

Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint

Xiaokun Xu

Department of Psychology, University of Southern California,
Los Angeles, CA, USA



Department of Psychology, University of Southern California,
Los Angeles, CA, USA, &

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



Irving Biederman

Face recognition involves collaboration of a distributed network of neural correlates. However, how different attributes of faces are represented has remained unclear. We used functional magnetic resonance imaging-adaptation (fMRIa) to investigate the representation of viewpoint, expression, and identity of faces in the fusiform face area (FFA) and the occipital face area (OFA). In an event-related experiment, subjects viewed sequences of two faces and judged whether they depicted the same person. The images could vary in viewpoint, expression and/or identity. Critically, the physical similarity between view-changed and between expression-changed faces of the same person were matched by the Gabor-jet metric, a measure that predicts almost perfectly the effects of image similarity on face discrimination performance. In FFA, changes of identity produced the largest release from adaptation followed by changes of expression; but the release caused by changes of viewpoint was smaller and not reliable. OFA was sensitive only to changes in identity, even when image changes produced by identity variations were matched to those of expression and orientation. These results suggest that FFA is involved in the perception of both identity and expression of faces, a result contrary to the hypothesis of independent processing of changeable and invariant attributes of faces in the face-processing network.

Keywords: face perception, FFA, fMRI-adaptation, identity, OFA, viewpoint, expression

Citation: 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.

Introduction

We readily extract a number of attributes from a face, such as identity, expression, orientation, age, sex, and attractiveness. Whether there are different cortical loci for the processing of these attributes remains unclear. Three areas that have received the most investigative attention with respect to the representation of information from faces are the Fusiform Face Area (FFA) (e.g., Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996), the Occipital Face Area (OFA) (Rossion et al., 2003; Rotshtein, Henson, Treves, Driver, & Dolan, 2005), and the Superior Temporal Sulcus (STS) (Fang, Murray, & He, 2007b; Winston, Henson, Fine-Goulden, & Dolan, 2004).

Bruce and Young (1986) proposed a hierarchical model with distinct pathways for processing facial identity, emotional expression and eye gaze. A corresponding neural model was proposed by Haxby, Hoffman, and Gobbini (2000) based on a review of single unit and fMRI studies in both humans and monkeys (Hoffman & Haxby, 2000; Puce et al., 1996). The model assumed that the changeable aspects of faces, such as facial expression, eye-gaze and mouth movement, are processed in STS,

whereas the invariant properties of faces such as facial identity were processed in FFA. OFA, according to Haxby et al.'s scheme, receives input from early visual stages and feeds the output to both FFA and STS. This model is further supported by temporal and anatomical segregation with evidence from MEG and EEG studies in which the early (~90 ms from stimulus onset) and late (~170 ms) signatures in the time course of face processing were located in inferior occipital and temporal cortices, respectively (Liu, Harris, & Kanwisher, 2002; Smith, Fries, Gosselin, Goebel, & Schyns, 2009; Sugase, Yamane, Ueno, & Kawano, 1999). Electrophysiology in monkeys also suggested a separation in which neurons in STS are tuned to expression and orientation, whereas those in the inferior temporal gyrus (ITG) are tuned to identity (Eifuku, De Souza, Tamura, Nishijo, & Ono, 2004; Hasselmo et al., 1989).

In fMRI-adaptation (fMRI-a) the repetition of identical stimuli leads to a decreased BOLD signal compared to non repeated stimuli (Grill-Spector et al., 1999; Krekelberg, Boynton, & van Wezel, 2006). Winston et al. (2004) reported that a change of identity but not expression between two sequentially presented faces produced a release from adaptation compared to identical faces in FFA while a change of expression but not identity resulted

in a release of adaptation in mid-STS, suggesting a functional segregation of identity and expression. Some evidence for independent coding of facial identity and expression has been derived from visual aftereffects, in which, for example, prolonged viewing of the face of person A results in a biased perception of a face with ambiguous identity (i.e., a morph between the faces of persons A and B) as person B. Fox and Barton (2008) found that the aftereffect of facial identity could be largely transferred across different emotional expressions. Similarly, the aftereffect of face orientation was found to be independent of identity. That is, adaptation to a face oriented to the left results in a frontal-oriented face appearing to be oriented to the right with only a small cost when the depicted individuals are changed (Fang, Ijichi, & He, 2007a), but not across exemplars from different categories like faces and paperclips (Fang & He, 2005). Based on such results, Fang et al. suggested that the processing of facial identity and viewpoint are largely independent.

There are challenges to the notion of independent processing of changeable and invariant properties of faces (Calder & Young, 2005). In the identity-dependent expression aftereffect experiments the shift in the perception of emotion only partially transferred across faces of different people (Campbell & Burke, 2009; Ellamil, Susskind, & Anderson, 2008; Fox and Barton, 2008; Vida & Mondloch, 2009). The existence of identity-dependent and identity-independent emotion aftereffects suggests at least partially overlapped representations of identity and expression. Single units in monkey STS are tuned to features such as iris size, inter-eye distance, face aspect ratio and their interaction (Freiwald, Tsao, & Livingstone, 2009). The combination of these features contributes not only to facial expression, but also to the identification of faces (Hasselmo et al., 1989). The level to which the representation of face identity, expression and view are separated thus needs to be further clarified.

A difficulty in interpreting prior research is that the physical similarities between stimuli for the different classes of image change are often unspecified. How does one render the physical change in, say, a change in expression to be equal to a change in a person? Without such scaling, one doesn't know if the release of an adapted BOLD response is an effect of high-level properties, such as identity or expressed emotion, or merely the physical changes, per se, that would be produced by any change in the image. Fox and Barton (2008) used the contrast thresholds of an ideal observer to scale stimulus similarity for different attributes for face (i.e. different identity vs. different expression). Although an ideal observer scaling does offer some control over low-level stimulus factors, it assumes a pixel-based representation that might be more characteristic of retinal and lateral geniculate bases of similarity than early cortical stages. (It would be possible to implement an ideal observer that would take as input the output of the Gabor-jet model which would overcome this problem.) Although there are some special cases of

subcortical face processing that are done with inputs directly from the lateral geniculate (e.g., Adolphs, Tranel, Damasio, & Damasio, 1994), they remain special cases.

The Gabor-jet model (Biederman & Kalocsai, 1997; Lades et al., 1993) goes beyond a pixel representation in that it captures the essential characteristics of the multi-scale, multiorientation of V1 filtering. This scaling model is justified by its extremely high prediction of the psychophysical similarity of face discrimination (Yue, Biederman, Mangini, & Malsburg, 2010a).¹ We used the Gabor-jet model to equate the magnitude of the image changes produced by changes in expression, individuation, and rotation in depth.

