Differences of Face and Object Recognition in Utilizing Early Visual Information

Peter Kalocsai and Irving Biederman

Department of Psychology and Computer Science
University of Southern California
Los Angeles, California 90089, U.S.A.
{kalocsai, ib}@selforg.usc.edu

Abstract. The first cortical stage in both object and face recognition by humans is generally presumed to be a filtering of the image by cells which can be approximated as oriented spatial frequency kernels. Are the outputs of these filters mapped in the same manner to the separate patches of tissue in extrastriate cortex presumed to code faces and objects? Complementary images of objects and faces were produced by dividing the Fourier spectrum of each image into 8 frequency bands and 8 orientation bands. In the inverse Fourier transform, half the 8 x 8 values (analogous to all the red squares of a checkerboard in a row by column representation of frequency and orientation) were contained in one member of a complementary image pair and the remaining combinations of values (e.g., black squares) were contained in the other member. In a naming task original and complementary images produced equivalent priming (equal RTs and error rates) for objects, but name verification for famous faces showed less priming for the complementary image. One possible explanation for these results is that faces are represented as a direct mapping of the outputs of early filter values whereas objects are recognized by means of intermediate primitives (e.g., parts), in which the same primitives can be activated by many patterns of filter activations. Two additional experiments using nonface but highly similar shaped objects (chairs) and unfamiliar faces confirmed the above hypothesis.

Keywords. Face recognition, object recognition, complementary pairs, Fourier filtering

1 Introduction

Four experiments will be described which were designed to assess whether the identification or matching of faces and objects would be directly dependent on the early spatial filter representation (V1) of the visual system. There is considerable evidence in the literature that the priming of objects cannot be dependent on a representation that retained the similarity space of the activation values of spatial filters (Fiser, Biederman, & Cooper, 1997). For example, if contour is deleted from a line drawing of an object so that the geons cannot be recovered from the image, recognition becomes impossible (Biederman, 1987). The same amount of contour deletion, when it permits recovery of the geons, allows ready recognition. Fiser et al showed that the Lades et al. (1993) model recognized the two kinds of

stimuli equally well. Similarly, the Lades et al. (1993) model failed to capture the differences in matching objects that did or did not differ in a NAP in the Cooper and Biederman (1993) experiment.

Biederman and Cooper (1991) showed that members of a complementary pair of object images in which every other line and vertex was deleted from each part (so that each image had 50% of the original contour) primed each other as well as they primed themselves. The measure of priming was the reduction in the naming reaction times and error rates from the first to the second brief exposure of an object picture. The priming was visual, and not just verbal or conceptual, because there was much less priming to an object that had the same name but a different shape (e.g., two different shaped chairs). In this case, humans treated the members of a complementary pair as equivalent although the two members would have different spatial filter activation patterns (Fiser, Biederman, & Cooper, 1997).

To test whether faces retain and objects do not retain the original spatial filter activation pattern, the first two experiments employed a similar design comparing the magnitude of priming of identical to complementary images. Rather than deletion of lines as in the Biederman and Cooper experiment, complementary pairs of gray-level images of objects and faces of celebrities were created by having every other Fourier component (8 scales X 8 orientations) in one member and the remaining 32 components in the other, as illustrated in Fig. 1[1].

2 Object naming experiment

Subjects named pictures of common objects on two blocks of trials (Exp. I). On the second block, for each object viewed on the first block, subjects would see either the identical filtered image that was shown on the first block, its spatial complement, or a different shaped exemplar with the same name, as illustrated in Fig. 2. The results of this experiment are shown in Fig. 3. Visual priming was evidenced on the second block of trials because the same shaped object was named more quickly and accurately that an image with the same name but a different shape. However, naming RTs and error rates for identical and complementary images were virtually equivalent, indicating that there was no contribution of the original Fourier components compared to their complements to the magnitude of visual priming.

---

[1] Complementary image pairs were created by the following procedure: 8-bit gray scale images were Fourier transformed and bandpassed filtered cutting off the highest (above 181 cycles/image) and lowest (below 12 cycles/image) spatial frequencies. The remaining part of the Fourier domain was divided into 64 areas (8 orientations x 8 spatial frequencies). The orientation borders of the Fourier spectrum were set up in succession of 22.5 degrees. The spatial frequency range covered 4 octaves in step of 0.5 octaves. By this operation the two complementary images had no common information about the objects in the Fourier domain.

