes to a face and perceptual similarity. Our secondary aim was to use the paradigm proposed in this paper to explore this and we found a main effect of familiarity but no interaction effect. Although this work is just the first step in exploring the factors that mediate a change in perceived identity, future work can be easily related to these findings because of the relationship between the image-based spaces. Regardless of the transformations, datasets and images used, it can be quantified and compared.

In conclusion, our study shows the predictive nature of Gabor, Pixel and DNN distances in the estimation of perceptual changes in faces. This relationship can be used in the exploration of the dimensionality of face space and facilitate in stimulus creation. For example, in research related to variability and learning, the experimenter is often limited in creating a high and low variability set based on observation. Any of the three image spaces can be used to quantify variability and study the effects on a more gradual scale. Lastly, Experiment 2 presents a framework on how to use controlled multidimensional face transformations to explore differences in the perception of familiar and unfamiliar faces. The ability to quantify, and thus control, these transformations is a powerful tool in exploring the factors that mediate a change in perceived identity.

Acknowledgments

The authors would like to thank Dr. Anthony Lee for advice and guidance on the statistical analyses.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

P.J.B.H. is supported by EPSRC grant number EP/N007743/1, Face Matching for Automatic Identity Retrieval, Recognition, Verification and Management. R.S.S. is supported by the Dylis Crabtree Scholarship grant number 1.

Data availability statement

The data related to the findings in this study are openly available at <https://osf.io/52kp4/> [view only for the time of review].

ORCID

Rosyl S. Somai  <http://orcid.org/0000-0002-8644-1282>

Peter J.B. Hancock  <http://orcid.org/0000-0001-6025-7068>

References

- Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision*, 16(3), 40, 1–18. <https://doi.org/10.1167/16.3.40>
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57(3), 217–239. [https://doi.org/10.1016/0010-0277\(95\)00669-X](https://doi.org/10.1016/0010-0277(95)00669-X)
- Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51(3), 256–284. <https://doi.org/10.1016/j.cogpsych.2005.06.003>
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102(4), 943–958. <https://doi.org/10.1111/j.2044-8295.2011.02039.x>
- Burton, A. M., Kramer, R. S. S., & Ritchie, K. L. (2016). Identity from variation: representations of faces derived from multiple instances. *Cognitive Science*, 40(1), 202–223. <https://doi.org/10.1111/cogs.12231>
- Craw, I., & Cameron, P. (1991). Parameterising images for recognition and reconstruction. *BMVC*, 91, 367–379. <https://doi.org/10.5244/c.5.52>
- Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, 12(7-8), 1053–1074. [https://doi.org/10.1016/S0893-6080\(99\)00050-7](https://doi.org/10.1016/S0893-6080(99)00050-7)
- De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*.
- Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 995–1002. <https://doi.org/10.1037/0096-1523.33.4.995>
- Hancock, P. J. B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments, and Computers*, 32(2), 327–333. <https://doi.org/10.3758/BF03207802>
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (1998). Testing principal component representations for faces. *4th neural computation and psychology workshop*, London, 9–11 April 1997, 84–97. https://doi.org/10.1007/978-1-4471-1546-5_8
- Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition*, 24(1), 26–40. <https://doi.org/10.3758/BF03197270>
- Hancock, P. J. B., Somai, R. S., & Mileva, V. R. (2020). Convolutional neural net face recognition works in non-human-like ways. *Royal Society Open Science*, 7(10), 1–7. <https://doi.org/10.1098/rsos.200595>
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323. <https://doi.org/10.1016/j.cognition.2011.08.001>
- Johnston, R. J., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory (Hove, England)*, 17(5), 577–596. <https://doi.org/10.1080/09658210902976969>
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6), 1187–1211. <https://doi.org/10.1152/jn.1987.58.6.1187>
- Kämäräinen, J., Kyrki, V., & Kälviäinen, H. (2002a). Fundamental frequency Gabor filters. *Proceedings of ICPR2002 16th international conference on pattern recognition*, 1, 628–631.
- Kämäräinen, J., Kyrki, V., & Kälviäinen, H. (2002b). *Gabor features for invariant object recognition*. Lappeenranta University of Technology.
- Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), 103–108. <https://doi.org/10.1109/34.41390>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., Malsburg, C. v. d., Würtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3), 300–311. <https://doi.org/10.1109/12.210173>
- MATLAB. (2019). *MATLAB r2019a*. The Mathworks Inc.
- Meytlis, M. (2011). A model of face space. *Visual Cognition*, 19(1), 13–26. <https://doi.org/10.1080/13506285.2010.512507>
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*.

- O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory and Cognition*, 26(1), 146–160. <https://doi.org/10.3758/BF03211378>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, 26(3), 179–195. <https://doi.org/10.1080/13506285.2017.1405134>
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88(1), 455–463. <https://doi.org/10.1152/jn.2002.88.1.455>
- Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, 70(5), 897–905. <https://doi.org/10.1080/17470218.2015.1136656>
- Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2019). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *CORTEX*, 122, 108–114. <https://doi.org/10.1016/j.cortex.2018.12.017>
- Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, 21(5), 42–50. <https://doi.org/10.1109/38.946630>
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86. <https://doi.org/10.1162/jocn.1991.3.1.71>
- Wiskott, L., Fellous, J. M., Krüger, N., & Von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 775–779. https://doi.org/10.1007/3-540-63460-6_150
- Yue, X., Biederman, I., Mangini, M. C., Malsburg, C. v. d., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, 55, 41–46. <https://doi.org/10.1016/j.visres.2011.12.012>