The stimulus scaling allowed us to investigate the loci sensitive to different aspects of faces, namely, expression, viewpoint and identity, without possible confounds of low-level image similarity in an event-related fMRI adaptation paradigm (Kourtzi & Kanwisher, 2000). Our previous study (Xu, Yue, Lescroart, Biederman, & Kim, 2009) used Gabor-jet scaling and had demonstrated that FFA was as sensitive to a change in the identity of a face as to a change in its viewpoint, when image similarities were scaled to be equal for these two types of variation. The present study was designed to further explore the representation of viewpoint, identity, and more importantly, the expression of faces at different loci in the human face-selective network.

Method

Subjects

Nine subjects (three females) with a mean age of 22 ± 2.5 years participated in the experiment. All subjects reported normal or corrected-to-normal vision and had no known neurological or visual disorders. They gave written informed consent prior to the experiment, in which all procedures and protocols were approved by the Institutional Review Board of the University of Southern California.

Magnetic resonance imaging setup

Subjects were scanned in a 3.0-T Siemens MAGNETOM Trio Scanner equipped with a 12-channel head coil at the Dana and David Dornsife Cognitive Neuroscience Imaging Center at the University of Southern California.

Stimuli

All stimuli in the adaptation experiment were generated by the FaceGen Modeller 3.2 (Singular Inversions, Vancouver, Canada). The face models were chosen to be

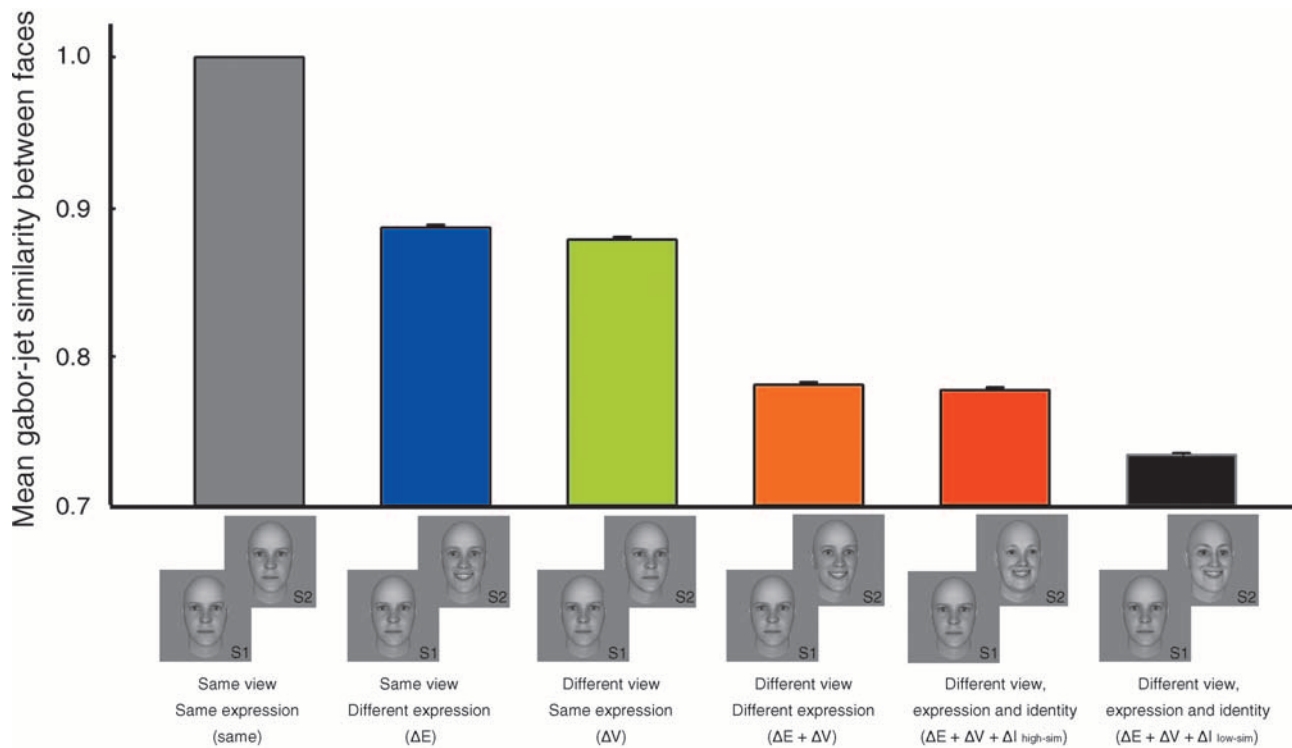


Figure 1. Sample stimuli for the six conditions (here illustrated with the same S1) and mean Gabor-jet similarity values for each condition with error bars indicating the standard error of the mean. 1.0 is the highest possible similarity value which would be the value for identical images. The stimuli in the triple change condition were broken down into a low-similarity half, $\Delta E + \Delta V + \Delta I_{low-sim}$, and a high-similarity half, $\Delta E + \Delta V + \Delta I_{high-sim}$, with the latter matching the similarity of the $\Delta E + \Delta V$ condition.

middle-aged, Caucasian males and females, without hair on a gray background (Figure 1). Identity was varied by modification of both local features (the shape of the eyes, nose, mouth, jaw and cheekbones) and their spatial configuration, for example the distance between the eyes, nose and mouth. The expression of each face could be neutral, happy or angry. The orientations were frontal to approximately 13° to the right to allow equivalent Gabor-jet similarity scaling of orientation and expression changes. The rotation angle was estimated by exporting the model into 3DsMax (Autodesk, California, USA). Finally, all the stimuli were sized to $6^\circ \times 6^\circ$.

The combination of orientation (same vs. different) and expression (same vs. different) between two faces of the same person resulted in four conditions: Same Expression Same View (*Identical*), Expression Change (ΔE), View Change (ΔV), and Expression and View Change ($\Delta E + \Delta V$). In addition, Identity (person) could change, always accompanied by a change in expression and view ($\Delta E + \Delta V + \Delta I$). In this condition, the order of identity, viewpoint and expression for the first and second face was counterbalanced across all possible combinations. There were six persons of the same sex depicted in each of the four blocks. The experimental trials were equally distributed among those five conditions. The mean and the standard error of Gabor-jet similarity values (explained below) between a pair of faces in each trial were 1.00 ± 0

for *Identical*, 0.88 ± 0.001 for both ΔE and ΔV , 0.78 ± 0.002 for the $\Delta E + \Delta V$ condition, and 0.75 ± 0.002 for the $\Delta E + \Delta V + \Delta I$ condition as shown in Figure 1. (We further broke down the stimuli in the $\Delta E + \Delta V + \Delta I$ condition into a low-similarity half $\Delta E + \Delta V + \Delta I_{low}$, 0.73 ± 0.002 and a high-similarity half $\Delta E + \Delta V + \Delta I_{high}$, 0.78 ± 0.002 , with the latter matching the similarity of the $\Delta E + \Delta V$ condition.)