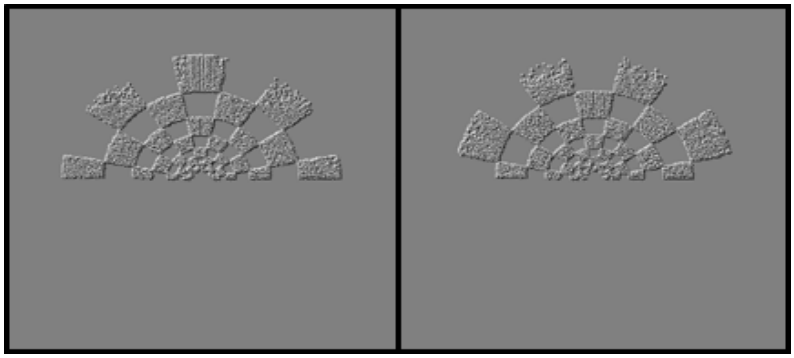

Fig. 1. Illustration of how the 8 scales X 8 orientations were distributed to the members of a complementary pair. If arranged as a checkerboard with rows the spatial frequencies and the columns the orientations, one image would have the specific scale-orientation values of the red squares, the other member the values of the black squares. Here the checkerboard is shown as two half radial grids, with scale varying with distance from the origin (low to high SF and orientation varying as shown. (The lower half would continue the upper half.)

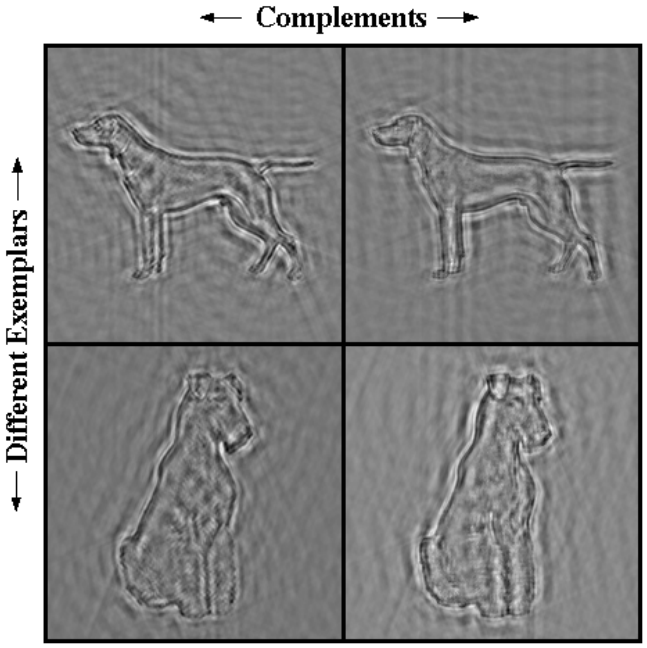

Fig. 2. Example images for the object naming task of Exp. I. Shown are the four images (2 exemplars X 2 complements) created for the entry level object "dog." In the priming paradigm one of the four images was displayed on the first block of trials and either the identical image, its complementary pair or a different exemplar image was displayed on the second block of trials.

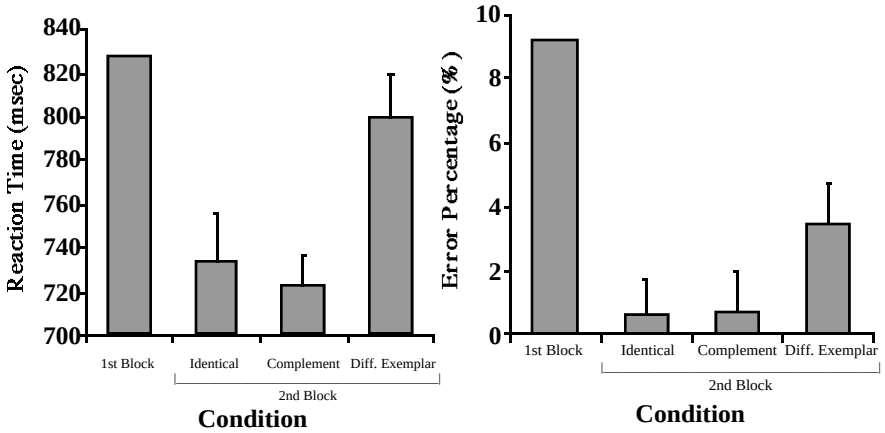

Fig. 3. Mean correct naming RTs and mean error rates for the object naming task of Exp. I. The second block data are for those trials where the object was correctly named on the first block. The second block data are for those trials where the object was correctly named on the first block.

