What makes faces special?

Xiaomin Yue[*], **Bosco S. Tjan**, and **Irving Biederman**
Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

Abstract

What may be special about faces, compared to non-face objects, is that their neural representation may be fundamentally spatial, e.g., Gabor-like. Subjects matched a sequence of two filtered images, each containing every other combination of spatial frequency and orientation, of faces or non-face 3D blobs, judging whether the person or blob was the same or different. On a match trial, the images were either identical or complementary (containing the remaining spatial frequency and orientation content). Relative to an identical pair of images, a complementary pair of faces, but not blobs, reduced matching accuracy and released fMRI adaptation in the fusiform face area.

Keywords

Face recognition; Fusiform face area; fMRI adaptation; Face vs. object recognition; Spatial representation; Face representation

1. Introduction

There are a number of striking differences in the recognition of faces and objects, even when the to-be-distinguished objects are as similar as faces. Faces are much more affected by contrast reversal (Galper, 1970; Subramaniam & Biederman, 1997) and orientation inversion (Yin, 1969). Faces show "configural" effects (Leder & Bruce, 2000; Tanaka & Farah, 1993). Differences between similar faces are extraordinarily diffcult to articulate whereas differences between similar objects tend to be readily describable in terms of their part differences (Biederman & Kalocsai, 1997).

Biederman and Kalocsai (1997) proposed that these differences could be understood if the representation of faces retained aspects of the original (V1 to V4) spatial filter representation, in a manner similar to that proposed by C. von der Malsburg's Gabor-jet model (Lades et al., 1993), although with translation and scale invariance (with scale expressed as cycles per stimulus rather than cycles per degree). Retention of the spatial frequency and orientation information allows storage of the fine metrics, pigmentation, and the surface luminance distribution important for the individuation of similar faces. However, such a representation is susceptible—as is human face matching performance—to variations in lighting conditions (Hill & Bruce, 1996; Liu, Colling, Burton, & Chaudhuri, 1999), viewpoint (Bruce, 1982; Hill, Schyns, & Akamatsu, 1997), and direction of contrast (Kemp, Pike, White, & Musselman, 1996).

*Corresponding author. Fax: +1 213 740 5687. *E-mail address:* E-mail: xyue@usc.edu (X. Yue).

Although non-face objects (as well as any visual stimulus) are similarly encoded in the early stages of visual processing, the ultimate representation would not be defined by the initial filter values, but "moderately complex features" (Kobatake & Tanaka, 1994) often corresponding to simple parts largely defined by orientation and depth discontinuities (Biederman, 1987; Grill-Spector, Kourtzi, & Kanwisher, 2001; Kobatake & Tanaka, 1994), allowing invariance to lighting conditions (Vogels & Biederman, 2002), viewpoint (Biederman & Bar, 1999), and direction of contrast (Subramaniam & Biederman, 1997). These features might be at a small scale when distinguishing among highly similar members of a subordinate-level class (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999).

The present study compared the matching of faces to the matching of novel, non-face objects designed to require the same kind of low-level discrimination of metric variation of smooth surfaces. To test the spatial-filter hypothesis of face representation, images of a set of faces and blobs, were filtered into eight frequencies and eight orientations (Fig. 2). Complementary pairs of these images were created in which every other combination of frequency and orientation in a radial checkerboard pattern in the Fourier domain was assigned to one image and the remaining combinations to the other image. Each image thus covered the entire range of frequencies and orientations, but in different combinations. Subjects performed a same-different matching task in which they had to judge whether a sequentially presented pair of images depicted the same or different person or blob, responding "same" when the images were either identical or complementary. On "different" trials, i.e., when the second stimulus was of a different face or blob than the first stimulus, the average physical differences were scaled by the Gabor-jet model (Lades et al., 1993) to be equally dissimilar, for faces and blobs. This is a needed control to insure that the demands for discriminating the stimuli were equivalent with respect to their physical similarity. If the representation of faces, but not blobs, retain aspects of the initial spatial frequency and orientation content, then matching of their complements should be more diffcult (relative to matching of the identical images) than the matching of complements of the blobs.

Faces are said to be represented "configurally' or "holistically." Perhaps the only neurocomputational account of such a representation is offered by the Gabor-jet model (Lades et al., 1993). Biederman and Kalocsai (1997) argued that the coverage of the face by Gabor kernels, especially those with larger receptive fields, would produce configural effects insofar as variations in contrast in any one region of the face would affect activation of kernels whose receptive fields were not centered at that region and that, conversely, any one kernel would be affected by contrast from all regions of the face within its receptive field. Such activation would be extremely diffcult to articulate, thus contributing to the ineffability of describing the differences between similar faces, a diffculty rarely witnessed when people discriminate highly similar non-face objects (Biederman & Kalocsai, 1997; Mangini & Biederman, 2004). This proposed test of the sensitivity of faces to the specific combinations of orientation and spatial frequency is thus of importance for assessing this account of the configural representation of faces.

Some researchers have suggested that mechanisms (and the representation) underlying face and object recognition are the same, differing only in their within-category expertise (Gauthier & Tarr, 1997). They have argued that experts who have achieved their expertise as adults distinguish objects in the domain of their expertise in the same way that faces are distinguished, viz., configurally, and that expertise for non-face objects is expressed at the same neural locus as faces (Tarr & Gauthier, 2000; Riesenhuber & Poggio, 2000). To address the possibility that a heightened sensitivity to the spatial frequency and orientation content of a face compared to those of a non-face object is a signature of expertise, we trained a group of subjects for over 8000 trials in discriminating a set of blobs (with a match-to-sample task) prior to the testing of blob discrimination in the scanner. Proponents of the expertise account have argued that

approximately 3000 discrimination trials are suffcient for the achievement of expertise (Gauthier, Williams, Tarr, & Tanaka, 1998).

The fusiform face area (FFA) is defined as a region in the fusiform gyrus that shows greater activation for faces than objects (Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995). While being scanned, blob experts and novices performed a same-different matching task of images of blobs and faces, responding same (by response button) to both Identical and Complementary images of the same faces or blobs. The sequential presentation allowed assessment of fMRI adaptation (Grill-Spector & Malach, 2001) and the extent to which a change in identity and/or the spatial content of the blob or face could produce a release from adaptation. The inclusion of a condition in which the same person's face is shown but with complementary spatial frequency and orientation content is important in determining the extent to which FFA is individuating faces as opposed to detecting faces. Prior experiments (e.g., Henson, Shallice, & Dolan, 2000) showing release from adaptation by presenting pictures of two different people compared to the identical image of the same person, confound a change of person with a change in low-level image properties (Grill-Spector, Knouf, & Kanwisher, 2004). The subjects were also administered localizer runs to define FFA and the lateral occipital complex (LOC), an area critical for object recognition (Fig. 3).

2. Methods

2.1. Stimuli

To capture the variations of the shapes of faces without the discontinuous edges and non-accidental properties readily distinguishing most objects, we adopted a method devised by Shepard and Cermak (1973) for 2D shapes to create 3D artificial objects that only differed from each other metrically. These objects were generated by adding a given orientation (one of eight) of the 2nd and 3rd harmonics of a sphere to the sphere and the 4th harmonic (Fig. 1a). The rotations produce a toroidal space of smooth, asymmetric 3D blobs (Fig. 1b). (Toroidal in the sense that the space curves around on itself so that columns 1 and 32 are identical and rows 1 and row 32 are identical. This characteristic avoids the problem that stimuli at the edges of a non-toroidal space would have no near neighbours on one side.) From this space, four equally spaced blobs were selected in which the harmonics were varied in size (mimicking, perhaps differences in the size of the nose, chin, etc. of faces.) The similarities between pairs of blobs were scaled using the Gabor-jet model, with values as illustrated in Fig. 1c. These similarities correlate almost perfectly with the speed and accuracy of judging whether two blobs are the same or different (Nederhouser, Mangini, & Biederman, 2002). Like other non-face objects, there is no effect of reversal of contrast when matching these blobs (i.e., matching a blob of positive direction of contrast to a blob with a negative direction of contrast), even when they are as pigmented as faces (Nederhouser, Mangini, Biederman, & Okada, 2003). We also matched the mean luminance of the face and blob stimuli, although their contrast was not matched. Blobs generally had a lower contrast than faces, but the difference in contrast in the super threshold range is unlikely to affect performance. In fact, in experimental conditions most critical for our purposes, subjects performed equally or better with the lower-contrast blobs than with the higher-contrast faces.

