

# A neurocomputational account of the face configural effect

**Xiaokun Xu**

Brain and Cognitive Science Program,  
Department of Psychology, University of Southern  
California, Los Angeles, CA, USA



**Irving Biederman**

Brain and Cognitive Science Program,  
Department of Psychology, University of Southern  
California, Los Angeles, CA, USA  
Neuroscience Program, University of Southern California,  
Los Angeles, CA, USA



**Manan P. Shah**

Neuroscience Program, University of Southern California,  
Los Angeles, CA, USA



A striking phenomenon in face perception is the configural effect in which a difference in a single part appears more distinct in the context of a face than it does by itself. The face context would be expected to increase search complexity, rendering discrimination more—not less—difficult. Remarkably, there has never been a biologically plausible explanation of this fundamental signature of face recognition. We show that the configural effect can be simply derived from a model composed of overlapping receptive fields (RFs) characteristic of early cortical simple-cell tuning but also present in face-selective areas. Because of the overlap in RFs, the difference in a single part is not only represented in the RFs centered on it but also propagated to larger RFs centered on distant parts of the face. Dissimilarity values computed from the model between pairs of faces and pairs of face parts closely matched the recognition accuracy of human observers who had learned a set of faces composed of composite parts and were tested on wholes (Which is Larry?) and parts (Which is Larry's nose?). When stimuli were high versus low passed the contributions of different spatial frequency (SF) bands to the configural effect were largely comparable. Therefore, it was the larger RFs rather than the low SFs that accounted for most of the configural effect. The representation explains why, relative to objects, face recognition is so adversely affected by inversion and contrast reversal and why distinctions between similar faces are ineffable.

## Introduction

The paradigmatic experiment documenting the face configural effect was one in which subjects learned the names of composite faces in which different shaped eyes, noses, and mouths could be swapped within an identical face context to produce pairs of faces differing in only a single part (Tanaka & Farah, 1993). The same operations were performed with houses and house parts (e.g., the door). In the recognition test, the subjects had to distinguish, Larry's face, say, from a composite foil that they had never seen before which differed by a single face part, the nose, for example. On other trials, they had to distinguish that single part, Larry's nose, from a nose that wasn't Larry's. Identification accuracy was higher for the composite faces, compared with the parts in isolation, although in both cases the target differed from the foil by the identical single part. Figure 1 shows three examples from the stimuli in the present experiment. In the Tanaka and Farah study, the advantage of the context with faces was not observed with houses and in a subsequent study by Farah (1995) the configural advantage with faces was not observed with a prosopagnosic.

An account of this phenomenon can be derived from an assumption that the representation of faces (but not nonfaces) retains aspects of the original simple-cell tuning characteristic of early cortical visual areas with the image coded by columns of cells tuned to multiple scales and orientations distributed across the visual field (Biederman & Kalocsai, 1997; Yue, Tjan, & Biederman, 2006). There would be a considerable

Citation: Xu, X., Biederman, I., & Shah, M. P. (2014). A neurocomputational account of the face configural effect. *Journal of Vision*, 14(8):9, 1–9, <http://www.journalofvision.org/content/14/8/9>, doi:10.1167/14.8.9.

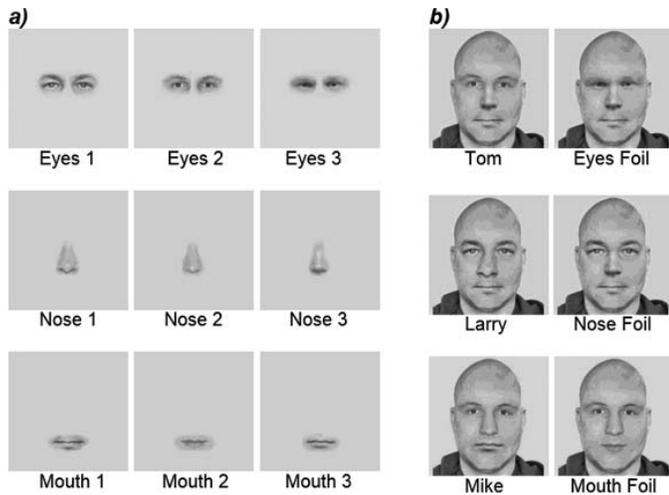


Figure 1. (a) Face parts and (b) composite target faces created from these parts for the current replication of the part-whole identification experiment of Tanaka and Farah (1993). Notice that pairs of composite faces differ only in a single face part, the eyes (top pair), nose (middle pair), and mouth (bottom pair) yet the differences appear greater than the individual parts.