3 Face verification experiment

Experiment II (face verification) employed the general priming design of Exp. I except now the stimuli were images of famous people and subjects verified rather than named the images. Before each trial the subject was given the name of a famous person. If the image was that person the subjects were to respond 'same'. On half the trials the picture did not correspond to the target. In these cases the picture was a face of the same general age, sex, and race as the target and the subjects were to respond 'different'. The verification task was used, rather than a naming task, because the naming of faces is slow and error prone. As in experiment I, two pictures with the same name but a different shape (differences in pose, expression, orientation, etc.), as illustrated in Fig. 4, were used to assess that the priming would be visual and not just verbal or conceptual. As in experiment I, for the 'same' trials on the second block, for each face viewed on the first block, subjects would see either the identical image, its complement or the different image of the same person as illustrated in Fig. 5. In contrast to the result for object naming, in this experiment complementary images were verified significantly more slowly and less accurately than those in the identical condition, as shown in Fig. 6. The difference between the complementary and the different exemplar faces was not significant, indicating that the visual system represented complementary face images almost as differently from the original as it did the different exemplar images. This result indicates that the representation of a face, unlike that of an object, is specific to the original filter values.

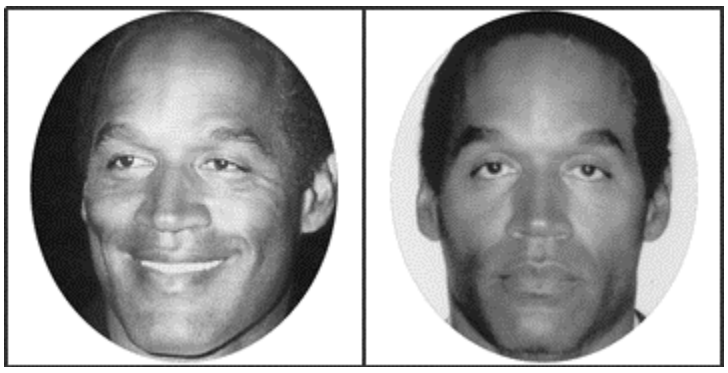

Fig. 4. Example of two original gray level images of a famous person (O. J. Simpson). illustrating differences in expression and pose used in the face verification task of Exp. II. The images were collected such that the expression and/or the orientation of the two face images of a person were different.