Eight-bit grey-level images of faces (ten male faces with hair and ears cropped) and blobs (ten blobs) were Fourier transformed and filtered by two complementary filters (Fig. 2). Both filters cut off the highest (above 181 cycles/image) and lowest (below 12 cycles/image, which is about 8 cycle/ face) spatial frequencies. The remaining part of the Fourier domain was divided into 64 areas, 8 orientations by 8 spatial frequencies. The orientation borders of the Fourier spectrum were set up in successive 22.5° steps. The spatial frequency range covered four octaves in steps of 0.5 octaves. The two filters comprised the complementary radial checkerboard patterns based on these divisions in the Fourier domain and shared no common combinations of spatial

frequency and orientation. The average similarity of the pairs of faces on different trials in the same-different matching task, as assessed by the Gabor-jet model, was equal to that of the blobs. This was true of both the complete images as well as their silhouettes.

2.2. Expertise training

Six subjects (*M*=26.3 years, *SD*=6.9), three females and three males, one left handed, were trained on discriminating a set of 64 blobs varying in the size of their harmonics for 8192 trials over 8 h-long sessions on one of four blob configurations (circled in Fig. 1b) to make them "blob experts." They were subsequently tested on a different set of blobs (i.e., with a configuration from one of the other circled blobs in Fig. 1b) which were identical to those presented to novices. If expertise is to be relevant to face recognition, then the training to expertise should transfer to new instances of a class. This did occur. The experts demonstrated markedly superior performance with a significantly lower error rate than the novice subjects when transferred, on their ninth session, to a new blob configuration (which matched that of the novices). This demonstration that the expertise was not to a specific set of blobs in general is important in that (Bukach, Gauthier, & Tarr, 2006) proponents of an expertise account of face recognition wish to address phenomena that all include effects, such as inversion, contrast reversal, and configuration, that are readily demonstrated with novel faces.

2.3. fMRI study

The main experiment used a rapid event-related (Burock et al., 1998), 2×2 design (same vs. different person/blob and same vs. different frequency-orientation combinations). A total of 1512 trials were run in 6 blocks, comprised of 3 blocks for faces and 3 for blobs. The sequences of blob and face blocks were counter-balanced across subjects. Subjects performed a same-different matching task on a sequence of two-stimuli (faces or blobs), each for 200 ms with a 300 ms ISI presented at the beginning of a 2-s trial on a screen viewable from within the bore of the magnet. In each trial, the images were presented at the center of the screen but at different sizes (4° for the large size and 2° for the small size) to reduce the possibility of using local features to perform the task. The order of the sizes of the presented images was randomized within each trial. The subjects were instructed to judge, by button press, whether the two images were of the same person or blob, regardless of the differences in size and spatial frequency content of the images. The subjects were the six blob experts, described previously, and six blob novices (*M*=27.7 years, *SD*=3.08, four females, and two males, one left handed). An additional two blob novices were run but excluded from the data analysis; one because of excessive head movement, the other because the face localizer failed to reveal a FFA (i.e., there was no region with stronger activation to faces than objects). The study was approved by the USC Internal Review Board. All subjects provided informed consent.

2.4. Imaging parameters

fMRI imaging was conducted in a 3T Siemens MAGNETOM Trio at the University of Southern California's Dana and David Dornsife Cognitive Neuroscience Imaging Center. Functional scanning used a gradient echo EPI PACE (prospective acquisition correction) sequence with 3D k-space motion correction (TR = 1 s; TE = 30 ms; flip angle = 65°; 14 slices; 64×64 matrix size with resolution $3 \times 3 \times 3$ mm) on the functional scans. The anatomical T1-weighted structural scans were done using MPRAGE sequence (TI = 1100, TR = 2070 ms, TE = 4.1 ms, Flip angle = 12°, 192 sagittal slices, 256×256 matrix size, $1 \times 1 \times 1$ mm voxels).

2.5. Region of interest (ROI)

Two localizer runs were administered, one in the beginning and the other at the end of the face matching trials. Each localizer run included five conditions (Fig. 3a): intact faces, objects, blobs, scrambled faces and textures. The faces were frontal views with neutral expressions,

randomly selected from a face database (http://pics.psych.stir.ac.uk) with equal numbers of male and female faces. The scrambled faces were created from the intact faces in such a way that the individual features were kept intact, altering only the relations among the features. The objects were grey-scale images photographed in our laboratory, with the background excluded by Photoshop (Adobe Systems Inc., San Jose, US). Textures were created by scrambling intact blobs with 8×8 patches of pixels so that no discernible blob structure was apparent. Each 18 s block of a localizer was composed of 36 images, each shown for 500 ms. The order of the five conditions was randomized in each of the localizer runs. A region of interest (ROI) for face activation (FFA) (Fig. 3b) was defined by a conjunction of two contrasts—faces vs. scrambled faces and faces vs. objects—at uncorrected $p < 10^{-4}$ (Fig. 3a). LOC (Fig. 3c) was defined by a conjunction of two contrasts—objects vs. textures and blobs vs. textures—at uncorrected $p < 10^{-6}$. No region with a minimum cluster-size of 10 voxels was differentially activated for blobs vs. objects, even with a low threshold.

2.6. fMRI data analysis

Brainvoyager QX (Brain Innovation BV, Maastricht, The Netherlands) was used to analyze the fMRI data. All data from a scan were preprocessed with 3D motion-correction, slice timing correction, linear trend removal and temporal smoothing with a high pass filter set to 3 cycles over the run's length. No spatial smoothing was applied to the data. Each subject's motion corrected functional images were coregistered with their same-session high-resolution anatomical scan. Then each subject's anatomical scan was transformed into Talairach coordinates. Finally, using the above transformation parameters, the functional scans were transformed into Talairach coordinates as well. All statistical tests reported were performed on this transformed data.

For the rapid event-related runs, a deconvolution analysis was performed on all voxels within each subject's localizer-defined ROI's to estimate the time course of the BOLD response. Deconvolution was computed using the BrainVoyager software by having fifteen 1-s shifted versions of the indicator function for each stimulus type and response (correct vs. incorrect) as the regressors in a general linear model.

In order to quantify the statistics between the deconvolved hemodynamic responses for the four conditions, the peak (average of the percent signal change for time points 6 and 7 s) for each correct response was computed for each condition for each subject. The peak location was found to be 6.75 s by fitting a double-gamma function (Boynton & Finney, 2003) on each subject's deconvolved hemodynamic response. The statistical analyses were performed on these values.

3. Results

3.1. Behavioural results

Fig. 4 shows the error rates on the same-different tasks performed in the scanner by Novices (a) and Experts (b), respectively. The expertise training had its expected effect, with the experts showing a significant advantage over novices when matching blobs; $F(1,10)=15.48$, $p < .01$, but not when matching faces; $F(1,10)=1.02$, ns. The decline in error rates over the eight sessions (see Section 2) for the experts trained on one set of blobs transferred to the different set of blobs used in the scanner.

As shown in Figs. 4a and b, matching complementary-images of a person's face produced markedly higher error rates than when matching identical-images of that person's face; $F(1,10)=58.57$, $p < .001$. This large cost of matching face complements was absent for blobs, where there was no significant difference in the error rates for Identical and Complementary matches;

$F(1,10)=3.76$, ns. These results held for both blob experts and novices; $F(1, 10)=2.28$, ns, for the interaction between stimulus type and expertise. The small cost of complementation on the blobs was confined to the novices, an effect opposite to what would be expected if expertise led to a representation of blobs that resembled that of faces. The cost of complementation when matching faces was not a consequence of a general increase in difficulty for matching faces: When the matching face was identical to the sample, error rates were lower than what they were when matching identical blobs.

The inclusion of a full range of spatial frequencies in each member of a complementary pair was an important feature of the complementary images. Face recognition has been shown to be more sensitive to the middle range of frequencies, approximately 8–13 cycles per face, than higher or lower frequencies (Costen, Parker, & Craw, 1994, 1996; Nasanen, 1999). Although both members of a complementary pair covered the entire range of spatial frequencies and orientations, it is possible that particular combinations of orientation and frequency are more critical for face recognition and the removal of such a combination from one member of a complementary pair disrupted recognition. We found no evidence for this. In the behavioural task, we compared performance with matching the images of faces in the upper circular checkerboard in the Fourier domain in Fig. 2 to that with matching the images of faces in the lower circular checkerboard. There was absolutely no difference in performance with the two sets of images, with the mean error rates for the two checkerboards equal to 10.35 and 10.57%.

