fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception

JS Winston¹, RNA Henson¹,², MR Fine-Goulden¹, RJ Dolan¹

1. Wellcome Department of Imaging Neuroscience, University College London
2. Institute of Cognitive Neuroscience, University College London

Running head: Representations of facial identity and expression

Figures: 5
Tables: 1

Correspondence should be addressed to:

Joel Winston
Wellcome Department of Imaging Neuroscience
12 Queen Square
London WC1N 3BG
UK

Telephone: +44 (0)20 7833 7477
Fax: +44 (0)20 7813 1420
e-mail: j.winston@fil.ion.ucl.ac.uk

Copyright © 2004 by the American Physiological Society.
Abstract

The distributed model of face processing proposes an anatomical dissociation between brain regions that encode invariant aspects of faces, such as identity, and those that encode changeable aspects of faces, such as expression. We tested for a neuroanatomical dissociation for identity and expression in face perception using a functional magnetic resonance imaging (fMRI) adaptation paradigm. Repeating identity across face pairs led to reduced fMRI signal in fusiform cortex and posterior superior temporal sulcus (STS), whereas repeating emotional expression across pairs led to reduced signal in a more anterior region of STS. These results provide neuroanatomical evidence for the distributed model of face processing, and highlight a dissociation within right STS between a caudal segment coding identity and a more rostral region coding emotional expression.
Introduction

Models of face perception propose a dissociation between the representation of identity and other aspects of human faces, for example emotional expression (Bruce and Young 1986). The “distributed model” (Haxby et al. 2000) posits an anatomical basis for the stages of face perception hypothesised by Bruce and Young (Bruce and Young 1986). In this model, fusiform cortex – a region known to be activated during simple face perception (Sergent et al. 1992; McCarthy et al. 1999; Puce et al. 1995; Kanwisher et al. 1997) - represents the identity of a perceived face, whereas superior temporal sulcus (STS) represents “changeable aspects” of the face, such as eye gaze and facial expression. Single unit recordings in monkeys (Hasselmo et al. 1989) and studies of human patients with discrete brain lesions (Adolphs et al. 1996; Young et al. 1993) support this model.

Functional imaging studies comparing familiar versus unfamiliar faces also support a role for fusiform cortex in representing facial identity (Henson et al. 2000; Henson et al. 2003; George et al. 1999). However, differences between familiar and unfamiliar faces, such as increased attention to familiar faces (Wojciulik et al. 1998), may confound these findings. Support for dissociable roles of fusiform and STS comes from studies that directed attention to different aspects of face stimuli (Winston et al. 2003a; Narumoto et al. 2001; Sergent et al. 1992; Sergent et al. 1994; Hoffman and Haxby 2000). Nevertheless, it is difficult to draw definitive conclusions about the nature of the representations in these regions using such task manipulations, because differences in activation might reflect
processing involved in directing visual attention to the specific aspects of the face required by each task, rather than representations of those face components themselves.

Functional imaging studies of emotional facial expression have reported data that seem inconsistent with the distributed model, showing enhanced fusiform activity to emotional compared to neutral faces (Breiter et al. 1996; Dolan et al. 1996; Morris et al. 1998; Pessoa et al. 2002; Vuilleumier et al. 2001; Winston et al. 2003a; Winston et al. 2003b; Surguladze et al. 2003). If fusiform cortex is specialised for identity it is unclear why it should show this enhanced response to emotional faces. This effect has been attributed to modulatory effects from amygdala, reflecting enhanced attentional processing associated with emotive stimuli (Dolan 2002), but direct evidence for this proposal is sparse (for exceptions, see Morris et al. 1998 and Pessoa et al. 2002).

fMRI adaptation (fMRI-A) is a technique used to infer regional specialisation with greater specificity than the subtractive methodology used in the above imaging studies. The logic of fMRI-A, outlined previously (Grill-Spector and Malach 2001; Henson 2003; Naccache and Dehaene 2001), can be summarised as follows: if a region contains sub-populations of neurons excited by distinct aspects of stimuli, then when two stimuli are shown sequentially in which one of these aspects is repeated, the firing neurons will habituate, and decreased fMRI signal from that region will be seen compared to when that aspect is not repeated. We report here the use of an fMRI adaptation paradigm to test the hypothesis, derived from the distributed model (Haxby et al. 2000), that fusiform cortex would show adaptation when identity was repeated relative to when it changed,
and STS would show adaptation when an emotional expression was repeated relative to when it changed. Given our factorial design (figure 1b), the prediction was that fusiform cortex should show a significant main effect of identity repetition, STS would show a significant main effect of expression repetitions, but that neither area would show a significant interaction, as the interaction in this experimental design looks for areas showing a co-dependence between identity and expression repetitions.
Methods

We applied the technique of fMRI-A using a 2X2 factorial design (Figure 1) to examine the neural basis for extraction of identity and expression from faces. Faces were presented in pairs where identity and expression of the second face could independently repeat or change with respect to the first face. Such immediate repetition induces robust fMRI adaptation (Epstein et al. 2003; Kourtzi and Kanwisher 2001). We refer to trials in which identity was repeated as “SI” and use “DI” to indicate a change in identity across a face pair. Similarly, trials where expression is held constant or changed are labelled as “SE” or “DE” respectively. Thus, a trial in which identity was unchanged but expression varied is referred to as “SIDE”.