degree of overlap of those cells with medium and large receptive fields (RFs) (Figure 2). These cells are thus responsive to image variation of a face, whether such variation arises from changes in part shape or part distances. Any one RF of such a cell would be activated by variation over a large area of the face and any one area of the face would be coded by cells with RFs centered at varied positions on the face. Therefore, the pattern arising from the combination of the local features, such as those from the eyes or nose, and the contextual face background create additional visual

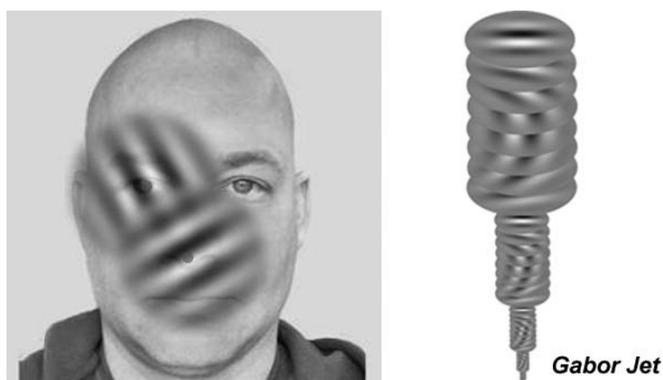


Figure 2. Left. Illustration of overlap of medium-sized receptive fields of two kernels centered on different parts of the face. The largest receptive field covers much of the face. Note that the activation of these kernels would be affected by many of the same face regions and variation in the shape of the same face parts. Right. Illustration of a Gabor “jet” (from Lades et al., 1993) with five scales and eight orientations. A jet models aspects of the tuning of a V1 hypercolumn.

features, which would be picked up by the RFs covering those areas, especially those with larger receptor fields.

Would a model of simple-cell tuning predict the advantage of a whole-face context with the larger receptive fields (and lower spatial frequencies) accounting for a greater proportion of the variance? That it would be simple cells—and not complex cells—is suggested by the great sensitivity of face recognition—but not object recognition—to the direction of contrast (e.g., Nederhouser, Yue, Mangini, & Biederman, 2007). Whereas larger RF sizes have lower spatial frequencies in V1, the linking of RF size and SF may be less true in face-selective areas. A second experiment tested whether the magnitude of the whole-face advantage would be witnessed to equal extents with high-passed compared to low-passed stimuli.

## Experiments

### Experiment 1: Does RF size account for the face configural effect?

To test whether the part-whole configural effect reported by Tanaka and Farah (1993) could be explained by a spatial representation of faces we employed the Gabor-jet model proposed by Lades et al. (1993). This is a model based on the multiscale multiorientation tuning characteristic of V1 hypercolumns, with different columns having receptive fields (RFs) centered on different parts of the visual field. Image dissimilarities based on the Euclidean distance of Gabor-jet activation values have been shown to be an excellent predictor of human psychophysical discriminability of metrically varying complex shapes, such as faces or complex blobs. Yue, Biederman, Mangini, von der Malsburg, and Amir (2012) reported essentially a perfect correlation between the Gabor-jet distances of two faces (or blobs) and subjects' discrimination accuracy in a match-to-sample task. In that experiment, subjects viewed three faces in a triangular display, with one face on top (the sample) and two test faces below. One of the two test faces was identical to the sample; the other was a foil that differed from the matching stimulus. It was the dissimilarity—computed with the Gabor jet model—of the distractor to the matching stimulus that predicted performance so well.

In fMRI adaptation experiments, the adaptation magnitudes of the BOLD signal in the fusiform face area (FFA) have also been shown to be proportional to the Gabor-jet similarity of two face images shown in succession (Xu & Biederman, 2010; Xu et al., 2009).