3.2. Ideal observer analysis

It is possible that the greater cost of matching complements of faces compared to blobs was a result of intrinsic differences between these classes of stimuli. To assess whether this was the case, an ideal observer model (Gold, Bennett, & Sekuler, 1999; Pelli, Farell, & Moore, 2003; Tjan, Braje, Legge, & Kersten, 1995) was defined, which was assumed to have perfect pixel knowledge of all 20 images. White noise was added to the images to limit performance (Pelli, 1990) and match it to the level of humans. The ideal observer sums the likelihoods of all possible combination of images that correspond to the "same" response and compares the sum against the sum-of-likelihood for the "different" response. There was no difference in the ideal observer's performance between faces and blobs on the identical and complementary conditions (Fig. 4c), suggesting that the results with the human observers were not a consequence of inherent differences in the stimuli that rendered the matching of complements of faces more susceptible to error than the matching of complements of blobs.

3.3. fMRI results

3.3.1. FFA—Activation of both the right and left FFA was markedly greater when matching faces than blobs as shown in Fig. 5a and b; $F(1,10)=79.03$, $p < .001$ for right FFA and $F(1,10) =78.30$, $p < .001$ for left FFA. The magnitude of this difference was the same for experts and novices (i.e., there was no interaction between Expertise and Stimulus Class) $F(1,10) < 1.0$. All the above effects were manifested equally by blob experts and novices; $F(1, 10) < 1.00$, for all comparisons (Fig. 5a and b). There was no reliable differential activation between experts and novices for matching blobs at FFA as a consequence of changes in frequency and individuation (both $Fs(1,10) < 1.00$). However, the three way interaction of Frequency-Orientation Combination Change $\times$ Blob Change $\times$ Expertise was significant, $F(1,10)=9.10$, $p < .05$. The pattern shown by the experts (compared to the novices) for this interaction did not resemble the pattern for faces, i.e., greater release from adaptation to a change in the spatial content or identity of the blobs. Instead, the condition showing the smallest percent signal change for the experts was when both blob and frequency content changed. For the novices, this condition showed the greatest release from adaptation, consistent to what was found for faces.

Fig. 6, shows the combined response of blob experts and novices to faces on correct trials in right FFA. A change in the spatial content of an image of a face and/or a change in the person resulted in a significant release from adaptation in right FFA, compared to the response to an identical image (for frequency, $F(1, 10) = 42.91$, $p < .001$; for individuation, $F(1,10)=7.10$, $p < .05$).

3.3.2. LOC—The release-from-adaptation results for faces in right FFA were not witnessed in either the right or left lateral occipital complex (LOC) nor was there a difference between experts and novices so the data are shown in a combined figure (Fig. 7). That is, a change in identity and/or spatial frequency and orientation content of a face did not result in a greater BOLD response (i.e., a greater release-from-adaption) in LOC than repetition of the identical image of a face. However, presentation of blobs to blob experts did produce a greater BOLD response bilaterally in LOC compared to novices ($F(1,22)=7.18$, $p < .05$), as shown in Fig. 8. In addition, a change in the identify of a blob for the blob experts produced a release from adaptation at right LOC, $F(1,5)=9.76$, $p < .05$, (Fig. 9). However, no significant release from adaptation was found for blob experts when the second blob was a complement of the first in either right or left LOC (right LOC: $F(1,5)=1.21$, ns; left LOC: $F(1,5) = 0.40$, ns) (Fig. 9). Despite the lack of significance, there was a small release from adaptation of a change in the frequency-orientation combination of a blob in right LOC for the experts. However, this non-significant effect was only about one-third of that witnessed for faces in FFA: In terms of percent signal change in the BOLD response for the experts, it was .036 for faces in right FFA [$t(5)=3.44$, $p = .02$] and .011 for blobs in right LOC. We also note that BOLD responses from different cortical regions are not strictly speaking directly comparable.

4. Discussion

The results provide strong support for the hypothesis that the representation of faces, but not objects, retains aspects of their original spatial frequency and orientation coding, both behaviourally and in their activation of FFA. What is disruptive about matching complements of images of faces compared to non-face objects in the present experiment is not the absence of particular frequencies or particular orientations but that these combinations are *different*.

Expertise for making within-category discriminations of non-face objects, i.e., the bobs, did not result in greater activation of the FFA, either in the magnitude or in the manner of faces. By employing blobs that required discriminating the metrics of smoothly varying surfaces, we attempted to engage FFA, if such engagement depended on processing those kinds of low-level features. But blob experts and novices produced the same magnitude and pattern of FFA activation and the same pattern of release from adaptation (or the lack thereof) when discriminating faces and blobs. Although the experts performed substantially better than our novices, they showed the same invariance to the spatial complementation of the blobs. Insofar as sensitivity to the spatial composition may be an indicant of a configural representation, our results are consistent with the recent results of Robbins and McKone (2006) who failed to find that expertise produced face-like—especially configural—processing of dogs in dog experts, thus failing to replicate the results of Diamond and Carey (1986).

Where might expertise be expressed? A difference that did emerge between experts and novices in the present investigation was in the right LOC, where there was a greater response to the blobs for the experts than for the novices (Fig. 8) and a significant release from adaptation to a change in blob identity (Fig. 9). But in contrast to what was observed for faces, there was only a small and non-significant release from adaptation due to spatial complementation for the experts (Fig. 9). Previous studies (Gauthier, Skudlarski, Gore, & Anderson, 2000) suggesting FFA involvement in expertise did not report any differential activity in LOC. Nonetheless, the lack of a reliable release from adaptation with the presentation of a

complementary image of a blob suggests that whatever the representation experts developed for discriminating blobs, it did not as strongly reflect the sensitivity to spatial frequency/ orientation content evident with the representation of faces in right FFA.

Our results differ from those of Rotshtein (Rotshtein et al., 2005) et al. who reported that the right FFA was sensitive to identity but not to physical image variations. Rotshtein et al. used famous faces in their experiment, which might be processed differently from the unfamiliar faces used in the present investigation (Malone, Morris, Kay, & Levin, 1982). The effects of expertise that are under examination here are not that for particular faces but for faces in general. Further research is needed to resolve the role of facial familiarity as well as the neural bases of facial familiarity.

We confined our discussion of our fMRI results to FFA and LOC, but not to the occipital face area (OFA) and superior temporal sulcus (STS). Some investigators have suggested that OFA might be engaged in encoding physical information of faces and STS with the coding of gaze and expression (Hooker et al., 2003). However, the face localizers (see Section 2) did not produce significant differential activation in either of these areas. Some recordings from human patients (McCarthy, Puce, Belger, & Allison, 1999) have suggested that single facial features may be suffcient for FFA activation. Our scrambled face localizer, which contained intact face features, such as the eyes and nose, but with scrambled relations so that it was impossible to achieve a coherent perception of a complete face, failed to produce significant activation in either OFA or STS.

The results of the present study are consistent with FFA's involvement in individuation of faces in that the release produced by a change in individuation, despite its smaller physical image change (as assessed by the Gabor-jet model) compared to the change of spatial content, nonetheless produced as much release as a complete change in spatial content. What is required to determine if FFA accomplishes individuation beyond that produced by any facial image change is a study in which release from adaptation produced by different images (say poses) of the same face are compared with pictures of different individuals where the two conditions of image change are equated for their physical differences.

Our results provide strong evidence, both behavioural and neural, that what makes the recognition of faces special vis-à-vis non-face objects is that the representation of faces retains aspects of combinations of their spatial frequency and orientation content extracted from earlier visual areas.

Acknowledgments

References

Biederman I. Recognition-by-components: a theory of human image understanding. Psychological Review 1987;94:115–147. [PubMed: 3575582]

Biederman I, Bar M. One-shot viewpoint invariance in matching novel objects. Vision Research 1999;39:2885–2889. [PubMed: 10492817]

Biederman I, Kalocsai P. Neurocomputational bases of objects and face recognition. Philosophical Transactions of the Royal Society of London: Biological Sciences 1997;352:1203–1219.