Stimuli

The stimuli were a selection of five male faces from the KDEF database (Lundqvist and Litton, 1998; The Karolinska Directed Emotional Faces; Department of Clinical Neuroscience, Psychology Section, Karolinska Institute). A major advantage of using this database is that it contains two exemplars of each expression for each identity. This allowed the use of different images for the first and second faces in the adaptation session for all trial types, even SISE. The five specific identities (males 13, 14, 16, 23 and 24) were selected on the basis of successful emotion recognition by nine subjects who took part in a pilot study judging the emotions expressed across the entire database. All face stimuli showed head and eye-gaze direction that were forward-facing towards the viewer. Five emotions were used in the study: anger, disgust, fear, happiness and sadness. One
exemplar of each expression (series ‘A’) was nominated as the prime face and the second (series ‘B’) as the second face for the adaptation phase. Stimuli were converted to greyscale and equated for mean luminance in Matlab (The Mathworks, Natick, MA) and cropped to a standardised outline in Photoshop (Adobe, San Jose, CA). Faces for the filler trials were neutral male faces from a variety of sources, and were prepared similarly to emotional faces, as were chairs and female faces (the targets in the localiser and adaptation phases respectively). 100 scrambled faces for the localiser session were derived from the 50 emotional faces and 50 neutral male faces used in that session by permuting the phase of each spatial frequency in the image while maintaining a constant power density spectrum, and then cropping to the same outline.

Subjects

Sixteen right-handed healthy subjects gave informed consent to take part in the study, which was approved by the local ethics committee. Data were rejected from two of these subjects; one because of gross head movement during fMRI scanning, and the second due to an incidental structural abnormality that made normalisation of scans difficult. Age range of included subjects was 18-29 years (mean=23), with seven males. Subjects had normal or corrected-to-normal vision.

fMRI experiment

There were three sessions to the fMRI component of the study. The first was a face localiser session, designed to familiarise participants with face stimuli and provide a
generic map of face- and expression-responsive brain regions. The second and third were adaptation sessions and were split only for subject comfort. In the first ("localiser") session, subjects’ task was chair detection – they responded on an MRI-compatible button box when they saw a chair. The stimuli seen in this session were the 50 emotional faces from critical trials of the second and third ("adaptation") sessions, 50 neutral male faces, 100 scrambled faces and 20 chairs. In the adaptation phase, the task was to press the button when a female face appeared. In this way, trials of interest in all sessions were uncontaminated by motor response and repetition of faces was incidental to the subject’s task. A number of different trial types occurred in the adaptation phase. The four trial types of interest consisted of a pair of emotional faces that could exhibit the same or different identities, crossed with the same or different emotional expression (Figure 1B). In addition, between any pair of trials of interest, one or two “filler trials” occurred to reduce the predictability of repetition. Three types of filler trials were used – “repeated fillers”, in which the face pair consisted of two neutral faces of the same identity; “different fillers”, consisting of a pair of neutral male faces of different identities; and “target trials” containing a neutral male and a neutral female face (which could be either the first or second face of the pair). In total, 80 repeated filler trials, 60 different filler trials and 20 target trials occurred in each of the two sessions. 25 trials of interest occurred for each of the four trial types in each session. Because there is only one way of combining the faces to produce $S_dS_E$ trials, but multiple ways of combining to produce the other trial types, the face pairs used for such trials were counterbalanced across subjects, with the constraint that each first face occurred once in each condition and each second face once in each condition within each session.
**fMRI scanning**

A Siemens 1.5T Sonata system (Siemens, Erlangen, Germany) was used to acquire BOLD contrast weighted echoplanar images (EPIs) for functional scans. Volumes, which consisted of 24 horizontal slices of 2mm thickness with a 1mm gap, were acquired continuously every 2.16s. This sequence was sufficient to obtain coverage from above the corpus callosum to below the inferior temporal lobes, thus including all regions of interest in this study: fusiform, amygdala, STS, inferior frontal cortex, and orbitofrontal cortex (figure 2). In-plane resolution was 3 X 3mm. The first six volumes were discarded to allow for T1 equilibration effects. Subsequent to functional scans, a T1 weighted structural image (1 X 1 X 1mm resolution) was acquired for coregistration and display of the functional data. Because of the difficulty of normalising limited field-of-view EPIs, we additionally acquired whole brain EPIs in each subject for improved normalisation.