A Gabor-jet roughly mimics the multiscale multi-orientation, tuning properties of the cells in a V1

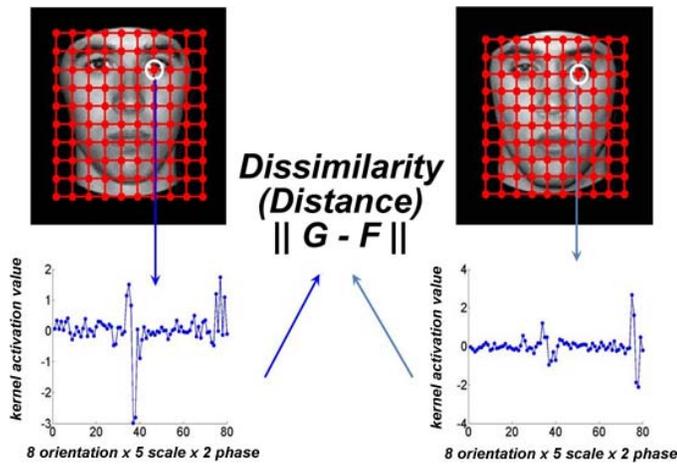


Figure 3. Illustration of the computation of dissimilarity for a corresponding pair of jets for a pair of face images (adapted from Yue et al., 2012). The Euclidean of the difference in the activation magnitudes, taken kernel by kernel within corresponding jets, summed over all 100 jets, provides a measure of dissimilarity.

hypercolumn. In the present implementation, each jet was composed of Gabor kernels at five scales and eight orientations and two phases (sine and cosine) comprising 80 filters, all with their receptive fields (RFs) centered at a common position in the visual field. The individual cells are modeled as Gabor filters that closely approximate the tuning profiles of V1 simple cells (De Valois & De Valois, 1988). Each Gabor cell (or kernel or filter or wavelet) within a jet is produced by the product of a sinusoidal grating at one of the five spatial frequencies (SFs) and one of the eight orientations and one of the two phases, and a Gaussian kernel envelope with an envelope that is a  $2\pi$  multiple of the corresponding sinusoid wavelength.

Given a pair of input images, their dissimilarity could be measured as the Euclidean distance between the vectors of the Gabor-jet coefficients of each image. Specifically, each pair of stimuli (faces or face parts) was filtered by a  $10 \times 10$  grid of jets (Figure 3). Each jet was composed of 80 Gabor kernels (each a convolution of a sinusoid and a 2-D Gaussian envelope) of eight equally spaced orientations (i.e.,  $22.5^\circ$  differences in angle) and five spatial frequencies (i.e., ranging from 8 to 32 cycles/face with half octave span), and two phases with  $90^\circ$  shift (sine and cosine), each centered on their jet's grid point. The coefficients of the kernels (the magnitudes and phases corresponding to an activation value for a V1 neuron) within each jet were then concatenated to an 8,000-element ( $100 \text{ Jets} \times 40 \text{ Kernels} \times 2 \text{ Phases}$ ) vector  $G$ :  $[g_1, g_2, \dots, g_{8000}]$ . For any pair of pictures with corresponding jet coefficient vectors  $G$  and  $F$ , the dissimilarity of the pairs is defined as the Euclidean Distance between the two vectors

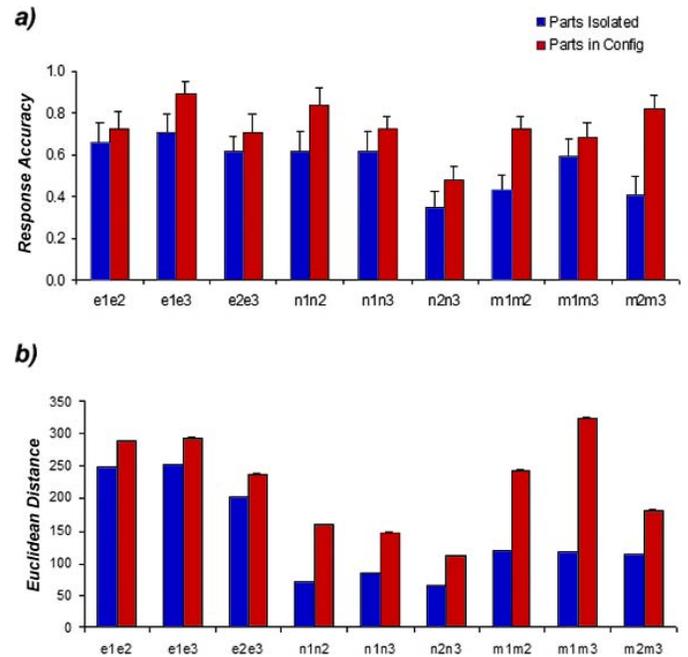


Figure 4. Behavioral results and stimulus similarity analysis in Experiment 1. (a) Accuracy in identification of all tested combinations (target and foil) of face features as a function of whether the features were shown isolated or in a face context. Each pair of bars indicates the specific exemplar pairing for eyes, nose and mouth, e.g., n1n3 indicates the discrimination of nose 1 and nose 3, with or without a contextual face background, respectively, in each bar. For every combination of face features, accuracy was higher for composite faces differing in a part than the isolated parts. (b) The image dissimilarity in Gabor Euclidean distance metric for the nine feature pairings and whether the features were shown isolated or in context. Note that the Euclidean distances reflect not only the greater dissimilarity of the faces in context but, to some extent, the difficulty of the individual feature combinations.

$$Dist(G, F) = \sqrt{\sum_{i=1}^{8000} (g_i - f_i)^2},$$

$$i = 1, 2 \dots 8000, \quad (1)$$

as illustrated in Figure 3.