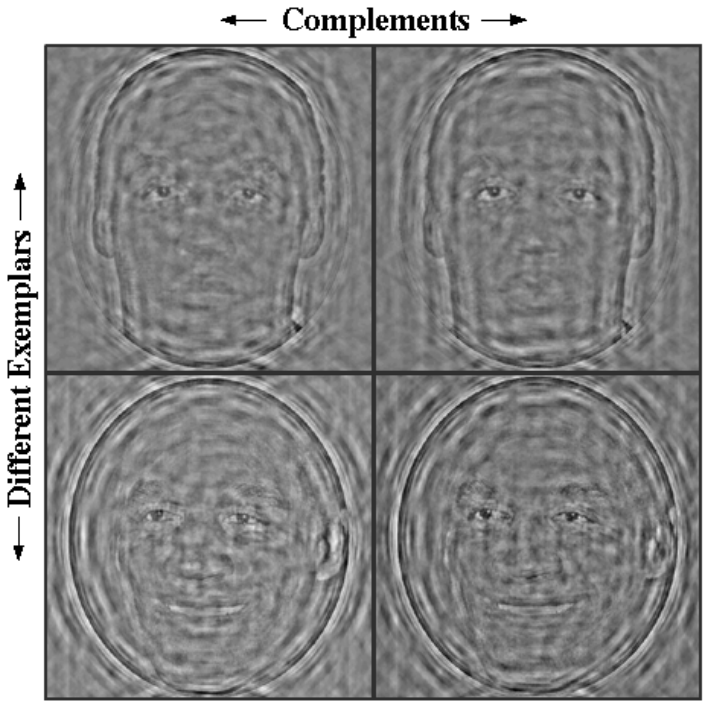

Fig. 5. Filtered complementary images for the famous face verification task of Exp. II. Shown are the four images (2 exemplars x 2 complements) created for the images of O. J. Simpson' shown in Fig. 12. In the priming paradigm one of the four images was displayed on the first block of trials and either the identical image, its complementary pair, or a different exemplar image was displayed on the second block of trials.

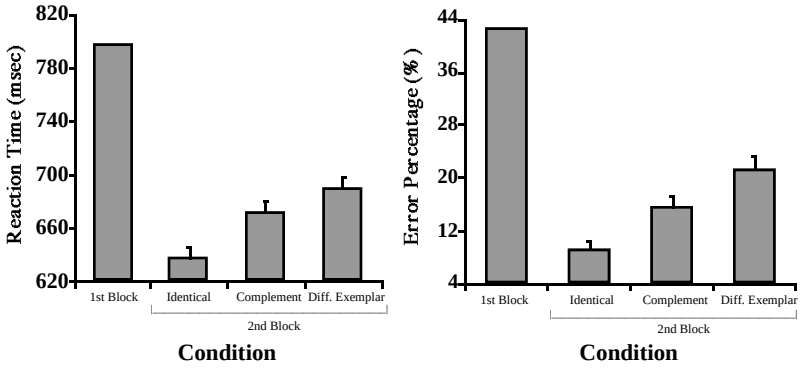

Fig. 6. Mean correct naming RTs and mean error rates for the face verification task of Exp. II. The second block data are for those trials where the object was correctly named on the first block. The second block data are for those trials where the face was correctly verified on the first block.

4 Same-different judgment of chairs and faces

One possible explanation for the above results is what we have been positing: Face representations preserve the activation pattern of early filter values, whereas object representations do not. Alternatively it could be that it is the necessity for distinguishing among highly similar entities, such as faces, that produces a dependence on the original early filter outputs. Two additional experiments were conducted to assess whether the dependence on the precise filter values were a consequence of the greater similarity of the face stimuli (or the verification task, itself) as opposed to being a phenomenon specific to the representation of faces. In these experiments, subjects viewed a sequence of two highly similar chairs (Experiment III, Figures 7 and 8 or two highly similar faces (Experiment IV) (Fig. 9). Subjects performed a same-different matching task in which they judged, 'same' or 'different,' whether the two chairs or persons were the same, ignoring whether the image was identical or complementary. The mean similarity of the complementary pairs of faces and objects were approximately equivalent as was the mean similarity of target and distractor faces and objects as assessed by the Lades et al. (1993) model[2] (Table 1). In both experiments III and IV, on half the same trials the second presented image was identical to the first and in the other half the trials it was the complementary image.

---

[2] A recent study (Subramaniam, Biederman & Kalocsai, 1997) provides strong documentation that the Lades et al. (1993) system can provide an a priori measure of shape similarity when the pairs of shapes only differ in metric properties. In a same-different sequential matching task, subjects judged whether two highly similar, blobby, asymmetric toroidal free-form shapes were identical or not. A family of 81 such shapes had been generated by Shepard and Cermack (1973). On different trials the shapes varied in similarity as assessed by the Malsburg system. For intermediate to highly similar shapes, RTs and error rates in judging that two shapes were different correlated .95 with the Malsburg similarity measure.