**Spatial pre-processing**

fMRI data were spatially preprocessed using SPM2 (Wellcome Department of Imaging Neuroscience; http://www.fil.ion.ucl.ac.uk/spm). The volumes were coregistered (Friston et al. 1995a) and then normalised to an EPI template corresponding to the MNI reference brain in Talairach space. The normalisation parameters for each subject’s EPIs were obtained by normalising the whole brain EPIs acquired after the experimental session. The limited field of view EPIs were coregistered with the raw whole brain images (Collignon et al. 1995) and normalised by applying the parameters calculated for
normalisation of the whole brain images. Normalised images were smoothed using an 8mm Gaussian kernel to account for residual inter-subject differences, and to allow statistical inference using Gaussian random field theory.

Data analysis

Data analysis used SPM2, applying a mass univariate general linear model (GLM) (Friston et al. 1995b). First, delta functions were constructed corresponding to the onset of each event type and for button presses in the adaptation phase, to accommodate false alarms. These delta functions were convolved with a synthetic haemodynamic response function to create regressors for the subsequent GLM. Also included in the model were six movement parameters estimated by the realignment stage, regressors representing session effects, and for the adaptation sessions, three regressors of no interest corresponding to the potential confound of similarity between face pairs (see below). Serial autocorrelations were modelled using an AR(1) process and the data were high pass filtered at 1/128Hz. Linear contrasts pertaining to the main effects and interaction of the factorial design were calculated. Consistent effects across subjects were tested using the resultant contrast images in one-sample t-tests (conforming to a “random effects” model). The model for the localiser session included separate regressors for the distinct facial expressions. As no significant (p<0.05 uncorrected) interactions with emotion type were detected in regions of interest (see also Winston et al. 2003a), we modelled the adaptation sessions without reference to different emotion types. Statistical threshold was set at p<0.05 corrected for multiple comparisons across a small volume of interest, using a mask derived from the localiser session (see below).
**Mask defining regions of interest**

Regions of interest were defined using two statistical results from the localiser session. First we tested for effects of faces>scrambled faces and thresholded the result at p<0.001 uncorrected. We next tested for emotional faces>neutral faces and thresholded this result at p<0.001 uncorrected. The two resulting masks were combined using an OR function, yielding the combined mask (figure 2) that was subsequently used for small volume correction (SVC) (Worsley et al. 1996).

**Visual similarity between face pairs**

It could be argued that changes in identity and expression across a pair of faces do not represent only categorical changes, but also a variation in a continuous spectrum of visual similarity. Thus, for example, a pair of faces with the same identity and expression are likely to be more visually similar than a pair where identity is the same but the expression different. In order to avoid this potential confound affecting fMRI data analysis three measures of visual similarity were collected and included in fMRI data analysis as covariates of no interest. The first two were image-based metrics, derived from mathematical analysis of image pairs. These were derived from normalised least squares measures of differences between face pairs. Briefly, faces were normalised for luminance and the second face subtracted from the first. The root mean square of the value at each pixel of this difference image was the difference score for a given image pair. In a refinement of the technique that accounted for minor differences in co-
registration of salient features, faces were allowed to move over one another by up to 25 pixels in either plane and the minimum resulting value taken as the difference score (Vogels et al. 2001). The third measure adopted derived from an independent group of subjects (see below) who each saw 200 face pairs (50 of each trial type) presented with the same parameters as the imaging component of the study and rated each pair for visual similarity. The three ratings were in good agreement (Table 1). All three were included in the main statistical model as regressors of no interest by generating a design matrix in SPM whereby all events of interest were modelled as one trial type that was modulated parametrically by expansions to model the three similarity confounds. The columns pertaining to similarity confounds were then extracted and utilised in the model described above.

**Analysis of eye-tracking data**

fMRI differences in regions such fusiform cortex could be attributable to variations in visual attention with trial type (Wojciulik et al. 1998) and differences in emotionally-responsive regions such as amygdala would be similarly attributable to variations in arousal (Critchley et al, 2002). To explore whether such differences might exist we used online eye-tracking. Data were acquired during scanning for a majority of subjects using an ASL504LRO eye-tracker (Applied Science Laboratories, Bedford, MA). Specifically, accurate pupillometry was achieved in nine subjects during the scanning session and accurate eye-gaze tracking in eight. Pupillometry data was analysed by defining a window of 1.2s after the second face and measuring the minimum, maximum and mean pupil diameter during averaged traces (low-pass filtered at 7.5Hz and baselined for the
onset of the second face) from this window for each subject. These three measures were entered into separate 2X2 ANOVAs. Eye-gaze direction was also assessed using a summary statistic approach. For each of the four critical trial types spatial maps of eye-gaze density were constructed. Each of these maps was compared to the mean map, and difference images constructed. The root mean squares of the density difference values for these latter maps were entered into a 2X2 ANOVA.

