Investigating the brain basis of facial expression perception using multi-voxel pattern analysis

Martin Wegrzyna,b,*, Marcel Riehlea, Kirsten Labuddac, Friedrich Woermanncc, Florian Baumgartnerd, Stefan Pollmannd, Christian G. Bienc and Johanna Kisslera,b

a Department of Psychology, University of Bielefeld, Bielefeld, Germany
b Center of Excellence Cognitive Interaction Technology (CITEC), University of Bielefeld, Bielefeld, Germany
c Epilepsy Centre Bethel, Krankenhaus Mara, Bielefeld, Germany
d Department of Psychology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

Article info
Article history:
Received 17 July 2014
Reviewed 1 September 2014
Revised 6 November 2014
Accepted 1 May 2015
Action editor David Carey
Published online 14 May 2015

Keywords:
Emotion
Face perception
Facial expressions
fMRI
MVPA

Abstract
Humans can readily decode emotion expressions from faces and perceive them in a categorical manner. The model by Haxby and colleagues proposes a number of different brain regions with each taking over specific roles in face processing. One key question is how these regions directly compare to one another in successfully discriminating between various emotional facial expressions.

To address this issue, we compared the predictive accuracy of all key regions from the Haxby model using multi-voxel pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data. Regions of interest were extracted using independent meta-analytical data. Participants viewed four classes of facial expressions (happy, angry, fearful and neutral) in an event-related fMRI design, while performing an orthogonal gender recognition task.

Activity in all regions allowed for robust above-chance predictions. When directly comparing the regions to one another, fusiform gyrus and superior temporal sulcus (STS) showed highest accuracies.

These results underscore the role of the fusiform gyrus as a key region in perception of facial expressions, alongside STS. The study suggests the need for further specification of the relative role of the various brain areas involved in the perception of facial expression. Face processing appears to rely on more interactive and functionally overlapping neural mechanisms than previously conceptualised.

© 2015 Elsevier Ltd. All rights reserved.
1. Introduction

Human faces are among the most complex and at the same time most frequently encountered stimuli in our daily life, carrying information important for survival and gene propagation (Little, Jones, & DeBruine, 2011). The human expertise in reading faces allows us to decode the static as well as changeable information they convey (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000). While static features include identity, gender, attractiveness or age, changeable features include gaze direction, utterance of speech and emotional expressions. Since humans can readily decode emotion expressions from faces, it should in turn be possible to decode how this ability is represented in the brain.

A seminal model of face perception, based on the distinction of invariant and changeable features, proposes distinct modules in the brain, each in charge of carrying out different tasks when observers perceive faces (Haxby et al., 2000). According to the Haxby model, the fusiform gyrus is the most important for processing the invariant aspects of the face, such as identity or gender, while the superior temporal sulcus (STS) is reported to be in charge of processing changeable features, like gaze or emotion expression (Hoffman & Haxby, 2000). This changeable information may be conveyed not only by dynamic stimuli like video clips, but crucially also by static configurations of muscle movement as represented by pictures of expressive faces (Haxby et al., 2000). Together with the occipital face area, from which they receive input signals, fusiform gyrus and STS constitute the core face network.

These core regions are supported by other brain areas located throughout the brain, which constitute the extended face system. They contribute to the more specific demands of face processing in a task-dependent manner. For example, amygdala and insula are supposed to be recruited when processing emotion expressions (Haxby et al., 2000; Hoffman & Haxby, 2000). This model and the dissociations it predicts are grounded in cognitive theories (Bruce & Young, 1986, 2012) and are supported by a number of neurological (Bruyer et al., 1983; Duchaine, Parker, & Nakayama, 2003) and neuroimaging (Hoffman & Haxby, 2000; Winston, Henson, Fine-Goulden, & Dolan, 2004) studies.

However, it is an ongoing question to what degree the neural basis of expression processing is restricted to STS, amygdala and insula (Calder & Young, 2005). While the Haxby model itself states that interactions do take place between regions, its emphasis is on functional dissociations.