Subjects were instructed to judge whether the images on a given trial were the same person or two different persons by pressing one of two buttons while ignoring possible changes in expression and orientation.

Gabor-jet similarity scaling

The Gabor-jet similarity value (Fiser, Biederman, & Cooper, 1996; Lades et al., 1993) for each pair of stimuli was computed from a 10×10 grid centered on each picture with each node of the grid corresponding to the center of the receptive field of one jet (corresponding to a complex cell hypercolumn). Each jet was composed of 40 Gabor filters (or kernels) at 8 equally spaced orientations (i.e., 45° differences in angle) \times 5 scales, centered on their jet's grid point. The coefficients of the kernels (the magnitude corresponding to an activation value for a complex cell, namely the square-root of the sum of square

of real and imaginary parts of the wavelet transformation coefficients) within each jet were then concatenated to a 4000-element ($100 \text{ jets} \times 40 \text{ kernels}$) vector G : $[g_1, g_2 \dots g_{4000}]$. For any pair of pictures with corresponding jet coefficient vectors G and F , the similarity of the pairs was defined as:

$$\text{Sim}(G, F) = \frac{\sum_{i=1}^{4000} g_i f_i}{\sqrt{\sum_{i=1}^{4000} g_i^2} * \sqrt{\sum_{i=1}^{4000} f_i^2}}, \quad i = 1, 2 \dots 4000. \quad (1)$$

This is a correlation between the vectors (corresponding to the cosine of their angular difference) and yields a similarity value between 0 and 1.00 (for identical images). The code could be downloaded through the link: http://geon.usc.edu/~biederman/GWTgrid_Simple.m.

Scanning parameters

For functional scanning, the BOLD contrast was obtained with a gradient-echo planar imaging (EPI) sequence. The parameters were: $TR = 2 \text{ s}$; $TE = 30 \text{ ms}$; flip angle = 60° ; field of view = $224 \times 224 \text{ mm}$, matrix size = 64×64 , in-plane resolution = 3.5 mm , slice thickness = 3 mm , between-slice gap = 0 . The scanning volume consisted of 32 continuous slices covering most of the lower cortex, including the temporal poles. For anatomical scans, a whole brain three dimensional T1-weighted structural scan was performed with MPRAGE sequences, with the following parameters: $TI = 1100$, $TR = 2.07 \text{ s}$, $TE = 4.1 \text{ ms}$, Flip angle = 12° , 192 sagittal slices, matrix size = 256×256 , voxel resolution = $1 \times 1 \times 1 \text{ mm}$.

Localizer runs

The conventional localizer for face-selective regions uses static images of faces and objects. We instead employed dynamic stimuli of faces and objects, in line with recent studies (Fox, Iaria, & Barton, 2009a; Schultz & Pilz, 2009; Trautmann, Fehr, & Herrmann, 2009) suggesting that such images are more reliable and consistent across subjects in revealing the face-selective network (Gobbini & Haxby, 2007; Haxby et al., 2000). We adopted the same stimuli and protocol as used by Fox et al. (2009a) with their consent. The face videos depicted a change in natural human facial expressions that went from neutral to happy or sad, while the corresponding dynamic object videos showed changes of state, such as a spraying fountain or a spinning wheel (for details see Fox et al., 2009a).

In the blocked localizer runs, participants performed a one-back task by pressing a button when the current clip was identical to the previous one. Identical fixation blocks started and ended the session and were interleaved with image blocks, with all blocks lasting 12 seconds. Eight blocks of each image category (object or face) were presented in a counterbalanced order. Each image block consisted of six video-clips (five novel and one repeated) presented centrally for 2,000 ms each, subtending a visual angle of $15^\circ \times 10^\circ$. Each subject underwent two runs of the localizer.

Event-related adaptation design

Four fast event-related fMRI-adaptation scans were used to test the sensitivity of FFA to changes in facial expression and/or viewpoint. In a given scan, a subject viewed a sequential pair (S1, S2) of faces on each trial, and responded on an MRI-compatible button box to indicate whether or not the two faces depicted the same or two different person. The duration for each trial was 2 seconds. S1 was presented for 300 ms, followed by a 400 ms blank screen with fixation, and then S2 was presented for 300 ms followed by a 1 s blank during which the subject responded. The timing parameters were the same as those used by Kourtzi and Kanwisher (2000) and Winston et al. (2004). No feedback was provided in the actual scanning sessions. Each run had a total duration of 7 minutes 36 seconds, consisting of 218 trials. There was an initial ten seconds fixation period with a black dot centered on a gray background screen to compensate for the initial magnetic field inhomogeneity and a final 10 s fixation period to accommodate for the lag of the hemodynamic response at the end of the scan session. Before going into the scanner, subjects were given 108 practice trials, using a different set of stimuli. During the practice trials, feedback was provided for incorrect responses, in which a red dot appeared when the subject's response was in error, or a yellow dot appeared when the subject's response missed the response interval.

Each run was composed of 36 trials per Expression-View condition in which images of the same person were depicted in S1 and S2, plus 36 target trials in which two different people, differing in both expression and view, were depicted in a trial, along with 38 non-trials in which a blank screen was presented throughout the trial. For each condition, the first face (S1) was centered on the screen, and the second face (S2) was randomly shifted 0.5° in a vertical or horizontal direction to avoid retinotopic adaptation. The trials was arranged such that the history of the two preceding trials for each trial consisted of equal numbers of all the conditions including the five experimental conditions as well as fixation trials. The jittering (Fox, Moon, Iaria, & Barton, 2009b) was achieved by randomized presentation of fixation trials

throughout the run. Subjects underwent four runs of this adaptation experiment, with six different individuals depicted in each run.

Data analysis

The imaging data were analyzed with Brainvoyager QX (Brain Innovation BV, Maastricht, Netherlands). All data from the 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. A 4 mm Gaussian kernel was also used in the spatial smoothing of the functional images prepared for ROI analysis in face-selective areas. Each subject's preprocessed image was then coregistered with their same-session, high-resolution anatomical scan. Then subject's anatomical scan was each transformed into Talairach coordinates. Finally, using the above transformation parameters, the functional image was transformed to Talairach coordinates as well. All statistical analyses were performed on the transformed data.

Face-selective regions were defined as regions with greater activation to dynamic faces than to dynamic objects. The threshold of voxel activation was set at Bonferroni corrected $p < 0.05$ for each subject, for the contrast of faces minus objects. The threshold of activation extension was set at 30 continuous voxels in unit of mm^3 .