Control data on explicit tasks for identity/expression detection

An important consideration is the possibility that subjects might not have noticed repetition of identity or expression of faces within each pair. Furthermore, if a change in one dimension (e.g., the expression of the faces) affected subjects’ ability to detect repetition of the other dimension (e.g., the identity of the faces), then interactions between the two dimensions detected by fMRI would be difficult to interpret (in that a decrease in the fusiform response for $S_iD_E$ trials relative to $S_iS_E$ trials, for example, could simply reflect a reduction in the number of trials in which subjects realised it was the same identity). Control behavioural experiments were conducted to test these possibilities. An independent group of 16 subjects (age range=22-36, mean age=28.5; 11 males; 2 left-handers) completed three behavioural tasks, using identical procedural parameters to those in the imaging study. In the first task, they rated pairs of faces presented for visual similarity using a computer-based visual analogue scale (providing the subjective measures of similarity mentioned above). In the second and third tasks, they classified face pairs as exhibiting either the same or different identity, or the same or different emotional expression, with the order of identity/expression task and the buttons
used for same/different responses counter-balanced over subjects. In all, each subject performed 50 trials of each type for each task. A short (25 trial) practice session preceded each task.
Results

Behavioral data during scanning

Subjects detected (mean ± standard deviation) 99±2% of targets (chairs) in the localiser (false alarm rate = 0.2±0.4%), and 88±7% of targets (female faces) in the adaptation phase (false alarm rate = 7±6%).

The two measures derived from the eye-tracking data from the fMRI scanning sessions showed no significant differences between the four trial types of interest (for gaze direction all $F(1,7)<0.125$, all $p>0.7$; for pupil diameter all $F(1,8)<2.1$, all $p>0.18$, except a marginal trend for an interaction between identity and expression repetitions in the minimum pupil constriction: $F(1,8)= 4.0$, $p=0.08$). These non-significant results suggest that there were no detectable differences in visual attention (indexed by eye-gaze direction) or arousal (indexed by pupil diameter changes) between the different experimental conditions.

Localiser phase

The results of the two T-tests performed on the data from the Localiser Phase, faces > scrambled faces and emotional > neutral faces, are shown in Figure 2. As expected, the activated regions included bilateral fusiform and more posterior occipital areas, as well as STS and amygdala. The two contrasts were then combined to create a mask of regions
that responded to faces and/or facial expression. This mask defined a search region for
the subsequent comparisons in the Adaptation Phase, allowing a principled means for
correcting for multiple comparisons over voxels.

**Adaptation phase: Main effect of repeated identity**

As predicted, a significant main effect of repeated identity (reduced response when the
second face exhibited the same identity as the first; \([D_I S_E + D_D E] > [S_I S_E + S_D D_E]\) was seen
in right fusiform cortex (\(x,y,z = 39, -60, -15\); \(Z = 3.76; p < 0.05\), one-tailed, small-volume
corrected (SVC) for the localiser mask) (Figure 3). In addition, adaptation for repeated
identity was seen in right posterior STS (STSp) (\(x,y,z = 63, -51, 15\); \(Z = 3.73; p < 0.05\)
SVC) (Figure 4). Because an interaction or main effect of repeated expression would
influence interpretation of these results, we examined for such effects at reduced
threshold. In the peak right STSp voxel, a marginally significant main effect of repetition
of expression was evident (\(Z = 1.74; p < 0.05\), one-tailed, uncorrected), whereas in fusiform
no such effect was evident (\(Z = 0.99, p > 0.1\) one-tailed, uncorrected). There was no
evidence for an interaction in the peak fusiform voxel (\(Z = 1.26, p > 0.2\) two-tailed,
uncorrected), nor in the right STSp (\(Z = 0.50, p > 0.2\) two-tailed). Although simple effects
are not conventionally inspected in the absence of a significant interaction, we checked
whether changing expression modulated fusiform responses either in the context of
identity remaining constant or changing. In neither case was there a significant effect
(simple effect of \(D_E\) relative to \(S_E\) with a change in identity, \(p = 0.93\); with identity
constant, \(p = 0.14\)). A more posterior region of right occipital cortex, possibly
corresponding to a face-responsive occipital region (FROR), showed uncorrected repetition effects but failed to withstand correction for multiple comparisons across the volume of the mask (x,y,z = 42,-75,-18; Z = 3.59; p=0.071 SVC) (see figure 3a).

**Adaptation phase: Main effect of repeated expression**

A region of right STS anterior to that described above was shown to be less active when the second face exhibited the same expression as the first face (Figure 5; $S_l D_E + D_l D_E > [S_l S_E + D_l S_E]$). This activation corrected for multiple comparisons across the volume of our mask (x,y,z = 57,-18,-12; Z = 3.80; p<0.05, one-tailed, SVC). Despite the apparent trend towards an interaction in Figure 5B, this was not significant when tested at a lenient statistical threshold (Z=1.06, p>0.2, two-tailed, uncorrected). Similar to the fusiform region, we checked for significant simple effects ($D_I$ relative to $S_I$) opposite to the detected main effect and found no significant differences in the context of expression changing (p=0.14) or being held constant (p=0.97).