## Method

**Participants:** Eleven students from University of Southern California (mean age = 24.3 years old, three female) participated in the experiment. All subjects reported normal or corrected-to-normal vision and normal face recognition ability.

**Stimuli and design:** The design followed that in the original Tanaka and Farah (1993) experiment. Three exemplars of each of three face features, eyes, and mouth (Figure 1), were chosen to create 27 composite

faces, with the face features maintained in the same spatial locations and embedded in a common head contour background, using the Morphases Editor (Morphases, Kajaani, Finland). (The original Tanaka & Farah, 1993, stimuli were no longer available, J. Tanaka, personal communication, November 9, 2012.) Six among the 27 faces were each given a name (Figure 1 shows three of them) while the others served as foils. The association between each of the six target faces and its name were acquired during a self-paced learning session, during which each face–name pair was presented six times in randomized order. Two of the six target faces shared one feature and differed in the other two, e.g., Larry and Bob had identical eyes, Mike and Derek identical mouths, such that no one part exemplar was unique to a particular face.

In each trial of the test session, subjects were presented with either a pair of isolated face features or a pair of composite faces, and prompted to perform an identification task. For example, given a pair of composite faces, say Larry’s face and a foil that differed from Larry (who had nose1) by only one feature (e.g., the foil had nose3), subjects pressed either the left or right arrow key to indicate which one was Larry. Similarly, when presented with nose1 and nose3 in isolation, subjects were asked to identify which one depicted Larry’s nose. Importantly, the foils were chosen outside of the six learned identities and were not seen during the learning session.

Subjects observed the presentation of the stimuli in grayscale on a CRT computer monitor, from a distance of approximately 57 cm. Each image subtended a visual angle of approximately  $6^\circ$  and was centered at  $4^\circ$  eccentricity, left and right, from central fixation. Subjects had to respond within 5 s from the onset of the target images.

### Results and discussion

On average, subjects’ identification accuracy was markedly worse for the isolated parts, 55%, compared to the 73% accuracy for identifying face composites,  $t(10) = 4.2$ ,  $p = 0.002$ , even though the face backgrounds were identical for the target and foil. Figure 4a shows identification performance for each of the three features (eyes, nose, and mouth), and for each target–foil combination of the exemplars, such as eyes1 versus eyes2 (indicated by e1e2), with the feature either in isolation or in the identical face context, respectively. All nine combinations showed an advantage of identification of the composite faces than face parts in isolation, confirmed by a paired  $t$  test:  $t(8) = 4.7$ ,  $p < 0.002$ . This perceptual effect was reflected in the image-based similarity analysis (Figure 4b). The Euclidean distance between the Gabor feature vectors of the