Biederman I, Subramaniam S, Bar M, Kalocsai P, Fiser J. Subordinate-level object classification reexamined. Psychological Research 1999;62:131–153. [PubMed: 10472199]

Boynton GM, Finney EM. Orientation-specific adaptation in human visual cortex. Journal of Neuroscience 2003;23:8781–8787. [PubMed: 14507978]

Bruce V. Changing faces: visual and nonvisual coding in face recognition. British Journal of Psychology 1982;73:105–116. [PubMed: 7059746]

Bukach CM, Gauthier I, Tarr MJ. Beyond faces and modularity: the power of an expertise framework. Trends in Cognitive Science. 2006in press.

Burock MA, et al. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional fMRI. Neuroreport 1998;9:3735–3739. [PubMed: 9858388]

Costen NP, Parker DM, Craw I. Spatial content and spatial quantisation effects in face recognition. Perception 1994;23:129–146. [PubMed: 7971093]

Costen NP, Parker DM, Craw I. Effects of high-pass and low-pass spatial filtering on face identification. Perception & Psychophysics 1996;58:602–612. [PubMed: 8934690]

Diamond R, Carey S. Why faces are and are not special: an effect of expertise. Journal of Experimental Psychology: General 1986;115:107–117. [PubMed: 2940312]

Galper RE. Recognition of faces in photographic negative. Psychonomic Science 1970;19:207–208.

Gauthier I, Skudlarski P, Gore JC, Anderson AW. Expertise for cars and birds recruits brain area involved in face recognition. Nature Neuroscience 2000;3:191–197.

Gauthier I, Tarr MJ. Becoming a "Greeble" expert: Exploring the face recognition mechanism. Vision Research 1997;37:1673–1682. [PubMed: 9231232]

Gauthier I, Williams P, Tarr MJ, Tanaka J. Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Research 1998;38(15–16):2401–2428. [PubMed: 9798007]

Gold J, Bennett PJ, Sekuler AB. Identification of bandpassed filtered letters and faces by human and ideal observers. Vision Research 1999;39:3537–3560. [PubMed: 10746125]

Grill-Spector K, Knouf N, Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. Nature Neuroscience 2004;7:555–562.

Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. Vision Research 2001;41:1409–1422. [PubMed: 11322983]

Grill-Spector K, Malach R. fMRI-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica 2001;107:293–321. [PubMed: 11388140]

Henson R, Shallice T, Dolan R. Neuroimaging evidence for dissociable forms of repetition priming. Science 2000;287:1269–1272. [PubMed: 10678834]

Hill H, Bruce V. The effects of lighting on the perception of facial surface. Journal of Experimental Psychology: Human Perception and Performance 1996;22:986–1004. [PubMed: 8756964]

Hill H, Schyns PG, Akamatsu S. Information and viewpoint dependence in face recognition. Cognition 1997;62:201–222. [PubMed: 9141907]

Hooker CI, et al. Brain networks for analyzing eye gaze. Brain Research: Cognitive Brain Research 2003;17:406–418. [PubMed: 12880911]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 1997;17:4302–4311. [PubMed: 9151747]

Kemp R, Pike G, White P, Musselman A. Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation. Perception 1996;25:37–52. [PubMed: 8861169]

Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. Journal of Neurophysiology 1994;71:856–867. [PubMed: 8201425]

Lades M, et al. Distortion invariant object recognition in the dynamics link architecture. IEEE Transactions on Computers: Institute of Electrical and Electronics Engineers 1993;42:300.

Leder H, Bruce V. When inverted faces are recognized: the role of configural information in face recognition. The Quarterly Journal of Experimental Psychology 2000;53:513–536. [PubMed: 10881616]

Liu CH, Colling CA, Burton M, Chaudhuri A. Lighting direction affects recognition of untextured faces in photographic positive and negative. Vision Research 1999;39:4003–4009. [PubMed: 10748933]

Malone DR, Morris HH, Kay MC, Levin HS. Prosopagnosia: a double dissociation between recognition of familiar and unfamiliar faces. Journal of Neurology, Neurosurgery, and Psychiatry 1982;45:820–822.

Mangini MC, Biederman I. Making the inefiable explicit: Estimating the information employed for face classification. Cognitive Science 2004;28:209–226.

McCarthy G, Puce A, Belger A, Allison T. Electrophysiological studies of human face perception II: response properties of face-specific potentials generated in occipitotemporal cortex. Cerebral Cortex 1999;9:431–444. [PubMed: 10450889]

Nasanen R. Spatial frequency bandwidth used in the recognition of facial images. Vision Research 1999;39:3824–3833. [PubMed: 10748918]

Nederhouser M, Mangini M, Biederman I. The matching of smooth, blobby objects–but not faces–is invariant to differences in contrast polarity for both naïve and expert subjects. Journal of Vision 2002;2:745a.

Nederhouser M, Mangini M, Biederman I, Okada K. Invariance to contrast inversion when matching objects with face-like surface structure and pigmentation. Journal of Vision 2003;3:93a.

Pelli DG, Farell B, Moore DC. The remarkable ineffciency of word recognition. Nature 2003;423:752–756. [PubMed: 12802334]

Pelli, DG. The quantum effciency of vision. In: Blakemore, C., editor. Vision: Coding and effciency. Cambridge: University Press; 1990. p. 3-24.

Puce A, Allison T, Gore JC, McCarthy G. Face-sensitive regions in human extrastriate cortex studied by functional MRI. Journal of Neurophysiology 1995;74:1192–1199. [PubMed: 7500143]

Riesenhuber M, Poggio T. Models of object recognition. Nature Neuroscience 2000;3:1199–1204.

Robbins R, McKone E. No face-like processing for objects-of-expertise in three behavioural tasks. Cognition. 2006in press.

Rotshtein P, et al. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nature Neuroscience 2005;8:107–113.

Shepard RN, Cermak GW. Perceptual-cognitive explorations of a toroidal set of free-from stimuli. Cognitive Psychology 1973;4:351–377.

Subramaniam S, Biederman I. Does contrast reversal affect object identification? Investigative Ophthalmology & Visual Science 1997;38:998.

Tanaka JW, Farah MJ. Parts and wholes in face recognition. The Quarterly Journal of Experimental Psychology 1993;46A:225–245. [PubMed: 8316637]

Tarr MJ, Gauthier I. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nature Neuroscience 2000;3:764–769.

Tjan BS, Braje WL, Legge GE, Kersten DJ. Human effciency for recognizing 3-D objects in luminance noise. Vision Research 1995;35:3053–3069. [PubMed: 8533342]

Vogels R, Biederman I. Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. Cerebral Cortex 2002;12:756–766. [PubMed: 12050087]