**Region-by-condition interaction**

To determine whether the differences in detectable main effects between the mid-STS region and fusiform were significant we undertook a region-by-condition interaction using a 3mm sphere centred on each peak. A significant 3-way interaction obtained (p<0.05) in the direction predicted by the distributed model (adaptation to identity in fusiform and to expression in STS).
Adaptation phase: Interaction

No areas within the mask defining our regions of interest showed an interaction between identity and expression.

Control Experiments

One potential confound in our design is the presence of differences in the visual similarity between face pairs of different trial types. To account for this confound we obtained mean subjective and objective similarity measures for the four different trial types (Table 1). The subjective data were obtained from an additional behavioural experiment (see Methods). All three measures showed significant differences between the four trial types (all F(1,13)>190, all p<0.001). $S_iS_E$ pairs were more similar than the other three types, despite our use of different images in this condition. Unsurprisingly, and consistent with the concept of identity as an invariant feature of the face, trials with same identity had greater similarity than trials with different identity. To account for these differences, we included all three measures as covariates of no interest in the analysis of the fMRI data (see Methods), which removed any linear contribution of similarity to the above fMRI findings. A random effects analysis of the contribution of these regressors to the model (using an F-contrast spanning the three regressors in an ANOVA model) suggested that they were explaining effects in visual regions, though not within the mask used for SVC (e.g. peaks at: $x,y,z = -51,-60,-6$, $Z = 3.76$; $x,y,z = 9,-78,9$; $Z = 3.46$; $x,y,z = 54,-45,-15$, $Z = 3.39$; $x,y,z = 30,-60,-15$, $Z = 3.38$; all p<0.001 uncorrected).
An additional concern, noted above, is that subjects might not notice repetitions of identity or expression, or that the presence of repetition in one dimension would affect behaviour to the other dimension. Data from a behavioural experiments on a separate subject cohort (see Methods) showed that the mean accuracy in an identity discrimination task was 88% for trials when expression was held constant and 83% for trials when expression changed (paired t-test: \( Z=3.78, p<0.001 \)). This was paralleled by slower reaction times (RTs) when judging identity in the context of expression changes (806ms vs 773ms, \( Z=3.33, p<0.001 \)). Mean accuracy for the emotion discrimination task was 87% when identity was unchanged across the face pair and 80% when identity changed (\( Z=3.53, p<0.001 \)). Again, RTs were slower on trials where the task-irrelevant dimension (identity) changed, compared to being held constant (888ms vs 848ms, \( Z=4.41, p<0.001 \)). These data demonstrate that people’s ability to detect repetition of identity or expression with these stimuli was generally high. The data also suggest that changes in one dimension do affect sensitivity to repetition of the other dimension. This behavioural interaction does not, however, confound our findings of two orthogonal main effects in the fusiform/STSp and mid-STS regions.
Discussion

In this study we used event-related fMRI-adaptation to identify the neuroanatomical basis for coding different aspects of faces, specifically identity and expression, in the human brain. By presenting pairs of faces in which the identity and emotional expression of a second face could accord or vary with respect to the first, we demonstrated that discrete brain regions show a reduced BOLD signal when a specific dimension was repeated relative to when it changed. Specifically, posterior lateral right fusiform cortex and posterior right STS exhibited adaptation for identity, whereas right mid-STS showed adaptation for emotional expression. These differences do not relate to any obvious measure of visual similarity between faces in each pair, given that we covaried out both objective and subjective measures of similarity. In addition, we found no evidence that the effects we observed could be attributed to differences in eye-movement or arousal. Control data showed that subjects’ explicit ability to detect changes in identity or in expression was generally high. Although performance was reduced when the other dimension changed, which might confound any interaction between identity and expression on the levels of adaptation, this observation cannot explain the simultaneous finding of two orthogonal main effects in the imaging data.

In the distributed model of face processing (Haxby et al. 2000), a dissociation is posited between processing of invariant and changeable aspects of faces. Specifically, it is suggested that invariant features are coded in ventral occipital and temporal cortex in the lateral fusiform region (also known as the “face area” (Kanwisher et al. 1997), whereas
changeable aspects are coded by right STS. Our data broadly support this model. Within the framework of fMRI-A, our demonstration of a main effect of repeated identity in right fusiform cortex indicates this region represents identity, an invariant aspect of human faces. Although previous studies have shown repetition decreases to faces in fusiform cortex (Henson et al. 2000; Henson et al. 2003; George et al. 1999; Gauthier et al. 2000), to our knowledge the present experiment is the first to show repetition effects in fusiform cortex across dramatically different views of the same identity (i.e., with different expressions). In our view, this finding is important because it suggests that face representations in this region encode not just a specific visual image but a more abstract representation of facial identity (see also Vuilleumier et al 2003a; Eger et al. 2004).