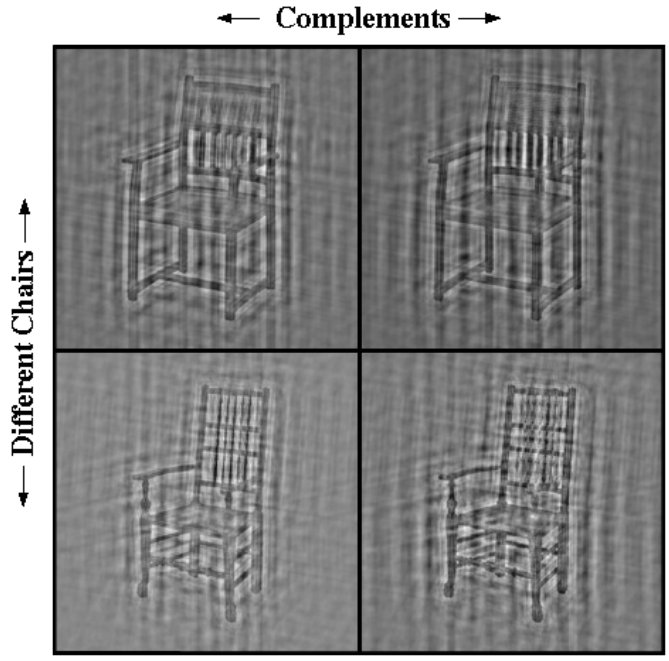

Fig. 7. Same images for the chair matching task of Exp. III. Shown are the four images (2 exemplars X 2 complements) created for two chair images from the stimuli set.

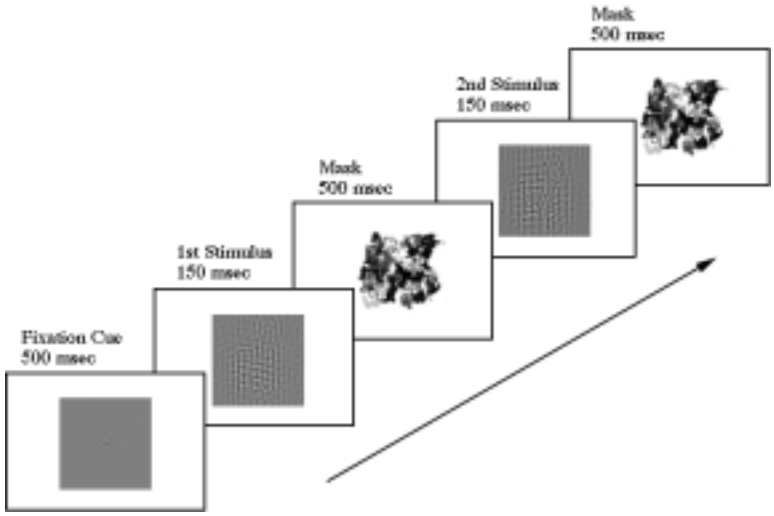

Fig. 8. Sequence of images presented in the chair matching task of Exp. III. The correct response to this sequence is 'same' because both pictures are of the same chair, though different members of a complementary pair.

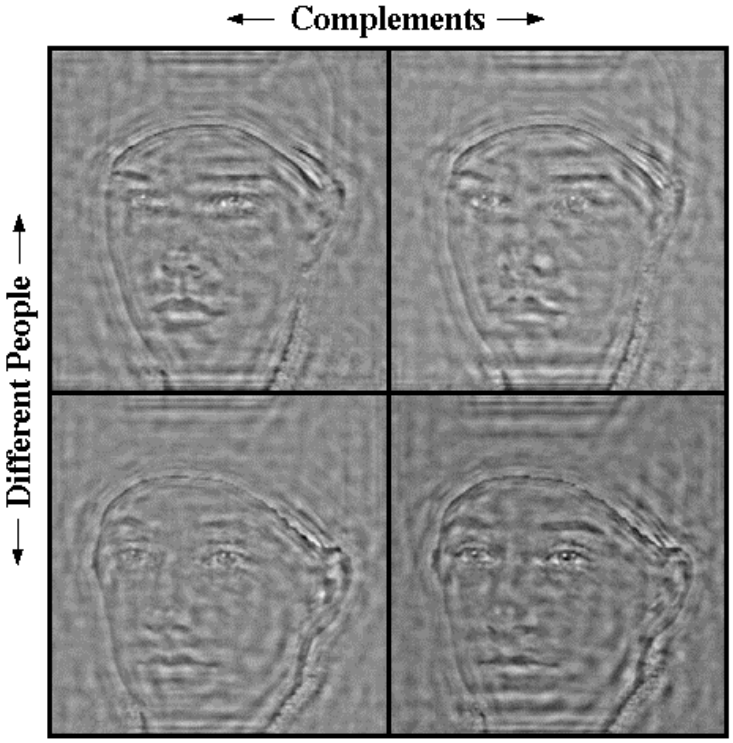

Fig. 9. Example images (from the Faces I set) for the unfamiliar face matching task of Exp. IV. Shown are the four images (2 exemplars x 2 complements) created for two face images from the stimuli set.