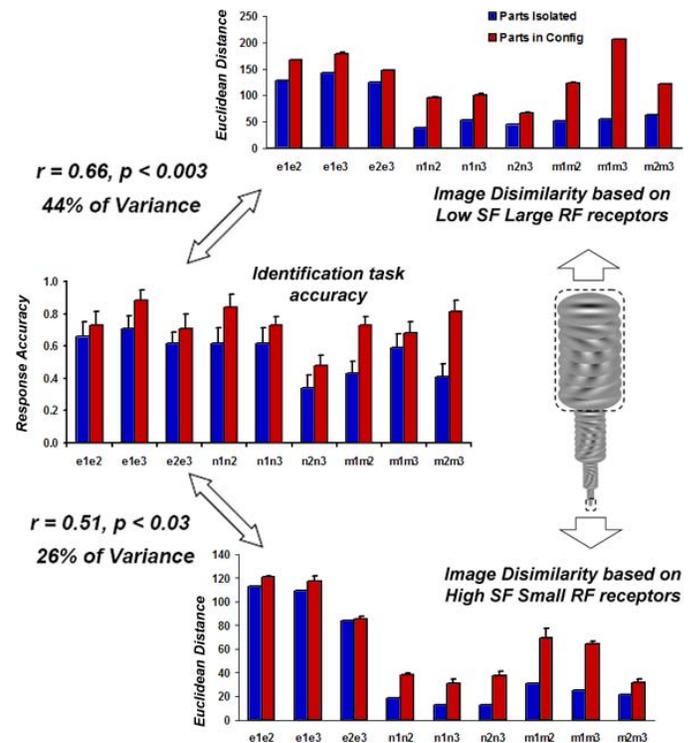


Figure 5. Predicting response accuracy from large RFs (low SF) versus small RFs (high SF) components in the Gabor-jet representation of faces. A greater proportion of the variance is predictable from the large RF (low SF) components.

target and foil was also larger for the part in face context than part in isolation for every feature and exemplar,  $t(8) = 4.3$ ,  $p < 0.003$ .

Because the composite faces were made by placing different feature exemplars in the same spatial configuration, a pixel-intensity based representation would yield exactly the same distances between parts and between whole faces. Why did the Euclidean of the Gabor representation yield the advantage of the composite over the isolated parts? We suggest that the advantage of the composite arises from the overlap of the receptive fields of face neurons, as illustrated in Figure 2. The interaction between local features such as those arising from the eyes and nose and the contextual face background created additional visual features, which were picked up by the kernels covering those areas, especially those with larger receptor fields. To test this hypothesis, we assessed the proportion of the variance of the whole face advantage that was predictable from the largest RFs (with the lowest SF of 8 cycles/face) versus the smallest RFs (with the highest SF of 32 c/f) components of the Gabor features of each face, separately. The results are shown in Figure 5. The mean distances between the isolated parts were significantly lower than that between composite faces in both components, both  $t(8) > 4$ ,  $p < 0.005$ . However,

the mean difference in the distances between the parts and wholes (e.g., distance between isolated Mouth 1 and Mouth 2, minus the distance between composite faces with Mouth 1 and Mouth 2) was 19 for the high SF and small RFs components and 56 for low SF and large RFs components, a sizable differences in the distances, confirmed as reliable by a post-hoc  $t$  test:  $t(8) = 3.6$ ,  $p < 0.01$ . The identification accuracy was better correlated with the distance measurement in the large RF band (low SF) ( $r = 0.66$ ,  $p < 0.003$ ) than in the small RF (high SF) band ( $r = 0.51$ ,  $p < 0.03$ ). This result suggests that the configural effect was mediated to a larger extent by the neural encoding of larger receptor fields or lower spatial, or a combination of the two.

## Experiment 2: Is the configural effect a function of RF size or spatial frequency or both?

Although a V1 type of representation necessarily links RFs and SF with larger RFs associated with lower SFs, subsequent face (and object) selective areas need not manifest this linkage. In fact, the general observation is that later ventral pathway areas are associated with large RFs composed of all SF (e.g., Kobatake & Tanaka, 1994). Our analyses of the configural effect in Experiment 1 showed that it was primarily accounted for by large RFs coupled with low SFs rather than small RFs coupled with high SFs. By varying SF independent of RFs, Experiment 2 was designed to investigate the extent to which the configural effects are a consequence of larger RFs, independent of SF.

### Method

**Participants:** Fifteen students from the University of Southern California (mean age = 23.3 years,  $\pm 3.8$  SD, eight female). All subjects reported normal or corrected-to-normal vision and normal face recognition ability.

**Stimuli:** The same set of stimuli in Experiment 1 was filtered by two 2-D Gaussian filters. The cutoff frequency was set at 8 cycles per face (cpf) for the low-pass filter and 32 cpf for the high-pass filter. The two filters therefore had negligible overlap in the frequency domain, as shown in Figure 6a. The original images went through Fourier transformation into the frequency domain, multiplied by the corresponding filter, respectively, and reverse transformed back into the image domain. Finally, the pixel intensities of each filtered image were standardized to match the mean luminance and RMS contrast of the corresponding original image, as shown in Figure 6b. The experimental procedure was identical to that in Experiment

1: Subjects learned six individual faces and performed the two alternative forced choice (2AFC) task given a pair of composite face, or isolated face parts, high passed and low passed, or all-passed (original) in three runs. The order of trial conditions in each run was counterbalanced in each run to eliminate the potential carry over effect of different conditions.