A consistent finding in neuroimaging studies of emotional face perception is activation of fusiform cortex in perception of emotional relative to neutral faces (Breiter et al. 1996; Dolan et al. 1996; Morris et al. 1998; Pessoa et al. 2002; Vuilleumier et al. 2001; Winston et al. 2003a; Winston et al. 2003b; Surguladze et al. 2003). This has been interpreted as relating to enhanced attentional processing associated with arousing emotional faces relative to non-arousing neutral faces (Dolan 2002). However, an alternative explanation is that this region encodes the emotionality of the face, resulting in enhanced activation when expressive faces are presented. The use of an adaptation paradigm in this study enables us to potentially dissociate between these possibilities. If this region coded for specific expressions, it should have shown adaptation for expression, akin to that for identity. The lack of evidence for adaptation for repeated expressions is consistent with the former interpretation that fusiform modulation is mediated by an amygdala-associated
effect (though we note that this inference is based on a null result). At the very least, it seems that any bottom-up effects of expression in right fusiform cortex are of less importance than those of identity, i.e. it exhibits relative preference for identity processing from faces.

In contrast to right fusiform, a focus in right mid-STS showed a main effect for repetition of emotional expression, with repeated expressions associated with reduced activation relative to differing expressions. This accords with a role for this region in coding the specific emotion expressed in a face. We were surprised by the anterior locus of this activation, which fell at –18 on the anterior-posterior axis. Previous studies concerning facial expression have reported activation in right STS in a more posterior locus (around –35 to –60mm) (Critchley et al. 2000; Iidaka et al. 2001; Narumoto et al. 2001; Winston et al. 2003a). This more anterior locus is, however, within the portion of STS reported as activated in studies of social cues (Allison et al. 2000; Martin and Weisberg 2003; Ojemann et al. 1992; Saxe and Kanwisher 2003). We have additionally checked our previous data for activation in this area and found it was activated in an explicit emotional judgement task relative to a gender judgement task (x,y,z = 52,-16,-18; Z = 3.96; see figure 5A in Winston et al. 2003a). Note also that this region fell within our face localiser mask and by definition is responsive to faces or facial expression.

Posterior STS, like the fusiform, showed adaptation to repeated identity. This is contrary to a previous study that failed to observe repetition effects in this region (Henson et al. 2003), though that study used much longer repetition lags. A role for posterior STS in
processing personal identity is however consistent with a recent human lesion study describing a patient with an infarct in the vicinity of left STS who described novel faces as familiar (Vuilleumier et al. 2003b). Unlike fusiform though, posterior STS showed a trend towards an additive main effect for repeated emotion, implying that its role in face processing may be multifaceted. Intriguingly, in a recent re-analysis of single neuron data from monkeys, Tiberghien and colleagues (Tiberghien et al. 2003) suggest that all facial features contribute to distinguishing identity, whereas only a subset determine facial expression. They hypothesise that as a consequence, inferior temporal regions in monkeys may contain identity-selective neuronal populations, whilst STS might contain populations sensitive to identity and expression. Such a view fits with our demonstration of identity-repetition in posterior STS and sensitivity to expression in posterior and mid-STS. However, with regard to the human lesion literature, the majority of reported prosopagnosic patients described have inferior occipitotemporal rather than lateral temporal lesions (see e.g. Damasio et al. 1990; Wada and Yamamoto 2001), presumably corresponding to fusiform rather than STS (but see Figure 1A in Tranel et al. 1997; see also Rossion et al. 2003). In addition, monkeys with STS lesions appear to have only minor identity discrimination deficits (Heywood and Cowey 1992), and recent evidence from multidimensional scaling analysis of single neuron data from monkey STS and inferior temporal (IT) cortex also suggests that STS is more concerned with analysis of facial view and the code in IT more concerned by facial identity (Eifuku et al. 2004). This apparent discrepancy with our result of posterior STS responsivity to identity might be explained in a number of ways. Firstly, it is possible that activation in this region is epiphenomenal and of no functional consequence for identity recognition. However, it is
also possible that our stimuli tax identity processing across different views of a face, and
the role of STS in processing different views of face stimuli is well known (Perrett et al.
1985; Perrett et al. 1991; Eifuku et al. 2004). In addition, it has been demonstrated that
STS neurons in the macaque monkey process identity, at least in the form of a population
code (Baylis et al. 1985), and there is evidence that single neurons in STS code for the
same identity across different face views and other STS cells code conjunctions of
identity and view (Perrett et al. 1991). It may be the case that the aspect(s) of identity
processing that occur in STS do not commonly lead to complaints of prosopagnosia, or
that tests designed to probe prosopagnosia are relatively insensitive to these aspect(s) of
identity processing. As an unpredicted, though significant activation we would like to
see this effect of adaptation for repetitions of identity across different views replicated
before drawing strong conclusions.

Previous work has implicated other brain regions in processing facial expressions, most
notably the amygdala (Breiter et al. 1996; Morris et al. 1996; Pessoa et al. 2002;
Vuilleumier et al. 2001; Winston et al. 2003a; Whalen et al. 1998). There are a number
of potential reasons why we did not detect adaptation in this region. One possibility is
that amygdala responses are emotion-specific, with greatest responses to fearful faces
(Calder et al. 2001), and thus our collapsing across different expression sub-types may
have obscured emotion-specific responses. Alternatively, the amygdala might code for
facial expression in a different manner from cortical regions such as STS, with a non-
specific code whereby responses are dependent upon the arousal engendered by the
emotion (Critchley, Rothstein & Dolan, unpublished observations). Another possibility
is that an expression-specific amygdala response may be insensitive to adaptation, though this seems unlikely, given positive findings concerning the amygdala and stimulus repetition (Ono and Nishijo 2000; Rotshtein et al. 2001).