That, at least on the macroscopic level, substantial overlap of function may be indeed present, is reflected by the neuroimaging literature which often finds pronounced emotion effects in fusiform gyrus (Fox, Moon, Iaria, & Barton, 2009; Ganel, Valyear, Goshen-Gottstein, & Goodale, 2005; Kawasaki et al., 2012; see Fussar-Poli et al., 2009 and Sabatinelli et al., 2011 for meta-analyses). For example, functional magnetic resonance imaging (fMRI) studies using adaptation designs found that both fusiform gyrus and STS are responsive to changes in identity and expression (Fox et al., 2009; Ganel et al., 2005).

Fox et al. (2009) investigated the occipital and fusiform face areas and the STS regarding their responsiveness to changes in expression versus identity of faces using an adaptation design. Participants were presented with morphed stimuli changing either along an identity or expression dimension. Release from adaptation was shown for the occipital face area when any structural changes occurred in a face, indicating that this region codes for low-level features. Both fusiform face area and STS showed release from adaptation when participants experienced a change of either identity or expression (e.g., the face switched from angry to happy). Therefore, the regions only reacted when a specific boundary in subjective experience was crossed. Since both regions were responsive to both feature dimensions, the results suggest the dissociation of expression and identity processing may not be as pronounced as previously thought.

Furthermore, re-entrant models of emotion processing state that areas in the ventral stream, including the fusiform gyrus, receive top-down input from regions like the amygdala to allow further detailed processing (Adolphs, 2002; Vuilleumier, 2005). This is often reflected in parallel activity in the amygdala and the fusiform gyrus during processing of emotional material (Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005) and in the fact that both areas show higher activity during perception of emotional than of neutral expressions (Fussar-Poli et al. 2009; Sabatinelli et al. 2011). While re-entrant models add to our understanding of how facial expressions are processed in the brain, they leave open the question of how different brain areas, such as fusiform gyrus, STS or amygdala, directly compare to each other regarding their relative roles in expression perception.

Most of the past fMRI studies addressing face processing in fusiform gyrus and STS have analysed task-induced increments of activity using univariate analyses. This approach assumes that differences in information content between experimental conditions are coded in linear increments or decrements in most or all voxels in the brain areas under investigation. However, advances in the analysis of fMRI data (Haxby, 2001) have allowed to investigate the multi-voxel patterns of perceiving emotional material in complex scenes (Baucom, Wedell, Wang, Blitzer, & Shinkareva, 2012), voice prosody (Ethofer, van de Ville, Scherer, & Vuilleumier, 2009), facial expressions (Harry, Williams, Davis, & Kim, 2013; Petro, Smith, Schyns, & Muckli, 2013; Said, Moore, Engell, Todorov, & Haxby, 2010) or of modality-independent emotion representations (Peelen, Atkinson, & Vuilleumier, 2010).

Multi-voxel pattern analysis (MVPA) may help to compare regions in regard to how much information they carry about expressions of emotion. While MVPA has previously been used to study facial expression perception, these studies focused only on single anatomical regions from the Haxby model (STS: Said et al., 2010; fusiform gyrus: Harry et al., 2013) or investigated the role of V1, an early visual area not traditionally incorporated in face processing models (Petro et al., 2013).

The study by Said et al. (2010) focused on face processing in the STS, using short videos depicting expressions of seven basic emotions. The authors showed that activity in both anterior and posterior parts of STS can be used to successfully predict which expression a participant has seen. However, the study did not compare STS with other regions from the Haxby model, due to the high-resolution acquisition protocol, which
prohibited full-brain coverage. The role of the fusiform gyrus has been targeted by the study of Harry et al. (2013). The authors used pictures of six facial expressions and were able to successfully predict the perception of most of them from activity in both the left and the right fusiform face area. Together with previous univariate analyses (Fox et al., 2009; Ganel et al., 2005; Kawasaki et al., 2012), this MVPA study provides evidence that the fusiform gyrus is indeed sensitive to information about emotional expressions.