### Results

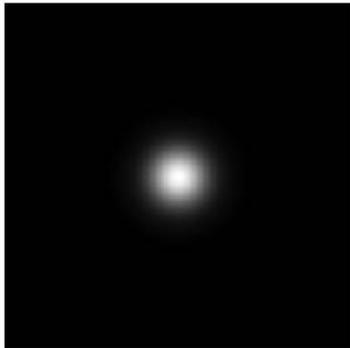
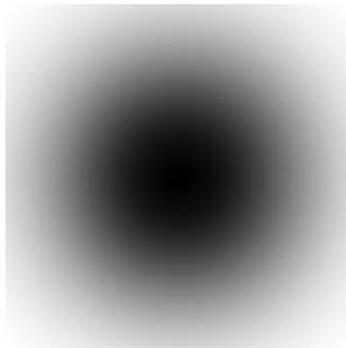
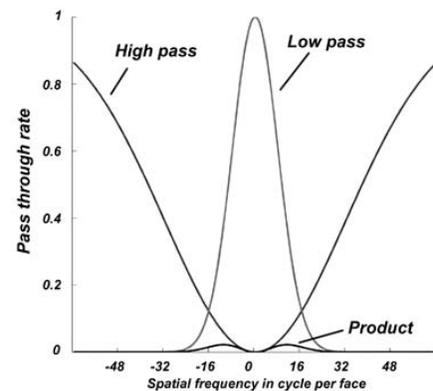
The configural effect, defined as the advantage in identifying the composite faces against isolated face parts, was evident in all filtering conditions, as shown in Figure 7. The mean accuracy for the all-pass, high-pass, and low-pass filtering conditions was 84.0%, 75.9%, and 79.2%, respectively, for the composite face, and 60.7%, 60.5%, and 58.7% for the isolated parts, respectively, for the three filtering conditions. A repeated measures  $3 \times 2$  analysis of variance (ANOVA) of the filtered images (All Pass vs. High Pass vs. Low Pass)  $\times$  Stimulus Type (Composite vs. Isolated Part) revealed a significant main effect of stimulus type,  $F(1, 14) = 84.6$ ,  $p < 10^{-7}$ , and to a markedly lesser extent, spatial frequency filtering,  $F(2, 28) = 5.5$ ,  $p < 0.01$ . However, the interaction between stimulus type and spatial frequency fell short of significance,  $F(2, 28) = 2.5$ ,  $p = 0.098$ , suggesting that the configural effect was not strongly modulated by spatial frequency.

### Discussion

The results of Experiment 1 show that the configural effect—the advantage in recognition of composite faces differing only in an individual part over the recognition of the isolated parts—could be better accounted for by the image component of large RFs coupled with low SF, compared with small RFs coupled with high SFs. To further tease apart the contribution of SF and RFs to the configural effect, we manipulated SF through high-pass and low pass filtering independent of RF size. The results showed that the advantage in identifying the composite whole face over isolated face parts was largely independent of spatial frequency. Therefore, the configural effect is to a large extent a function of receptor field size rather than spatial frequency band.

Goffaux, Hault, Michel, Vuong, and Rossion (2005) tested subjects in a 2AFC match-to-sample task while using spatial filtering procedures similar to those in Experiment 2. They reported that the configural effect relied more on the low-spatial frequency components. However, in their report the configural effect was qualitatively defined as the difference between “featural” and “configural” processing, where the target and foil differed in the shape of

a)

**LowPass with 8 cpf cutoff****HighPass with 32 cpf cutoff****Minimal overlap between the two**

b)

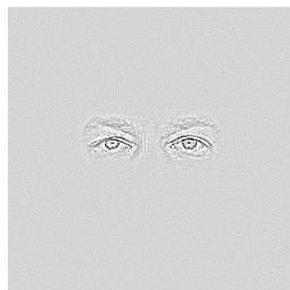
**Original Image****HighPassed Image****Low Passed Image**

Figure 6. Spatial filtering of the part and whole face stimuli in Experiment 2. (a) The low-pass and high-pass 2-D Gaussian filters in the frequency domain, with a cutoff threshold of 8 cpf (cycles per face) and 32 cpf, respectively. The 1-D silhouettes of the high-pass and low-pass filter, and their point-wise product are shown in the rightmost plot, showing minimal overlap between the two frequency channels. The negative spatial frequencies arise from the convention of the discrete Fourier transform. (b) Examples of a face part and a whole composite face before and after spatial filtering.

individual features or the distance between features, respectively. Importantly, the face images were further smoothed after the spatial filtering rendering the difference between target and foil extremely subtle when the differential cue was “featural,” compared to the “configural” condition. In our Experiment 2, the contrast of identifying learned individuals based on

parts or the whole face provided a more direct test of the configural effect. With the results from Experiments 1 and 2 taken together, we conclude that the configural effect is largely a function of the overlap in the encoding of multiple face features allowed with large RFs, rather than the information carried in low spatial frequency components.