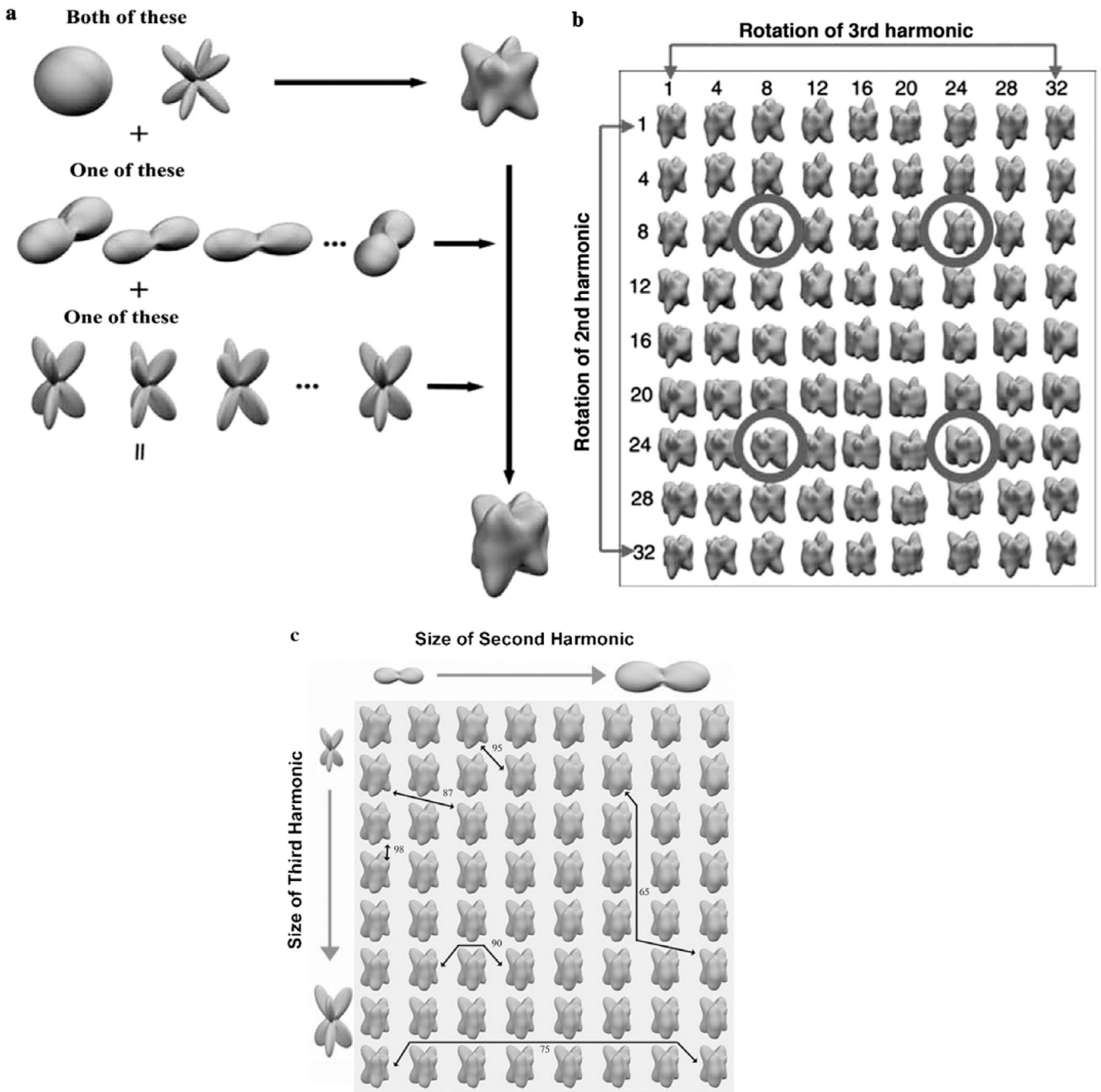Yin RK. Looking at upside down faces. Journal of Experimental Psychology 1969;81:141–145.

Fig. 1.
Generation of visual stimuli. (a) Blobs were generated by combining the 2nd and 3rd harmonics of a sphere in eight different orientations. (b) Blob space produced by combining different orientations of the 2nd and 3rd harmonics, as shown by the orientations shown above and to the left of the blob space. Proximate blobs were highly similar in shape and those distant were less similar, as confirmed by a Gabor-jet similarity measure. The four circled blobs were the seeds used to generate the blob spaces, defined by variation in the sizes of the 2nd and 3rd harmonics. (c) A blob space generated by holding constant the orientation of the harmonics but only varying their size, as shown above and to the left of the blob space. The illustrated space is generated from the upper left circled blob in (b). The variations in sizes of the