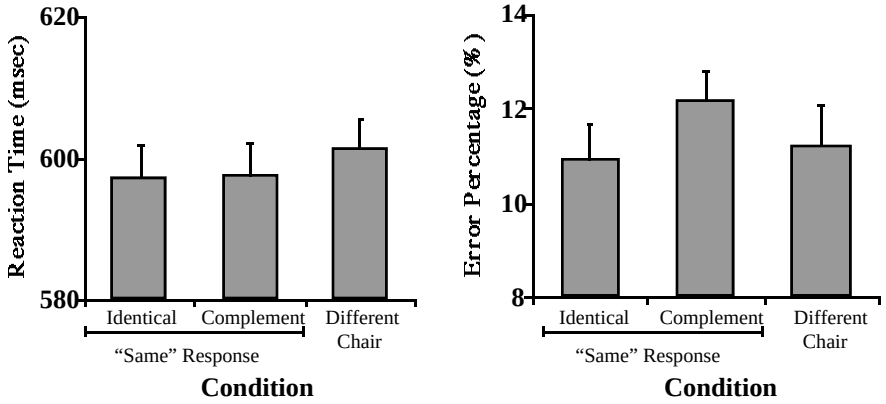

Fig. 10. Mean correct RTs and mean error rates for the chair matching task of Exp. III.

	Chairs	Faces I	Faces II
Complements	.73	.75	.78
Different ("No" trials)	.75	.76	.78

Table 1. Average similarity for stimuli in the three same-different judgment experiments (1.0 similarity would indicate perfect match).

Performance on identical and complementary chair images on same trials was virtually identical, as shown in Fig. 10, indicating that there was no effect of changing the specific spatial components of the chair images. However, for faces the complementary images were significantly more difficult to match than identical ones (Fig. 11), indicating a strong contribution of the specific spatial components in the image.

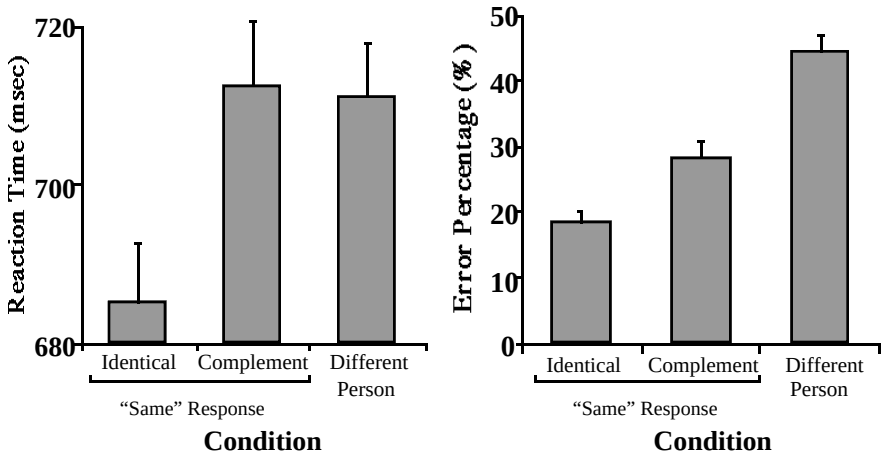

Fig. 11. Mean correct RTs and mean error rates for the unfamiliar face matching task of Exp. IV.

However, as Figures 10 and 11 shows subject made significantly more errors and their reaction time was also slower on the same-different judgment of faces than of chairs. Notice that the error percentage for Different Person trials was close to 50% which would be chance performance. This indicates that although the similarity of same and different chairs and faces was comparable to each other (Table) subjects still found the face judgment task a much more difficult one which could have altered the result. In order to test this possibility an additional experiment was run with a new set of face images (Faces II) with the purpose of making the same-different judgment of faces an easier task. As Fig. 12 indicates subject were faster and also made less error on this face judgment task compared to the previous one, but the difference between performance on the Identical and Complement conditions remained constant showing that the observed effect can not be contributed to the difficulty of the task.