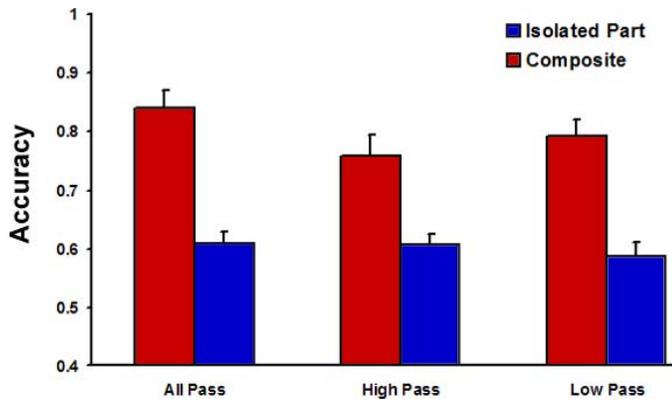


Figure 7. The configural effect (isolated part vs. composite) as a function of spatial frequency (all SF, high, and low pass).

## General discussion

We replicated the part-whole configural effect reported by Tanaka and Farah (1993): once in Experiment 1 and twice in Experiment 2 (with high- and low-passed images). In all cases, subjects identified face features better when they were placed in the context of a whole face than when in isolation. Critically, the contextual face background was identical for the target and foil composite faces and thus not informative by itself for identification. The contextual benefit thus had to arise from the configural (or “holistic”) processing of all face features; more specifically, it had to be attributed to the interaction of the parts with the whole face. To the best of our knowledge, no previous research has proposed a computational account of this phenomenon. This advantage of the whole over the parts could be derived from the image similarity analysis using a Euclidean distance metric on the Gabor feature output. The Euclidean distances between isolated face features were smaller than the distances between the whole face composites where each feature was embedded in the identical configural context. Therefore, it would be expected, assuming the Euclidean distances were relevant to perceptual discrimination, that the isolated features would be more difficult to discriminate given their proximity in Gabor feature space relative to the composite whole faces. Although we have employed Gabor wavelets—which are justified both by V1 tuning profiles (De Valois & De Valois, 1988) and optimal computational efficiency (Daugman, 1980), other kernels of varying scale and orientation could also predict an advantage of whole faces over the isolated parts.

An assumption underlying our explanation of the configural effect is that the individuation of faces (but not objects) retains aspects of the original spatial representation, with allowance for scale and translation

invariance. Evidence supporting this assumption was reported by Yue, Tjan, and Biederman (2006) who showed that the same–different matching accuracy (same person?) and fMRI release of adaptation (in FFA) of sequences of pairs of faces were sensitive to the particular combinations of eight scales and eight orientations comprising the spatial content of those faces. Nonface blobs resembling teeth, with differences between blobs (same blob?) scaled to be equal to the differences between the faces, did not show this sensitivity in discrimination performance or adaptation (in LOC) to the particular combinations of scales and orientations.

That the representation of faces retains aspects of the original spatial filtering does not mean that such a representation would not manifest position, scale, and saccade invariance. For example, although activation of V1 shifts with saccades, the perception of faces (and objects) remains stable. Similarly, we can readily recognize faces and objects with little or no cost when they are translated or varied in size. Indeed, the receptive fields assumed in the current study are defined in terms of cpf (cycles per face) rather than the cpd (cycles per degree) with which V1 contrast sensitivity thresholds are typically expressed. That the configural effects documented in the current experiments are manifestations of a face system rather than an earlier stage is supported by the finding of Tanaka and Sengco (1997) who failed to find configural effects with houses or inverted faces and reported larger configural effects (better recognition memory for face parts) in familiar rather than unfamiliar face contexts. Faces are likely the only class of stimuli in our evolutionary past where fine metric differences of complex shapes had to be individualized. Retention of the original spatial coding provides the sensitivity to achieve such discriminations. Given our extensive visual experience with upright intact faces lit from above during development, the prototype representation of faces (e.g., the norm-based face representation proposed by Leopold, Bondar, & Giese, 2006) could also be constructed with the Gabor representation that conforms to those regularities.