Although identity and emotion may be processed by partially dissociable neural pathways, the two pathways are likely to interact in production of behavioural responses. This would appear to be the case for our explicit identity and emotion detection tasks, in which a change in one dimension (identity or emotion) impaired ability to detect changes in the other dimension. In behavioural studies, other authors have also found evidence for the non-independence of identity and emotion processing (Schweinberger and Soukup 1998; Schweinberger et al. 1999; Ganel and Goshen-Gottstein *in press*). A further brain region may be responsible for the integration of distinct aspects of the face, which we failed to detect in this study. A more explicit behavioural task during fMRI may help to clarify this issue in future studies.

One issue that deserves consideration is the meaning of BOLD changes in adaptation paradigms such as this. It has been demonstrated that local field potentials (LFP) correlate with the BOLD signal better than multi- or single-unit activity in the macaque monkey (Logothetis et al. 2001). Thus, a region showing fMRI-adaptation may not be transmitting fewer spikes but may either be showing a reduced afferent input or reduced local processing. This highlights one possible dissociation between fMRI-adaptation and response suppression as recorded in single unit work in monkeys (see Desimone 1996). We do not consider that fMRI experiments based upon adaptation are uniquely
problematic in this regard, but that this is a more general interpretational issue for unifying electrophysiological and fMRI work. See Henson and Rugg (2002) for a more extensive discussion of haemodynamic decreases and response suppression.

In conclusion, we have shown that fusiform cortex demonstrates fMRI adaptation when the identity of a face is repeated, and a region of STS shows adaptation when the emotional expression of a face is repeated. The response profiles of these two regions were significantly different in the directions predicted by the distributed model of face processing (Haxby et al. 2000), and we suggest that our findings are generally consistent with this model. However, an adaptation response in posterior STS to repeated identity suggests that STS may also manifest a degree of functional segregation in face perception.
Acknowledgements

We thank J. O’Doherty and the radiographers at the FIL for assistance with scanning, and P. Bentley, J. Gottfried, J. Kilner, P. Vuilleumier and P. Rotshtein for helpful discussions. This work was carried out under a Wellcome Trust Programme Grant to RJD. RNAH is supported by the Wellcome Trust.
Table

**Table 1: Means (standard deviations) of similarity measures for different event types**

Measures for pairs of stimuli were scaled from 0 to 1 and averaged across the trials used in the fMRI experiment. Higher values represent more similar face pairs. “Computer measure 1” represents the minimum value for the RMS difference between image pairs allowing a 25 pixel displacement in any direction; measure 2 represents the value with no displacement. The human measure results from a cohort of subjects who rated the similarity of face pairs on a visual analogue scale.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>$S_1S_E$ (std. dev.)</th>
<th>$S_1D_E$ (std. dev.)</th>
<th>$D_1S_E$ (std. dev.)</th>
<th>$D_1D_E$ (std. dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computer measure 1</strong></td>
<td>0.86 (0.002)</td>
<td>0.63 (0.013)</td>
<td>0.44 (0.012)</td>
<td>0.38 (0.019)</td>
</tr>
<tr>
<td><strong>Computer measure 2</strong></td>
<td>0.84 (0.003)</td>
<td>0.61 (0.011)</td>
<td>0.44 (0.020)</td>
<td>0.39 (0.014)</td>
</tr>
<tr>
<td><strong>Human measure</strong></td>
<td>0.88 (0.005)</td>
<td>0.53 (0.028)</td>
<td>0.46 (0.014)</td>
<td>0.24 (0.036)</td>
</tr>
</tbody>
</table>
Figure legends

Figure 1: Study design and example stimuli

(a) In the localizer phase, subjects pressed a button when they saw chairs amongst a run of face and scrambled face stimuli. Stimuli were presented for 500ms with a 2,400ms gap during which a static checker-board within a face outline was presented.

(b) 2 X 2 factorial design of adaptation phase. Trials in this phase consisted of a pair of faces which could display the same or different identity and (independently) the same or different emotion. Note that even $S_I S_E$ trials do not consist of identical stimuli, as the stimulus set offers two photographs of each individual displaying each emotion, taken on different days. Five different expressions (anger, disgust, fear, happiness and sadness) and five different identities were used. Note that the interaction in this factorial design does not pertain to regions showing specialisation for identity or expression processing; instead it tests for an adaptation to expression that depends on identity changes (or vice versa).