For the event-related experimental scans, a deconvolution analysis was performed on each subject's localizer-defined ROI to estimate the time course of the BOLD response, for each trial type. The trials with error response in each condition were modeled as covariate of non-interests, respectively. Deconvolution was computed by having ten 2 s shifted versions of the indicator function for each trial type and response type (correct or incorrect) as the regressor in a fixed-effect general linear model. The percent signal change was computed as the beta values for each regressor, divided by the mean activation values (value of beta zero in the general linear model) of the whole ROI. Finally the peak amplitudes of the signal change, were computed for each condition (Identical, ΔE , ΔV , $\Delta E + \Delta V$ and $\Delta E + \Delta V + \Delta I$, in which Δ represents change along the specific dimension, such as Expression, Viewpoint and Identity) in each ROI. Generally, the peak of the BOLD curve occurred at the 4th TR point (6 seconds from the onset of the stimulus). There were very few occasions when it peaked at the 3rd TR in certain conditions and certain subjects. Considering the group means also clearly showed a peak at the 4th TR, we always took the 4th TR for the analysis.

A 2×2 (Expression [same/different] \times Viewpoint [same/different]) repeated measures analysis of variance (ANOVA) was performed on the peak signal change on

correct responses of each condition. In addition, the $\Delta E + \Delta V + \Delta I$ condition was compared with the double change conditions $\Delta E + \Delta V$ in a paired t-test. Reaction times (RTs) and response accuracy were analyzed in a similar fashion.

Results

Neuroimaging results

Region of interest localization

Figure 2 shows the activation pattern for the contrast of dynamic images of facial expression compared with dynamic images of objects of one subject, superimposed on the his anatomical image in Talairach coordinates. The face-selective ROIs were consistently found in the bilateral occipital-temporal cortex and the superior temporal sulcus across subjects, and were consistent with those in previous studies (Fang et al., 2007b; Grill-Spector, Knouf, & Kanwisher, 2004). Table 1 shows the average Talairach coordinates and number of voxels within each ROI from the dynamic localizer.

Analysis of event-related BOLD responses within each ROIs

A three-way ANOVA of percent BOLD signal change in response to Viewpoint (same-different) \times Expression (same-different) \times Hemisphere (left-right) performed separately for each of the three ROIs (FFA, OFA, and

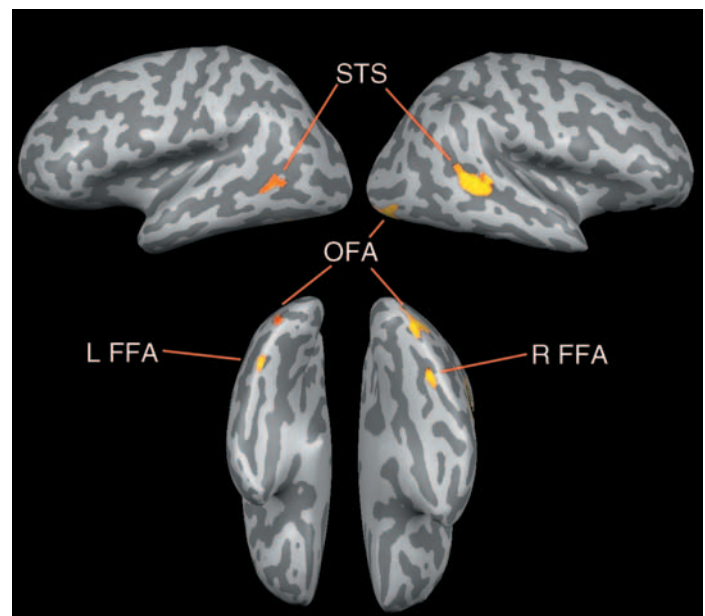


Figure 2. The activation of face-selective ROIs in a typical subject's Talairach normalized brain, for the contrast of face minus objects, with a threshold of $p < 0.05$, Bonferroni corrected.

ROI	X	Y	Z	t-value of Peak Voxel \pm SEM	Mean No. of Voxels \pm SEM	No. of subjects ($n = 9$ max)
rFFA	35 ± 1	-51 ± 3	-16 ± 1	11.4 ± 1.11	766 ± 122	8
lFFA	-40 ± 2	-52 ± 3	-16 ± 1	10.4 ± 1.13	499 ± 110	8
rOFA	30 ± 2	-83 ± 2	-7 ± 2	8.5 ± 1.15	1190 ± 400	7
lOFA	-36 ± 3	-86 ± 2	-7 ± 4	6.5 ± 0.80	250 ± 84	7
rSTS	48 ± 2	-42 ± 3	6 ± 2	11.2 ± 1.13	2340 ± 240	9
lSTS	-55 ± 3	-47 ± 4	9 ± 1	8.1 ± 1.15	1135 ± 223	9

Table 1. The average Talaraich coordinates (Mean \pm Standard Error of Mean: SEM) for each of the designated areas, the t-value of voxel in each ROI, the mean number of voxels of each ROI defined by the dynamic localizer, and the number of subjects in whom the ROIs could be localized. Each voxel extended 1 mm^3 .

STS), did not reveal a reliable main effect of Hemisphere. There was a trend for higher activation of the right hemisphere in FFA (consistent with previous findings), but this effect fell short of significance, $F(1, 7) = 3.25$, $p = 0.12$. None of the hemispheric differences interacted with viewpoint or expression in any of the ROIs, all $F_s < 1$. We therefore collapsed the data from both hemispheres across the Viewpoint/Expression/Identity condition and ran a repeated $2(\text{Viewpoint}) \times 2(\text{Expression})$ ANOVA for each ROI.

As shown in Figure 3, in FFA, a change of expression produced a greater BOLD response than that to identical faces; $F(1, 7) = 16.35$, $p < 0.01$, while there was no release from adaptation from a change in viewpoint. The F_s for viewpoint and the interaction of viewpoint and expression were both < 1 . In Bilateral OFA (see Figure 4), neither the change of expression nor the change of viewpoint produced a release from adaptation compared to identical faces (both $F_s < 1$).