harmonics are taken to mimic the variation in the sizes and distances of facial parts. Both experts and novices were tested with one of the four spaces, but the experts gained their expertise on a space defined by a seed diagonally opposite to their test space. Numbers along arrowed lines pointing to pairs of blobs show the Gabor-jet similarity values for those pairs as a percent of an identity match (= 100).
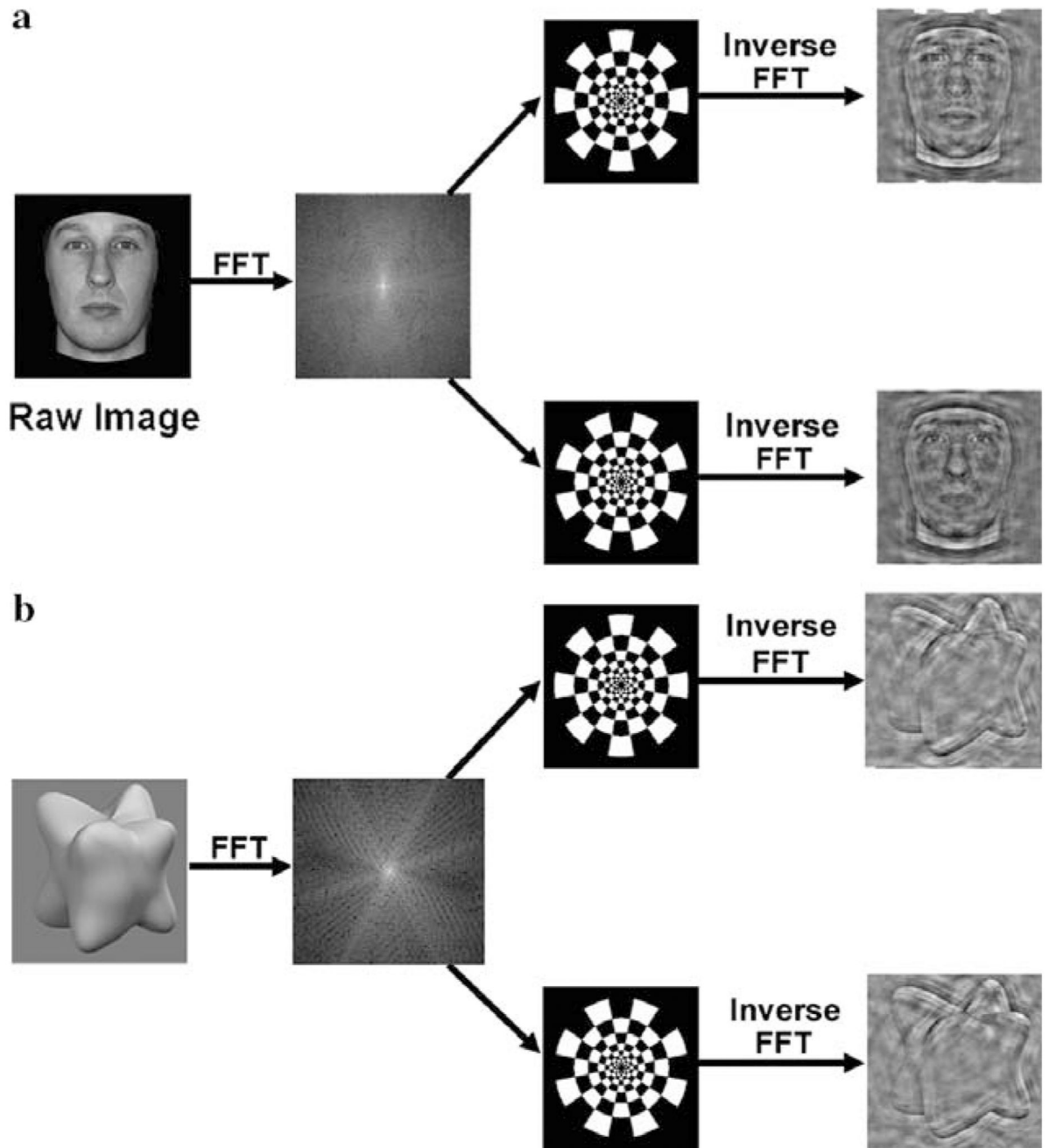
Fig. 2.
Complementary images were created by filtering an image in the Fourier domain into 8 orientations by 8 spatial frequencies. The content of every other 32 frequency-orientation combinations, as illustrated by the circular checkerboards, was assigned to one image of a complementary pair and the remaining content to the other member of that pair. Each member of a complementary pair thus had all 8 orientations and all 8 frequencies but in different combinations. The Fourier-domain images were then converted to images in the spatial domain by inverse FFT. (a) An example with a face. (b) An example with a blob.
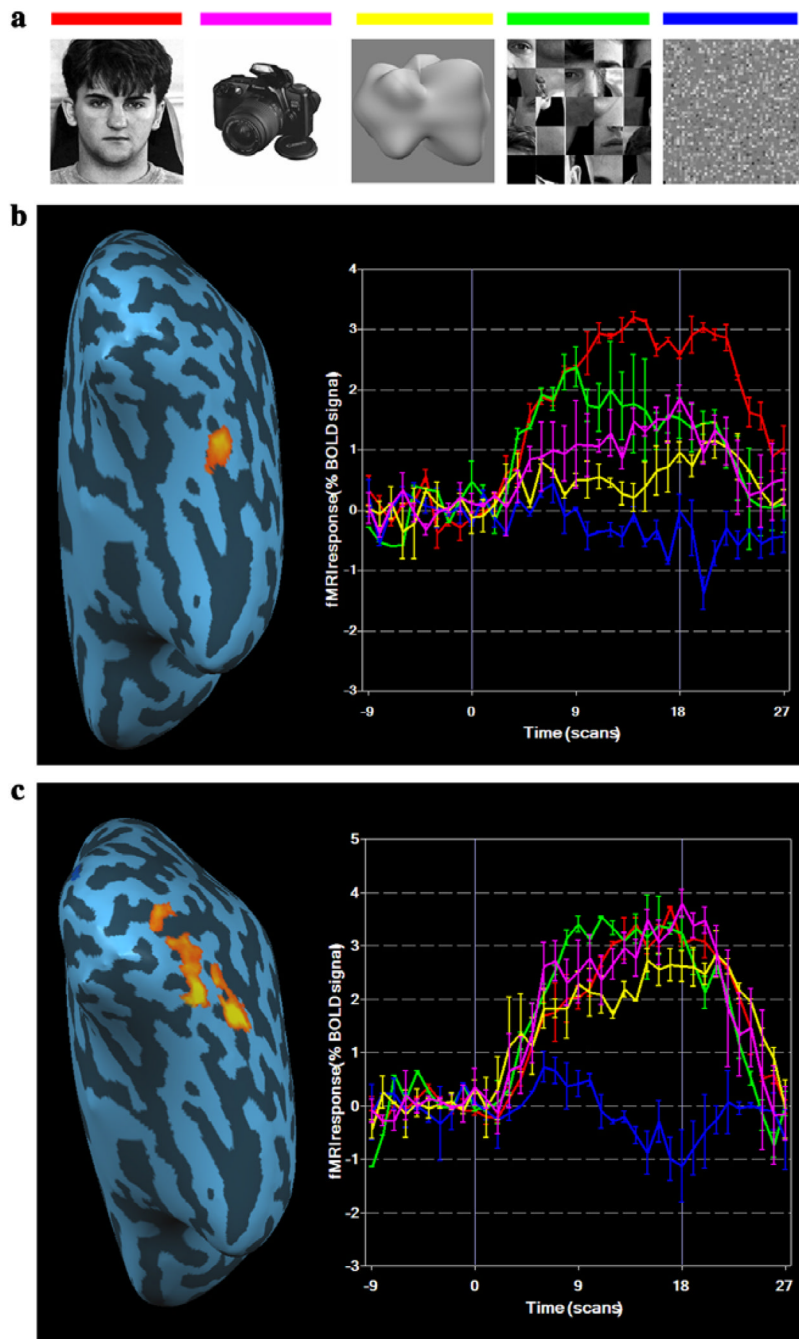
Fig. 3.
Localizer images and BOLD responses at the right FFA and LOC. (a) Examples of images were used in the localizer runs. (b) The left panel shows one subject's right FFA defined by a conjunction of Faces minus Object and Faces minus Scrambled-faces projected to an inflated brain. The right panel shows the event-related average