Because the matching of the Gabor activation values is done in a 2-D coordinate space (Mangini & Biederman, 2004; Yue et al., 2011), planar inversion is thus particularly disruptive to the individuation of faces (e.g., Yin, 1969). Because the Gabor activation values define the surface of the face there is, similarly, a marked cost to recognition if the direction of contrast is reversed (and a lesser cost if the direction of illumination is changed) as when matching a positive to a negative image of a face (e.g., Nederhouser, Yue, Mangini, & Biederman, 2007).

In contrast to faces, the individuation of objects is largely based on parts defined by edges at orientation and depth discontinuities so their recognition is

invariant to direction of contrast and direction of lighting (e.g., Nederhouser et al., 2007; Russell et al., 2007; Vogels & Biederman, 2002). These edges define a structural description specifying the nonaccidental shape properties of the object's parts and their categorical relations (Biederman, 1987; Lescroart & Biederman, 2012) rendering the recognition of objects much less sensitive to 2-D inversion, contrast reversal, and direction of illumination than faces. Faces undergo the same parts-like processing as objects, so we know that it is a face, but such processing does not yield the detailed metrics required to individuate similar faces (Biederman & Kalocsai, 1997).

The current account of face configuration effects as arising from the overlap of the receptive fields of larger spatial kernels challenges the characterization of 2-D inversion costs with faces as a “configural effect” (e.g., Freire, Lee, & Symons, 2000). As described above, inversion produces mismatches in the (Gabor) kernel activation values in the 2-D coordinate space between a stored representation of a face and an inverted probe without any requirement to posit interactions at a distance. Our model's account for the inversion cost in face recognition therefore did not rely on, and was not restricted by the explicit encoding of distance between face parts, as proposed by Freire et al., 2000.

By the present account, the essential matching operations in face individuation are performed on the activation values of the spatial kernels. We do not have cognitive access to these values so while we can often accurately distinguish similar faces, the basis of our discrimination typically remains ineffable.

*Keywords:* face configural effect, holistic processing, Gabor jet, high- versus low-pass filtering of faces

## Acknowledgments

We are indebted to Bosco Tjan for his expert critical analysis and conceptualization of the filtering operations in Experiment 2. Supported by NSF BCS 0617699 and the Dornsife Research Fund.

Commercial relationships: none.

Corresponding author: Irving Biederman.

Email: [bieder@usc.edu](mailto:bieder@usc.edu).

Address: Program in Neuroscience and Department of Psychology, University of Southern California, Los Angeles, CA, USA.

## References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences*, *352*, 1203–1219.
- Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, *20*, 847–856.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. Oxford, UK: Oxford University Press.
- Farah, M.J. (1995). Is face recognition “special”? Evidence from neuropsychology. *Behavioural Brain Research*, *76*, 181–189.
- Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 1159–1170.
- Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77–86.
- Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.
- Lades, J. C. V., Buhmann, J., Lange, J., Malsburg, C., Wurtz, R., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers: Institution of Electrical and Electronics Engineers*, *42*, 300–311.
- Leopold, D., Bondar, I., & Giese, M. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*, 572–575.
- Lescroart, M. D., & Biederman, I. (2012). Cortical representation of medial axis structure. *Cerebral Cortex*, *23*, 623–637.
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classification. *Cognitive Science*, *28*, 209–226.
- Nederhouser, M., Yue, X., Mangini, M. C., & Biederman, I. (2007). The deleterious effect of contrast reversal on recognition is unique to faces, not objects. *Vision Research*, *47*, 2134–2142.
- Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2007). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, *35*, 749–759.

Biederman, I. (1987). Recognition-by-components: A

- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A*, *46*, 225–245.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*, 583–592.
- Vogels, R., & Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. *Cerebral Cortex*, *12*, 756–766.
- Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity and viewpoint. *Journal of Vision*, *10*(14):36, 1–13, <http://www.journalofvision.org/content/10/14/36>, doi:10.1167/10.14.36. [PubMed] [Article]
- Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): image or person. *Vision Research*, *49*, 2800–2807.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.
- Yue, X., Biederman, I., Mangini, M. C., von der Malsburg, C., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, *55*, 41–46.
- Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2011). Lower-level stimulus features strongly influence responses in the fusiform face area. *Cerebral Cortex*, *21*, 35–47.
- Yue, X., Tjan, B., & Biederman, I. (2006). What makes faces special? *Vision Research*, *46*, 3802–3811.