The change of identity was not fully crossed with the other two factors (View and Expression) so the comparison of the triple change condition ($\Delta E + \Delta V + \Delta I$) with other conditions covaried with the response differences (since a different response was required), and more importantly, the overall Gabor-jet similarity differences with changes in Identity (which were always associated with changes in Viewpoint and Expression) were lower for the conditions in which identity did not change. Therefore, we broke the $\Delta E + \Delta V + \Delta I$ condition into low-similarity and high-similarity halves, such that the high-similarity half ($\Delta E + \Delta V + \Delta I_{\text{high-sim}}$) was comparable in image-similarity to the double change condition ($\Delta E + \Delta V$). In FFA (Figure 3) the $\Delta E + \Delta V + \Delta I_{\text{high-sim}}$ produced a higher BOLD response than the $\Delta E + \Delta V$ condition (although the paired t-test fell short of significance, $t(7) = 1.92$, $p = 0.09$). The same pattern held for OFA, $t(6) = 1.90$, $p = 0.11$. However, the high and low similarity halves of the $\Delta E + \Delta V + \Delta I$ did not differ in

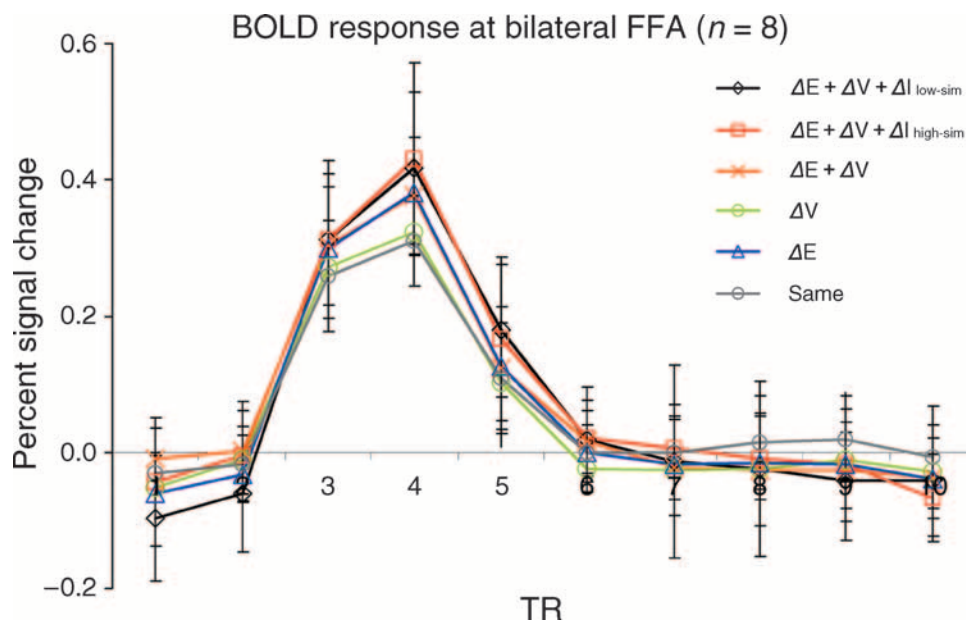


Figure 3. Average event-related BOLD response for 8 subjects in Bilateral FFA.

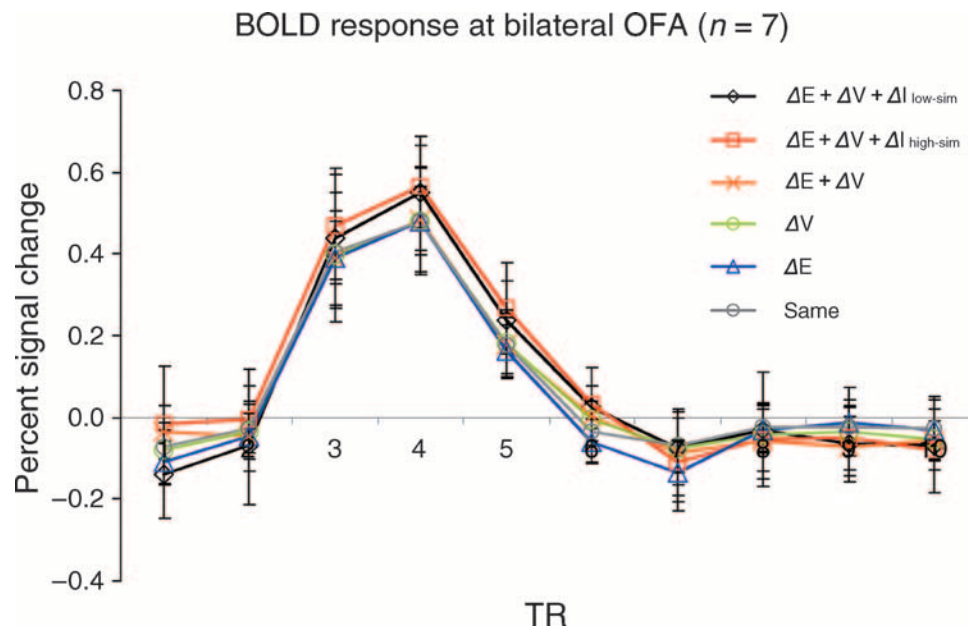


Figure 4. Average event-related BOLD response for 7 subjects in Bilateral OFA.

the magnitude of their mean BOLD response (all $t_s < 1$, $p > 0.5$).

In Bilateral STS, the deconvolution analysis produced a generally flat curve (Figure 5), with no differentiation among the conditions, a picture which was markedly different from the canonical hemodynamic function and the functions observed in FFA and OFA in the present experiment. This weak and non differentiated response in STS was possibly a consequence of our use of computer-generated face stimuli. We nonetheless carried out a statistical analysis similar to those in other ROIs. This

analysis did not reveal any significant effect of either viewpoint or expression, or their interaction in bilateral STS. To further explore the low-responsiveness in STS, on the suggestion of a reviewer, about 700 “top” voxels which showed the strongest face selectivity (relative to their responses to objects) in the dynamic localizer were chosen from only the right STS (which was more responsive to faces than left STS). The selection of these voxels was made by increasing the threshold to a Bonferroni corrected $p < 0.0001$ for the face minus object contrast. The result showed that the highly selective

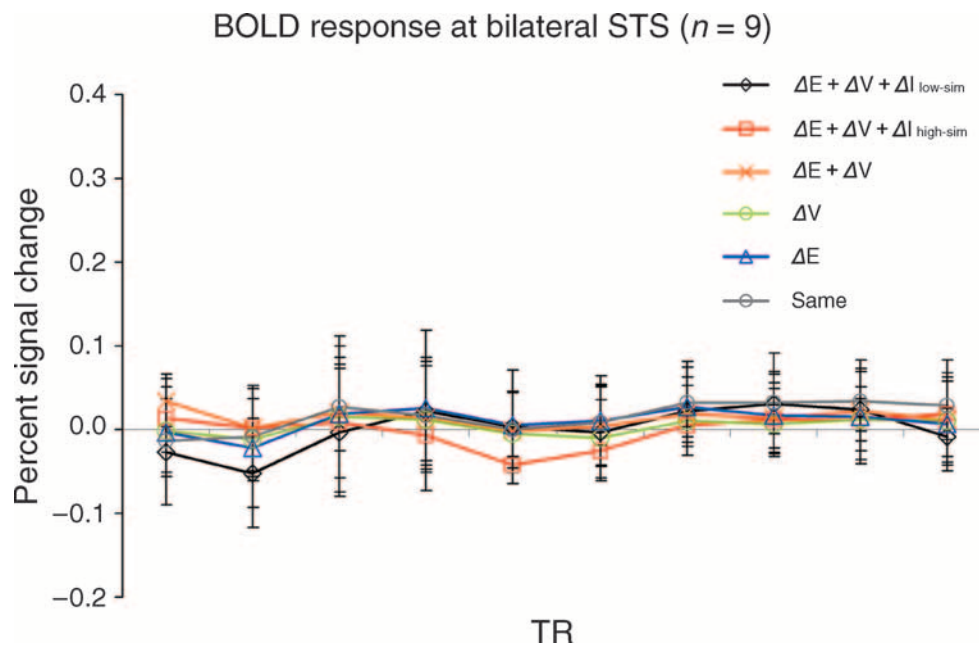


Figure 5. Average event-related BOLD response for 9 subjects in Bilateral STS.