percent BOLD signal change for the five classes of images. The colours of the lines are indicated in (a). (c) The left panel shows one subject's right LOC defined by a conjunction of Objects minus Textures and Blobs minus Textures. The right panel shows event-related average BOLD response for five classes of images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)
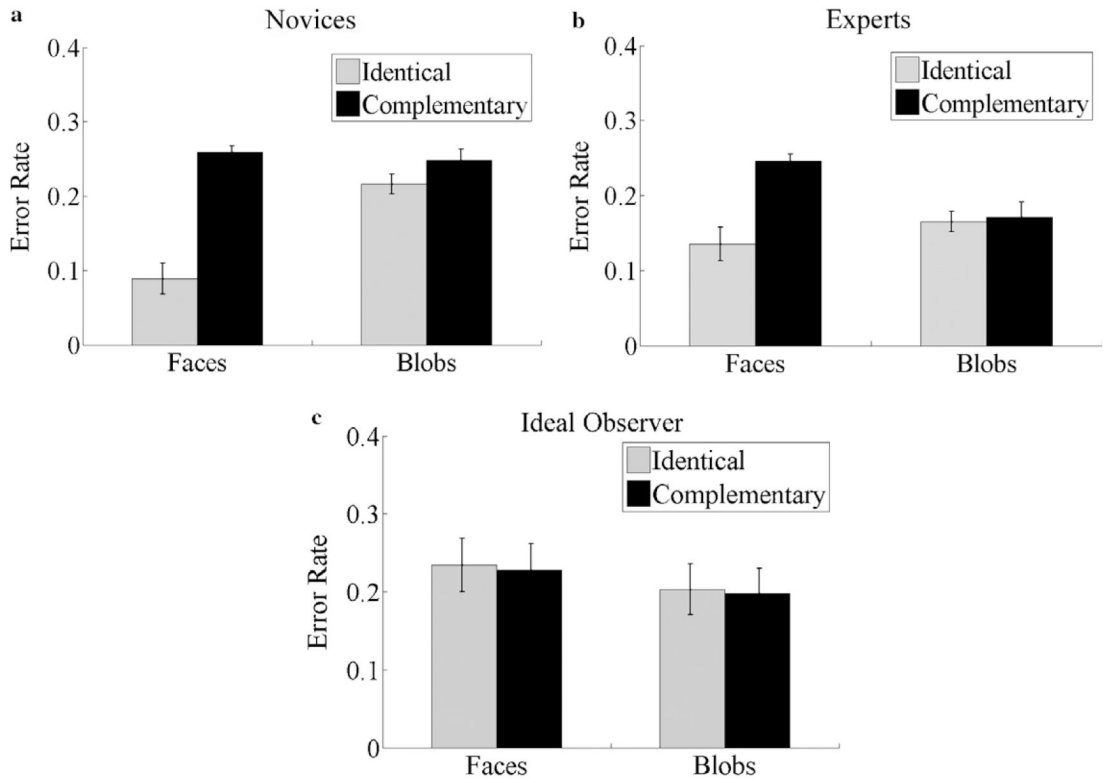
Fig. 4.
Same-different proportion error rates for matching Identical and Complementary faces and blobs for (a) novices, (b) experts, and (c) the Ideal Observer. Error bars are SEM. Because the main effect of identity (same vs. different person/blob) was not significant, they were combined in these figures.
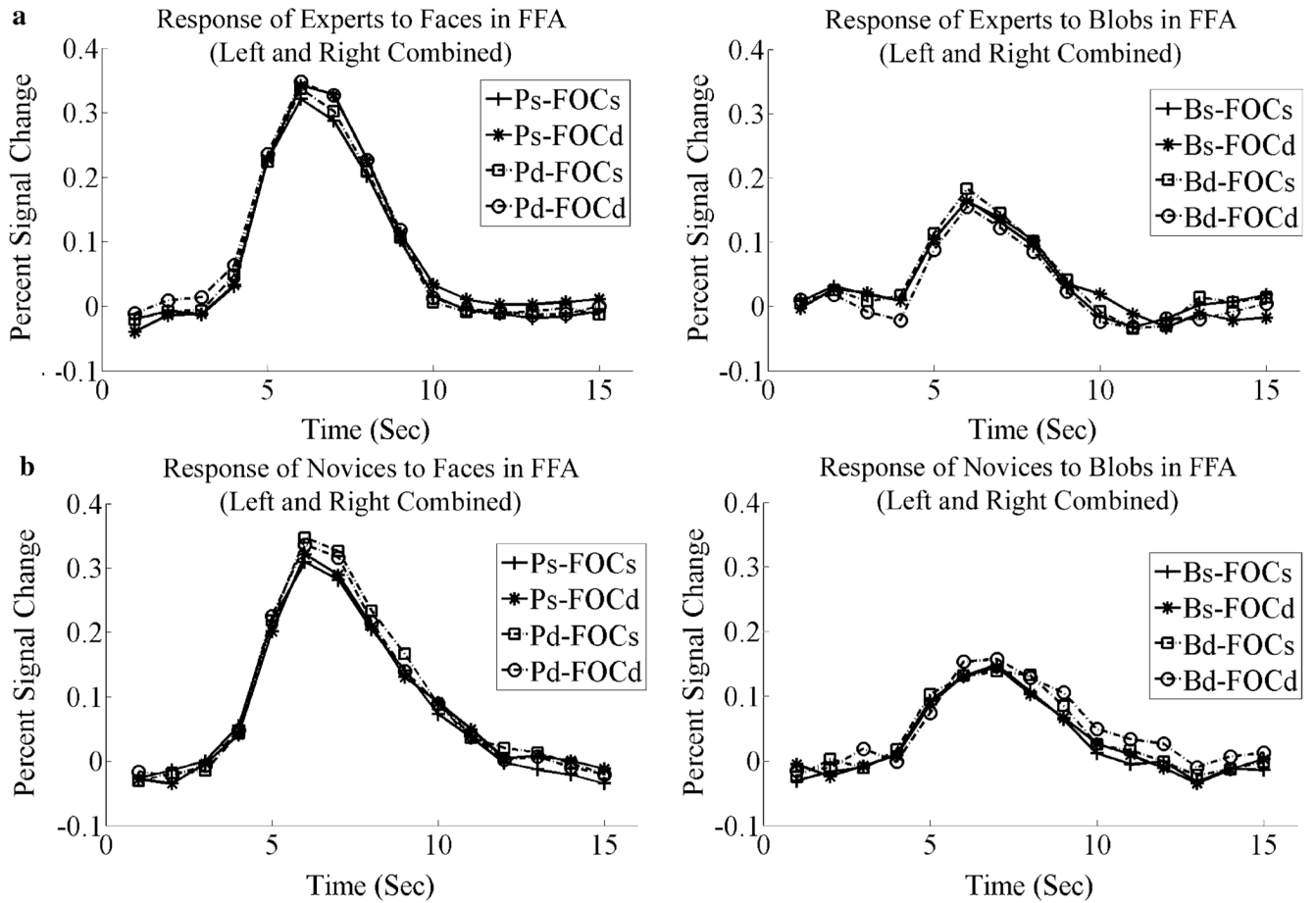
Fig. 5.
Hemodynamic response functions of blob experts and blob novices to blobs and faces at FFA (left and right hemispheres combined). (a) Hemodynamic response functions of blob experts to faces (left) and blobs (right). (b) Hemodynamic response functions of blob novices to faces (left) and blobs (right). Labels: Ps-FOCs, person same and frequency-orientation combination same; Ps-FOCd, person same and frequency-orientation combination different; Pd-FOCs, person different and frequency-orientation combination same; Pd-FOCd, person different and frequency-orientation combination different; Bs-FOCs, blob same and frequency-orientation combination same; Bs-FOCd, blob same and frequency-orientation combination different; Bd-FOCs, Blob different and frequency-orientation combination same; Bd-FOCd, blob different and frequency-orientation combination different.