We aim to investigate whether the claim made by the Haxby model, namely that STS, amygdala and insula predominantly code for expressions of emotion, can be corroborated using MVPA, directly comparing classification success across all key areas specified by the model. If so, STS, amygdala and insula should have better classification performance than other regions specified by the model, such as inferior occipital gyrus, fusiform gyrus, intraparietal sulcus and anterior temporal regions. On the other hand, a more interactive model might suggest the contribution of other regions to the perception of expression of emotion, in particular the fusiform gyrus (Calder & Young, 2005; Fox et al., 2009). Comparison of classification accuracies across regions allows for a comprehensive test of these alternatives.

The present study builds on this previous work and employs an event-related fMRI design with happy, angry, fearful and neutral facial expressions. A selection was made to keep the duration of the experiment within reasonable limits. Also, these expressions are most often used in studies of emotional face perception. We directly compare the response patterns in all key brain regions of the Haxby model, and quantify how well their activity patterns can be used for predicting the presence of a facial expression. This provides a measure of the representational content in each region in regard to the presence of a facial expression. Independent meta-analytical data from the Neurosynth database (Yarkoni, Poldrack, Nichols, van Essen, & Wager, 2011) that aggregates activation from thousands of previous fMRI studies, was used to define the respective regions of interest.

2. Methods

2.1. Participants

Fourteen healthy participants (7 female; age: M = 25.4, SD = 2.6) took part in the study. Participants had normal or corrected-to-normal vision and reported no history of psychiatric or neurologic illness. All participants gave written informed consent prior to data acquisition. The study was approved by the Ethics Board of the Department of Psychology, University of Bielefeld, and was conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

Faces were derived from the MPI Faces database (Ebner, Riediger, & Lindenberger, 2010). Forty-eight face identities (24 female) were identified from the pool of “young” faces and each face identity was selected to show all four expressions. Faces were cut out along their contours (preserving hair), converted into greyscale and manually adapted to reduce variability in size and eye position. Faces were presented inside a white rectangle, placed centrally on a black background. Stimuli subtended a vertical visual angle of 12° and a horizontal visual angle of 10°, with variations depending on participant’s head size. See Fig. 1 for two examples of experimental stimuli.

2.3. Design and procedure

Participants performed a gender discrimination task while viewing faces from the four categories and indicated face gender by pressing a button with the left or right hand (counterbalanced across participants). Equal numbers of male and female faces were included in each expression category, rendering the motor response patterns orthogonal to the classes of facial expressions.

To increase participant comfort, stimuli were presented in 6 runs (cf. Said et al., 2010), lasting 213 sec each. In each run, 8 faces per expression category were shown, resulting in an overall presentation of 48 faces per category. Each run started with a fixation cross, followed by a face stimulus presented briefly for 200 msec, minimizing variance in eye fixation patterns. The face stimulus was followed by a mask, consisting of an oval shaped, grey scaled, sinus vortex shown for 500 msec. The mask was followed by a question mark presented for 1000 msec, cuing participants to perform the gender discrimination task. Subsequently, a fixation cross of jittered duration (2000–12000 msec) was presented, yielding a minimum stimulus onset asynchrony (SOA) of 3700 msec between faces (cf. Fig. 1). Stimulus sequence and jitter times were determined using Chris Rorden’s fMRI Design Software (http://www.mccauslandcenter.sc.edu/CRNL/tools/fmrisim). Different stimulus sequences were created for each subject. The paradigm was created using Presentation software (www.neurobs.com) and back-projected onto a screen which participants viewed over a head coil-mounted mirror.

Fig. 1 – Schematic representation of the employed paradigm.
2.4. Selection of regions

Target regions were selected on the basis of the Neurosynth meta-analytical database (Yarkoni et al., 2011), comprising 9721 studies, as of 10 September 2014. To derive maps for each region of interest, the database was queried for the region names (e.g., ‘fusiform gyrus’) and the reverse inference maps were downloaded. These maps report the probability of a keyword being present in a study given activation in a particular voxel. Thereby, only voxels highly specific for the keyword are included in a map.