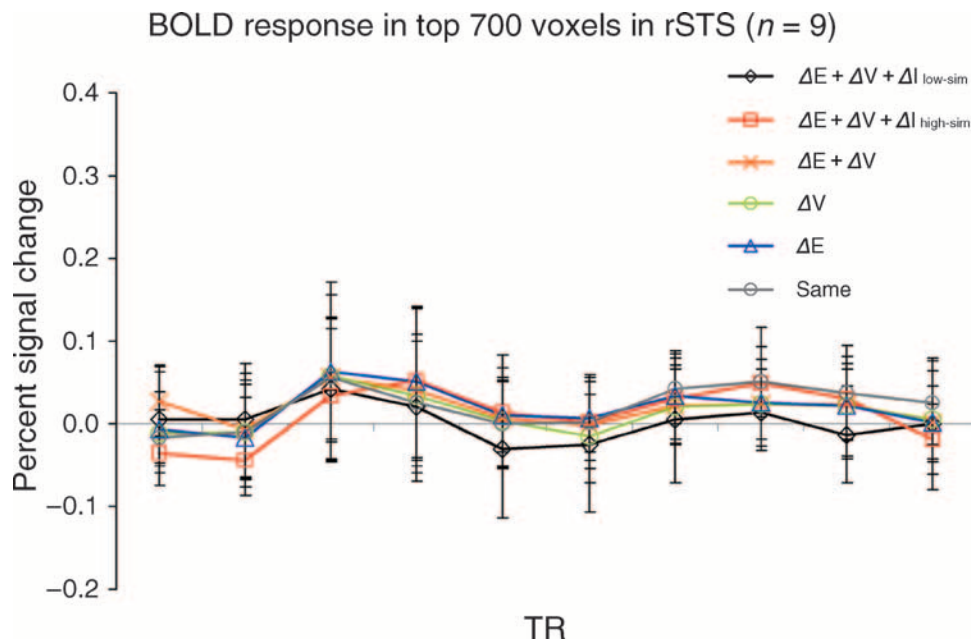


Figure 6. Average event-related BOLD response for 9 subjects in top 700 voxels of right STS.

voxels tended to cluster in the posterior part of STS. These data more closely resemble the expected hemodynamic function (Figure 6), but the overall magnitude was still too weak for a statistical comparison between the conditions.

Behavioral results for the same–different person judgment task

Even though the task did not require the subject to judge view or expression, there were, nonetheless, costs of changes in Expression or View on performance when the faces were from the same person (Table 2). That is, differences in Expression or View caused interference in the judgment of identity with the costs of a change of Expression being markedly greater than the costs associated with a change in viewpoint. Multiple changes increased these costs.

A 2×2 (Expression [same/different] \times Viewpoint [same/different]) repeated measures analysis on the RTs showed significant main effects for both Viewpoint $F(1, 8) = 57.03$, $p < 0.001$, and Expression $F(1, 8) = 232.38$, $p < 0.001$, and an insignificant interaction $F(1, 8) < 1.00$. For error rates, the analysis also revealed a

significant main effect for expression $F(1, 8) = 45.45$, $p < 0.001$, a marginally significant main effect for viewpoint $F(1, 8) = 5.07$, $p = 0.05$, and a significant interaction between the two variables $F(1, 8) = 12.60$, $p < 0.01$.

From one perspective, the interaction in the error rates between the judgment of changeable aspect of faces (i.e., expression and view) and an invariant aspect (i.e., identity) suggests that the processing of the two types of information was not fully segregated. Given that the faces in the triple change condition, $\Delta E + \Delta V + \Delta I$, had even lower similarity than those in the double change condition, $\Delta E + \Delta V$, it is possible that it was the physical difference rather than the identity change that produced higher behavioral costs in the former condition. A comparison between $(\Delta E + \Delta V + \Delta I_{high-sim})$ and the $(\Delta E + \Delta V)$ condition, which had equivalent similarities, showed that it was the change in identity (and/or response) that resulted in reliably longer RTs and higher error rates rather than the greater image change, per se. Paired t-tests for RTs $t(8) = 3.92$, $p < 0.05$ and for error rates $t(49) = 3.38$, $p < 0.01$ were both reliable, even though the two conditions did not differ significantly in Gabor-jet similarities. On the other hand, the $\Delta E + \Delta V + \Delta I_{high-sim}$ and $\Delta E + \Delta V + \Delta I_{low-sim}$ did lead to differences in error rates $t(8) = -2.83$, $p < 0.02$, but not in RTs $t(8) < 1.00$.

	Identical	ΔE	ΔV	$\Delta E + \Delta V$	$\Delta E + \Delta V + \Delta I_{high-sim}$	$\Delta E + \Delta V + \Delta I_{low-sim}$
RTs (ms)	500	570	522	592	663	650
SEM	20.6	24.1	21.2	21.5	32.8	30.7
%Error	2.7	16.5	2.5	21.1	41.1	35.5
SEM	1.05	2.60	0.58	2.64	3.71	3.96

Table 2. Behavioral results (RTs and % Error) as a function of Expression, Viewpoint and Identity change ($n = 9$).

Garner interference (Pomerantz & Garner, 1973) is the assessment of whether variation in the values of an irrelevant dimension produces a decrement in the efficiency of processing (speed and/or accuracy) the values of a relevant dimension. Schweinberger and Soukup (1998) investigated whether such interference would be observed between facial identity, expression and speech. They found that processing of identity was independent of expression and speech, but not the reverse, i.e., the Garner interference was asymmetrical. This result is not surprising because the two individuals depicted in the stimuli had different hair styles which likely provided a salient and highly discriminable cue for identity when compared to the relatively subtle differences when facial expression and speech were varied. The saliency of this cue was reduced in later studies (Baudouin, Martin, Tiberghien, Verlut, & Franck, 2002; Schweinberger, Burton, & Kelly, 1999) that confirmed the previous finding that individuation was independent of emotional expression. But it remains unclear whether that result was due to the higher discriminability between different identities than that between different expressions. The current experiment addressed this problem by matching low-level metric image differences for the ΔE and ΔV conditions as well as the $\Delta E + \Delta V$ and $\Delta E + \Delta V + \Delta I$ conditions. With this control for stimulus-similarity, we found that processing of identity was not completely independent of expression when expression was irrelevant to the task. We note, however, that the Gabor scaling is only sensitive to the metric coding characteristic of V1. It does not reflect the heightened sensitivity that humans have for nonaccidental differences (such as would occur with large differences in hairstyles) relative to metric differences. The highly similar variations in the present experiment were likely all metric.