(c) Example run and timing of four trials from the adaptation phase. Targets were rare (10%) trials containing female faces. Trials of interest were always separated by at least one filler trial consisting of neutral male faces which could either show the same identity or different identities. Similar to the localizer session (part a), but not shown for clarity, a face-outlined static checker-board separated all face stimuli.
**Figure 2: fMRI coverage and results of localizer session**

(a) Sagittal slice from group mean T1 structural image showing approximate location of 24 slices for functional imaging sequence. This sequence yielded coverage of inferior temporal cortex up to superior temporal sulcus, including amygdala, orbitofrontal cortex and fusiform cortices.

(b) Coronal (i) and horizontal (ii) sections of group mean EPI showing mask derived from contrasts of faces>scrambled faces and emotional face>neutral faces. Note that the mask covers bilateral fusiform cortex (FFA), face-responsive occipital region (FROR), superior temporal sulcus (STS), and right amygdala (AMY).

**Figure 3: Fusiform cortex shows fMRI-A for repeated identity**

(a) SPM overlaid on group mean EPI showing activation in posterior occipital and fusiform cortices. Fusiform peak at x,y,z = 39,-60,-15; Z = 3.76. The more posterior activation (FROR) is at x,y,z = 42,-75,-18; Z = 3.59 but does not correct for the volume of the mask. Red = p<0.01 uncorrected, yellow = p<0.001 uncorrected. Results are displayed masked using the results from localizer scan (with the latter thresholded at p<0.05 uncorrected).

(b) Mean response profiles for different event types from peak fusiform coordinates. Data derived from a finite impulse response (FIR) model with 3s time bins. Asterisks represent time points where the relevant main effect (adaptation for same identity trials) is significant at p<0.05 in the FIR model.

(c) Differential effects in peak fusiform voxel using data from the main model for the three contrasts tested (main effect of identity (i), expression (ii) and their interaction (iii)).
Bars represent the mean parameter estimate across subjects of the canonical haemodynamic response function; error bars represent standard error. p-values represent t-tests of the mean difference from zero. t-tests are one-tailed for main effects and two-tailed for the interaction term. Note that the interaction is not a comparison of the two main effects, and therefore would not be significant for an area selective for either identity or expression.

(d) Differential responses in peak fusiform voxel with time. Differential effects derived from 3s time bin FIR model and from fitted responses of main model are shown (FIR datapoints for the main effect of identity are shifted backwards by 0.75s and for the interaction, forwards by 0.75s, for legibility).

**Figure 4: Posterior STS shows fMRI-A for repeated identity**

(a) Posterior STS (x,y,z = 63,-51,15; Z = 3.73) shows greater response to trials with different identities for the second face than repeated identities. This region appears to be around the posterior horizontal segment, though the anatomy is somewhat variable from subject to subject (ii). Display as in figure 3(a). (ii) shows group-result data shown on single subject T1 structurals to aid localisation.

(b) Mean response profiles in peak posterior STS (STSp) voxel, derived from 3s time-bin FIR model. Display as in figure 3(b).

(c) Differential effects in peak STSp voxel using data from the main model for the three contrasts tested. Display as in figure 3(c).

(d) Differential responses in peak STSp voxel with time. Display as in figure 3(d).
Figure 5: Mid-STS shows fMRI-A for repeated emotion

(a) Mid-STS (x, y, z = 57, -18, -12; Z = 3.73) shows greater response to trials with different expressions for the second face than repeated expression. Display as in figure 3(a).

(b) Mean response profiles in peak mid-STS voxel, derived from 3s time-bin FIR model. Display as in figure 3(b), except asterisks now represent significant differences for the main effect of emotion repetition.

(c) Differential effects in peak mid-STS voxel using data from the main model for the three contrasts tested. Display as in figure 3(c).

(d) Differential responses in peak mid-STS voxel with time. Display as in figure 3(d).
References


Figure 1
Figure 2
Figure 3

**A**  

y = -60  

z = -15  

x = 39

**B**  

Mean responses in fusiform peak

<table>
<thead>
<tr>
<th></th>
<th>D_D_E</th>
<th>D_S_E</th>
<th>S_D_E</th>
<th>S_S_E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differential % signal change

Time (seconds)

**C**  

Main effects and interaction in fusiform

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>ii</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effect: Identity</td>
<td>(D_S_E + D_D_E)</td>
<td>(S_S_E + S_D_E)</td>
<td>(S_D_E + D_S_E)</td>
</tr>
<tr>
<td>Main Effect: Expression</td>
<td>(S_D_E + D_D_E)</td>
<td>(S_S_E + D_S_E)</td>
<td>(S_S_E + D_D_E)</td>
</tr>
<tr>
<td>Interaction</td>
<td>p=0.0009</td>
<td>p=0.160</td>
<td>p=0.208</td>
</tr>
</tbody>
</table>

**D**  

Event-related differential fusiform responses

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential % signal change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (seconds)</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Main effect: Identity

Main effect: Expression

Interaction
Figure 4

Mean responses in STSp peak

Main effects and interaction in STSp

Event-related differential STSp responses
Figure 5