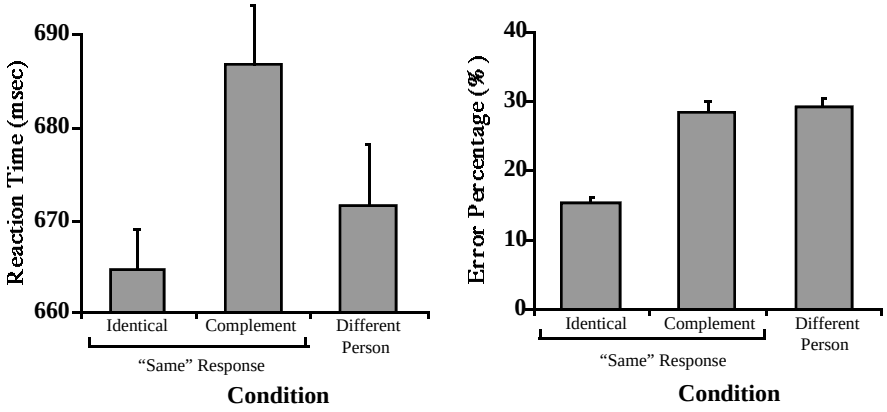

Fig. 12. Mean correct RTs and mean error rates for the unfamiliar face matching task of Exp. IV (version 2).

In summary this set of experiments showed equivalent priming and matching performance for identical and complementary images of objects. However, faces revealed a striking dependence on the original filter values. There was virtually no visual priming across members of a complementary pair of faces and face complements were far more difficult to match than identical images. These results indicate that faces are represented as a more direct mapping of the outputs of early filter values. One likely reason why the objects were unaffected by varying the filter values is that object representations employ nonaccidental characterization of parts or geons based on edges at depth or orientation discontinuities. Different spatial filter patterns can activate the same units coding edges, nonaccidental characteristics, part structures, and relations, as discussed by Hummel and Biederman (1992).

5 Conclusion

A series of experiments demonstrated that the recognition or matching of objects is largely independent of the particular spatial filter components in the image whereas the recognition or matching of a face is closely tied to these initial filter values. These results reveal crucial differences in the behavioral and neural phenomena associated with the recognition of faces and objects. Readily recognizable objects can typically be represented in terms of a geon structural description which specifies an arrangement of viewpoint invariant parts based on a nonaccidental characterization of edges at orientation and depth discontinuities. The parts and relations are determined in intermediate layers between the early array of spatially distributed filters and the object itself and they confer a degree of independence between the initial wavelet components and the representation. Individuation of faces, by contrast, requires specification of the fine metric variation in a holistic representation of a facial surface. This can be achieved by storing the pattern of activation over a set of spatially distributed filters. Such a

representation would also evidence many of the phenomena associated with face recognition such as holistic effects, unverbalizability, and great susceptibility to metric variations of the face surface, as well as to image variables such as rotation in depth or the plane, contrast reversal, and direction of lighting.

Acknowledgements This research was supported by ARO NVESD grant DAAH04-94-G-0065. Parts of this paper are excerpted from Biederman and Kalocsai (1997).

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

Biederman, I. & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393-419.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London B*, 352, 1203-1219.

Cooper, E. E., & Biederman, I. (1993). Metric versus viewpoint-invariant shape differences in visual object recognition. *Investigative Ophthalmology & Visual Science*, 34, 1080.

Fiser, J., Biederman, I., & Cooper, E. E. (1997). To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition? *Spatial Vision*, 10, 237-271.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480-517.

Lades, M., Vortbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., & Konen, W. (1993). Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers, 42,* 300-311.

Shepard, R. N., & Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-from stimuli. *Cognitive Psychology*, 4, 351-377.

Subramaniam, S., Biederman, I., & Kalocsai, P. (1997). Predicting nonsense shape similarity from a V1 similarity space. Unpublished ms., University of Southern California.