Consideration of response biases is grounds for caution before accepting the inference of non-independence of judging identity and expression (and orientation). When judging that two individuals are the same person, larger stimulus differences, either because of differences in category or physical similarity, would be expected to increase difficulty. The opposite effect of similarity should be witnessed when the correct response should be that the depicted faces are of different individuals. Here, larger physical image differences should *facilitate* performance in the triple change condition $\Delta E + \Delta V + \Delta I$. That was what was observed: the $\Delta E + \Delta V + \Delta I_{low-sim}$ condition had lower error rates and shorter RTs than the $\Delta E + \Delta V + \Delta I_{high-sim}$ condition.

Discussion

Adaptation to expression

A fast event-related fMRI adaptation experiment revealed that a change in facial expression of the same

person produced a release of adaptation of the BOLD signal in FFA when compared to that for identical faces beyond what would be expected from image dissimilarity alone, as gauged by the much smaller response to orientation changes. This result was consistent with previous reports of expression sensitivity in FFA (Fox et al., 2009b; Kadosh, Henson, Kadosh, Johnson, & Dick, 2010), but was not consistent with Winston et al.'s (2004) finding of a lack of a release from adaptation in FFA when expression was changed between faces. Instead of modeling the similarity between faces as covariates of no interest, as did Winston et al., we matched the similarity of view-changed and expression-changed faces through the Gabor-jet model (Lades et al., 1993). The Gabor-jet model provides a principled framework with which to scale image similarity. The justification for this scaling derives from a match-to-sample task in which the model predicts error rates and reaction times from the similarity between the target face and the distracter with exceedingly high accuracy (both $r_s > 0.95$) (Yue et al., 2010a). With this control for physical similarity, the sensitivity to facial expression was shown to be much greater than that for a face's orientation, as reflected in both behavioral performances and fMRI adaptation in FFA. In a task requiring detection of a change in identity, as in the present experiment, Ganel, Valyear, Goshen-Gottstein, and Goodale (2005) and Kadosh et al. (2010) both reported that a change of expression resulted in higher activation in FFA in a block-design fMRI adaptation paradigm, although image similarity was not controlled in these studies. The present results clearly showed that FFA is sensitive to both expression and identity, as revealed by the significant releases from adaptation from changes in these variables compared to orientation changes, when these three variables had been matched in similarities.

OFA was speculated to be an early stage in the face-processing hierarchy, feeding both FFA and STS, which were postulated to represent variable and invariable aspects of faces, respectively (Haxby et al., 2000). Consistent with this theory was the result of fMRI adaptation experiments in which the adaptation in OFA was released when either expression (Fox et al., 2009b) or identity (Fox et al., 2009b; Rotshtein et al., 2005) was changed between two faces. Our experiment, however, did not detect a release of adaptation in OFA as a result of an expression change. The absence of an expression effect could possibly be attributed to the individuation task performed by our subjects, which required attention to face identity rather than expression. A role of attentional modulation is supported by the finding that greater activation in OFA was found in a task where subjects made expression judgments, rather than a gender-identification task (Gorno-Tempini et al., 2001). More research is needed to clarify the sensitivity to expression in OFA, perhaps by using different tasks with scaling of stimulus similarity.

Adaptation to viewpoint

Whereas the sensitivity to expression was evident in FFA, a change of viewpoint of the same physical magnitude did not produce a reliable release from adaptation. Previous studies using the fMRI adaptation paradigm have demonstrated that FFA was sensitive to the rotation in depth of unfamiliar faces as small as 20° (Andrews & Ewbank, 2004; Ewbank & Andrews, 2008; Fang et al., 2007b; Xu et al., 2009; Yue, Cassidy, Devaney, Holt, & Tootell, 2010b). It is possible that the lack of release of adaptation from viewpoint change in the present study was due to the relatively small rotation angle of 13°, necessitated by the need to equate the image similarity values for orientation and expression. As shown in Figure 1, the mean Gabor-jet similarity between same-identity-different-viewpoint faces is about 0.9, while in our previous study (Xu et al., 2009) it is about 0.8 (as a result of a 20° rotation in depth). Consequently, it is possible that the magnitude of release from adaptation is nonlinearly related to the degree of rotation (Fang et al., 2007b), with 13° being below the threshold for an effect. A nonlinearity was evident at the upper end of rotation in the results of Fang et al. (2007b), in which the increase to 90° did not produce a greater BOLD signal in FFA than 60°. A second possibility is that there were task differences. In the prior study (Xu et al., 2009) subjects detected a change of the size of a face; here they judged identity. It is possible that the identification task biased the attention away from viewpoint.

We can only speculate as to why we found only a flat hemodynamic function across all conditions in bilateral posterior STS. It is possible that the voxels selected by our localizer, which contrasted the fMRI response to dynamic images of faces with those of dynamic images of objects, is less sensitive to the discrete static facial stimuli employed in the event-related adaptation runs (Fox et al., 2009a; Trautmann et al., 2009). Indeed the face patch discovered in monkeys' STS (O'Toole, Roark, & Abdi, 2002) is also sensitive to biological motion, such as that produced by body, hand and mouth movements. Our stimulus presentation procedure, where the second presentation of a face was always translated by 0.5° in all conditions, might have obscured an implied motion effect generated by the rotation of faces. Fang et al. (2007b) and Winston et al. (2004) did manage to obtain adaptation effects in STS with photographs so their faces were more realistic than those in the present experiment. However, Fox et al. (2009b) also used photographs and reported, nonetheless, STS adaptation of relatively small amplitude.

Adaptation to identity

The present study found that a change in identity produced an enhanced BOLD response over that produced

by changes in expression and viewpoint in both FFA and OFA. This response could not be attributable to physical changes in the image, since the effect was present even when Gabor similarity of the $\Delta E + \Delta V + \Delta I_{high-sim}$ condition was equivalent to the $\Delta E + \Delta V$ condition and was, therefore, an effect attributable to a change in category, rather than to a physical image change. In line with this result are prior fMRI adaptation studies that showed a release of BOLD adaptation in both OFA and FFA when subjects perceived a change in the identity between two faces (Gilaie-Dotan & Malach, 2007), for example, different degrees of morph between Marilyn Monroe and Margaret Thatcher (Rotshtein et al., 2005).