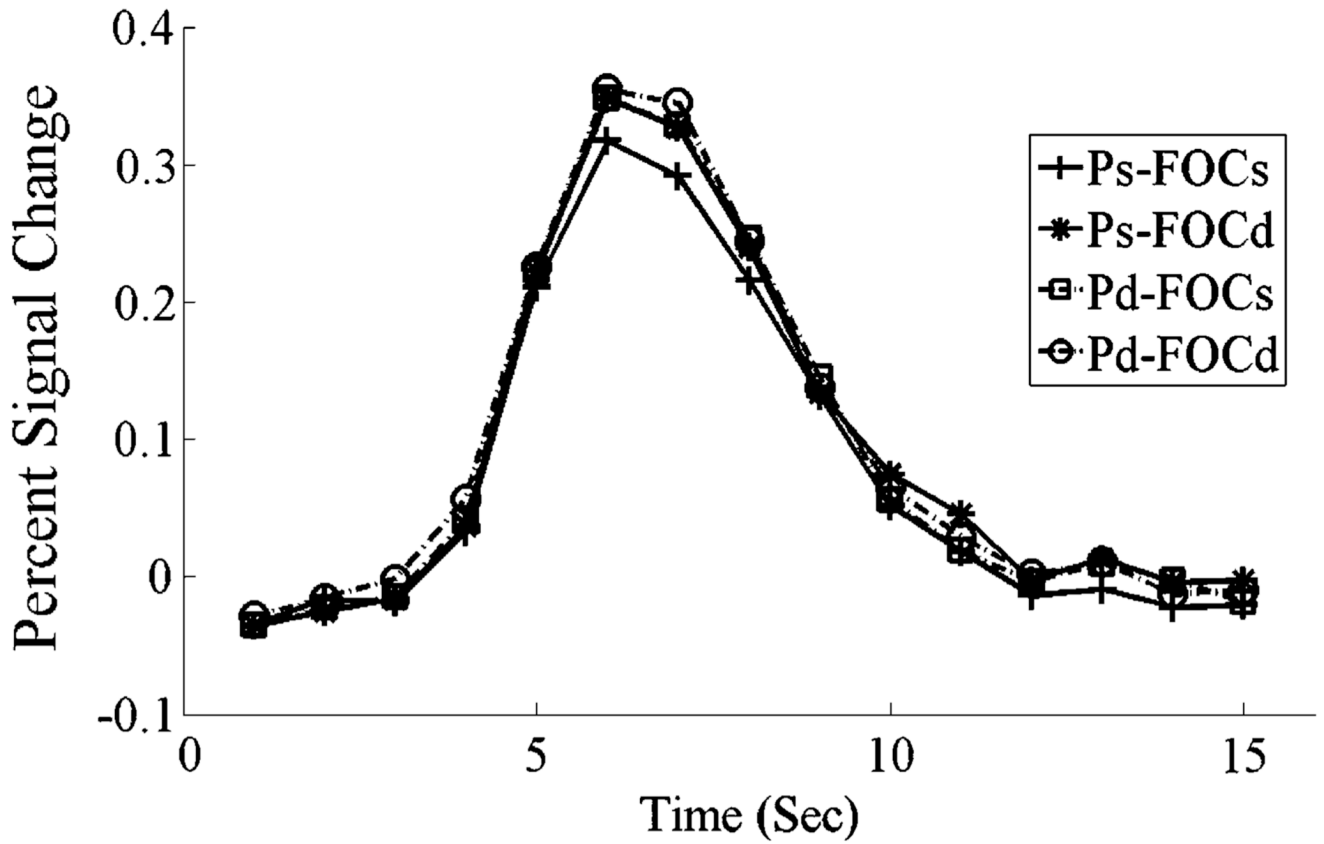
Fig. 6.
Hemodynamic response functions for correct trials when matching faces in right FFA, combined across blob experts and novices, showing a release from adaptation for a change in the person (identity) or a change in the frequency-orientation combination.
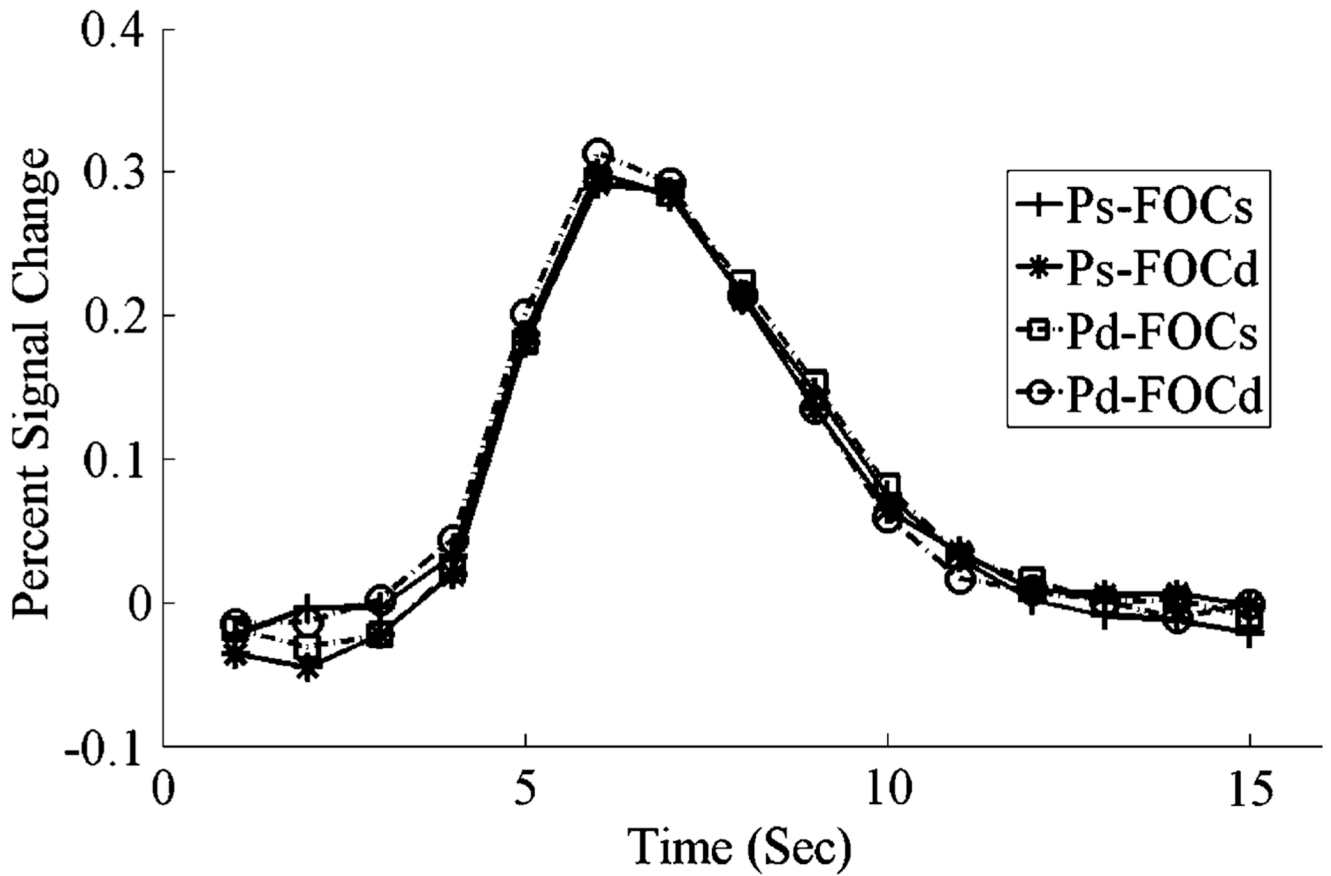
Fig. 7.
Hemodynamic response functions in LOC for changes in Identity and frequency-orientation combinations when matching faces. As there were no differences between left and right LOC and blob experts and novices, the data are shown collapsed over these variables. Neither person nor spatial content produced a release from adaptation (in contrast to the pattern produced in right FFA for faces shown in Fig. 6).
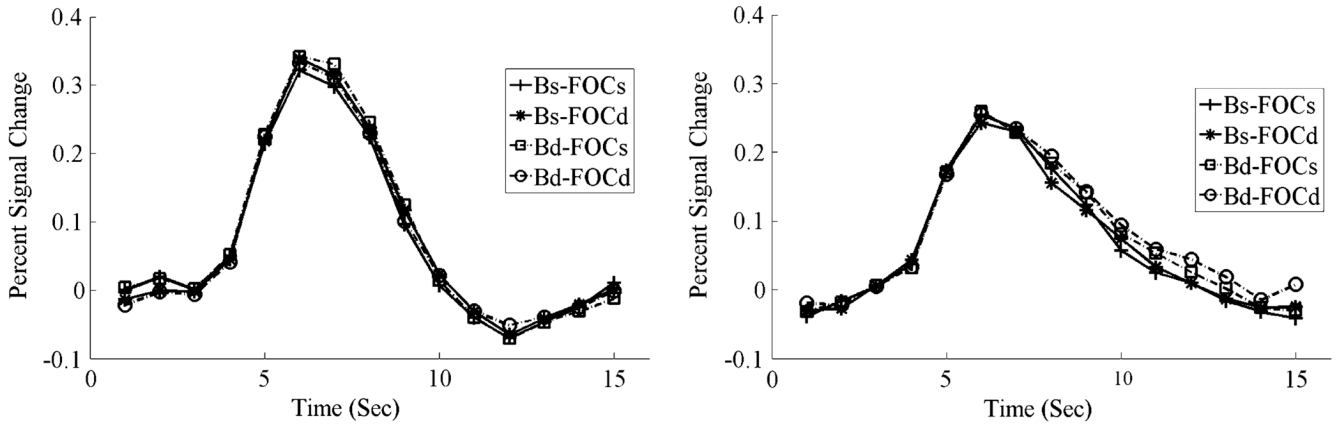
Fig. 8.
Hemodynamic response functions in LOC for blob experts and novices. Blob experts (left) showed larger response to blobs than blob novices (right) in all conditions.
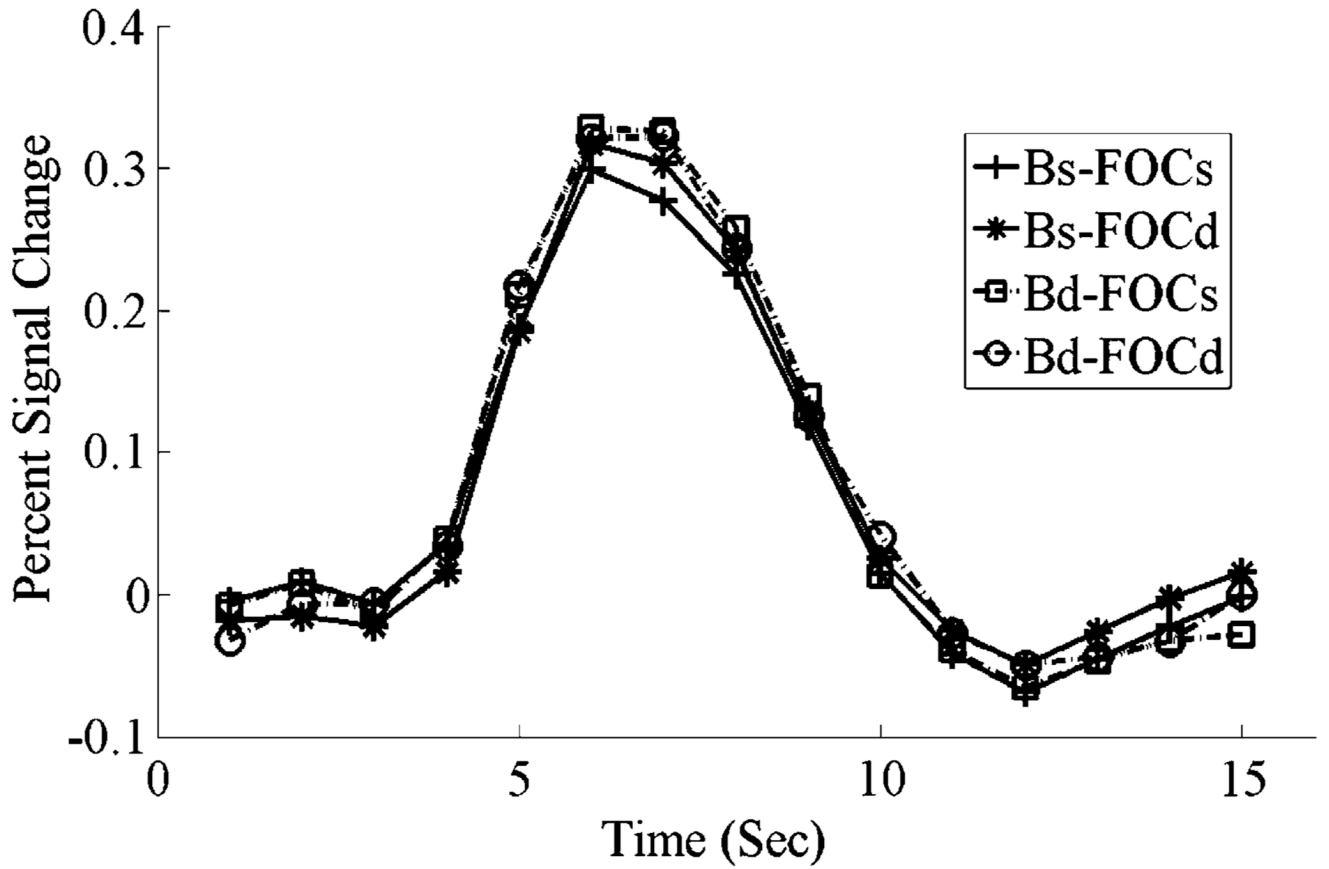
Fig. 9.
Hemodynamic response functions in the right LOC for blob experts when matching blobs.
There was a significant release from adaptation with a change in blob identity, but the apparent
release with a change in frequency-orientation content was non-significant.