Patients with lesions in FFA and/or OFA often exhibit deficits in face individuation. That a lesion to OFA can be sufficient to produce prosopagnosia is documented with P.S. (Rossion et al., 2003; Schiltz et al., 2006), a patient with ablated OFA and intact FFA, who has normal performance in detecting faces but cannot individuate faces. Unlike control subjects (or subjects in the present study), an identity change did not produce a release from adaptation relative to identical faces in FFA for this individual. Transcranial magnetic stimulation (TMS), which temporarily interrupts local neural activity, when applied to OFA, interferes with discrimination of individuals (Pitcher, Walsh, Yovel, & Duchaine, 2007; Pitcher, Charles, Devlin, Walsh, & Duchaine, 2009). Together, these results suggest that an intact network that includes both OFA and FFA is necessary for normal face processing for individuation (Rossion et al., 2003).

Conclusion

Bruce and Young (1986) and Haxby et al. (2000) proposed that facial identity and expression are functionally independent and neuronally represented in distinct brain areas. Our results challenge the idea of (complete) anatomical separation of the neural representation of facial expression and identity (Calder & Young, 2005). Rather the fMRI adaptation pattern in the present investigation suggests that FFA is sensitive to both facial identity and expression and OFA is sensitive only to an identity change. The lack of sensitivity of OFA to changes in viewpoint or expression, is also not in agreement with expectations from Haxby et al.'s (2000) claims that OFA would be sensitive to any shape change of a face. Taken together, our findings suggest a more complex picture in which the processing of changeable and invariant aspects of faces might not be clearly separated.

Acknowledgments

The authors thank Jiancheng Zhuang for assistance in the scanning and Christopher Fox for authorizing the use of his videos of dynamic face expression and moving objects, Xiaomin Yue and Bosco Tjan and two anonymous reviewers for their insightful comments on the manuscript. This study is supported by NSF 0420794, 0531177, and 0617699 to IB. The authors declare that they have no competing financial interests.

Commercial relationships: none.

Corresponding author: Xiaokun Xu.

Email: xiaokunx@usc.edu.

Address: University of Southern California, 3641 Watt Way, Hedco Neurosciences Building, Room 316, Los Angeles, CA 90089-2520, USA.

Footnote

¹It is not difficult to construct cases where the Gabor-jet measure of similarity reflects psychophysical similarity that is not reflected in pixel measures of similarity (Yue et al., 2010a, Figure 7) Link: http://geon.usc.edu/~biederman/publications/Yue_et_al_2010.pdf. However, with highly similar faces and other compact shapes, both the pixel and Gabor-jet scaling measures provide similarity measures that are highly correlated with each other and with psychophysical similarity.

References

- Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage*, *23*, 905–913.
- Baudouin, J., Martin, F., Tiberghien, G., Verlut, I., & Franck, N. (2002). Selective attention to facial emotion and identity in schizophrenia. *Neuropsychologia*, *40*, 503–511.
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences*, *352*, 1203–1219.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641–651.
- Campbell, J., & Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Research*, *49*, 1532–1540.
- Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H., & Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *Journal of Neurophysiology*, *91*, 358–371.
- Ellamil, M., Susskind, J. M., & Anderson, A. M. (2008). Examinations of identity invariance in facial expression adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 273–281.
- Ewbank, M. P., & Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *Neuroimage*, *40*, 1857–1870.
- Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron*, *45*, 793–800.
- Fang, F., Ijichi, K., & He, S. (2007a). Transfer of the face viewpoint aftereffect from adaptation to different and inverted faces. *Journal of Vision*, *7*(13):6, 1–9, <http://www.journalofvision.org/content/7/13/6>, doi:10.1167/7.13.6. [PubMed] [Article]
- Fang, F., Murray, S. O., & He, S. (2007b). Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. *Cerebral Cortex*, *17*, 1402–1411.
- Fiser, J., Biederman, I., & Cooper, E. E. (1996). To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition? *Spatial Vision*, *10*, 237–271.
- Fox, C. J., & Barton, J. J. (2008). It doesn't matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. *Journal of Vision*, *8*(3):11, 1–13, <http://www.journalofvision.org/content/8/3/11>, doi:10.1167/8.3.11. [PubMed] [Article]
- Fox, C. J., Iaria, G., & Barton, J. S. (2009a). Defining the face-processing network: Optimization of functional localizer in fMRI. *Human Brain Mapping*, *30*, 1637–1651.
- Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. (2009b). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage*, *44*, 569–580.
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*, 1187–1196.

- Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia*, *43*, 1645–1654.
- Gilaie-Dotan, S., & Malach, R. (2007). Sub-exemplar shape tuning in human face-related areas. *Cerebral Cortex*, *17*, 325–338.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*, 32–41.
- Gorno-Tempini, M. L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., et al. (2001). Explicit and incidental facial expression processing: An fMRI study. *Neuroimage*, *14*, 465–473.
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*, 187–203.
- Hasselmo, M., et al. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, *32*, 203–218.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, *10*, 14–23.
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*, 80–84.
- Kadosh, K. C., Henson, R., Kadosh, R. C., Johnson, M. H., & Dick, F. (2010). Task-dependent activation of face-sensitive cortex: An fMRI adaptation study. *Journal of Cognitive Neuroscience*, *22*, 903–917.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, *20*, 3310–3318.
- Krekelberg, B., Boynton, G. M., & van Wezel, R. J. (2006). Adaptation: From single cells to BOLD signals. *Trends in Neurosciences*, *29*, 250–256.
- 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.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.
- O’Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Science*, *6*, 261–266.
- Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. C. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology*, *19*, 319–324.
- Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. C. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, *17*, 1568–1573.
- Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics*, *14*, 565–569.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*, 5205–5215.
- Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*, 1–15.
- Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, *8*, 107–113.
- Schiltz, C., Sorger, B., Caldara, R., Ahmed, F., Mayer, E., Goebel, R., et al. (2006). Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. *Cerebral Cortex*, *16*, 574–586.
- Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, *194*, 465–475.
- Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, *61*, 1102–1115.
- Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1748–1765.

- Smith, M. L., Fries, P., Gosselin, F., Goebel, R., & Schyns, P. G. (2009). Inverse mapping the neuronal substrates of face categorizations. *Cerebral Cortex*, *19*, 2428–2438.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.
- Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotion in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research*, *1284*, 100–115.
- Vida, M. D., & Mondloch, C. J. (2009). Children's representations of facial expression and identity: Identity-contingent expression aftereffects. *Journal of Experimental Child Psychology*, *104*, 326–345.
- Winston, J. S., Henson, R. N., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*, 1830–1839.
- 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.
- Yue, X., Biederman, I., Mangini, M., & Malsburg, C. (2010a). Near perfect prediction of the psychophysical discriminability of faces and non-face complex shapes by image-based measures. Manuscript submitted for publication.
- Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2010b). Lower-level stimulus features strongly influence responses in the fusiform face area. *Cerebral Cortex*, in press.