The derived maps were thresholded at multiple levels. The smallest ROIs were created by including the top 25 voxels (of 2 mm³ size) and ROI size was subsequently doubled resulting in masks of 25, 50, 100, 200, 400, 800 and 1600 voxel. The extracted regions are described in detail in Table 1 and Fig. 2.

As some regions like the amygdala are not expected to be larger and as regions start to overlap once a certain size is reached, no larger ROIs were used. The auditory cortex was dismissed from the study as no auditory information was presented and because it is too close to the STS and overlaps quickly occur. A whole-brain mask was additionally used to minimize signal loss in the amygdalae (TR = 33 msec, Flip Angle 90°, Field of View 192 mm, slice thickness 4 mm, in-plane resolution 2.4 × 2.4 mm). 71 EPI volumes were acquired for each of the six runs.

2.5. Acquisition protocol

MRI data were collected using a 3T SIEMENS Verio Scanner with a 12-channel head coil.

A high-resolution MPRAGE structural scan was acquired with 192 sagittal slices (TR = 1900 msec, TE = 2.5 msec, 8 mm slice thickness, .75 × .75 in-plane resolution).

Functional echo-planar images (EPI) were acquired with 35 coronal slices, oriented perpendicular to the hippocampus to minimize signal loss in the amygdalae (TR = 3000 msec, TE = 33 msec, Flip Angle 90°, Field of View 192 × 192 mm, slice thickness 4 mm, in-plane resolution 2.4 × 2.4 mm). 71 EPI volumes were acquired for each of the six runs.

2.6. Preprocessing and beta extraction

Preprocessing and beta extraction were performed using SPM8 (www.fil.ion.ucl.ac.uk/spm/). Motion correction was carried out with the Realigned&Warp function. Motion that was associated with significant signal intensity changes was identified and corrected using the ArtRepair Toolbox (Mazaika, Whitfield, & Cooper, 2005). In the overall sample 0.4% of volumes were interpolated (maximum of 1.9% in a single participant). Interpolation of volumes was preferred over excluding trials to preserve equal case numbers across conditions and allow for straightforward interpretation of accuracies. No further pre-processing steps were employed. Pre-processed functional data were used as input for run-wise GLM first level designs, to derive beta images for every single event, yielding one regressor for the event of interest, a second regressor for all other events and a third regressor defining a block of instructions the participants read in each run (adapting the rationale of Mumford, Turner, Ashby, and Poldrack, 2012), creating one GLM per event. For rapid event-related designs this approach leads to more representative trial-by-trial estimates of the true activation magnitudes, as compared with beta-series regression, where each trial is a separate regressor in one GLM (Mumford et al., 2012). Data for each GLM were high-pass filtered and time and dispersion derivatives were included into the model as an alternative to slice timing correction, since the effect of slice timing depends on the reference slice chosen (Henson, Buechel, Josephs, & Friston, 1999) and may limit direct comparison of brain regions. Only the beta image of the canonical HRF for the event of interest was used as input for further analysis.

To transform the atlas ROIs into the participants’ native space, MPRAGE images were brain extracted using the BET algorithm (Smith, 2002) and co-registered to the mean functional image. Co-registered MPRAGE images were normalized using New Segment (Ashburner & Friston, 2005). Inverse deformation fields were applied to the atlas ROIs, which were then resliced into the participants’ native space.

To investigate which brain regions showed significant activations for face stimuli and to determine how this activity corresponded with the ROIs for the MVPA, overall brain activity evoked by the task versus rest was assessed using a mass-univariate analysis.

2.7. MVPA

Pattern analyses with PRoNTo v1.1 (Schrouff et al., 2013) were carried out separately for each participant and later used as input for group statistics. Classification of the four categories of facial expression was carried out, with each beta image

Table 1 – Description of selected ROIs.

<table>
<thead>
<tr>
<th>Name in Haxby Model</th>
<th>Neurosynth keyword</th>
<th>Number of studies in database</th>
<th>Peak voxel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior occipital gyrus</td>
<td>‘inferior occipital’</td>
<td>73</td>
<td>48  87  12</td>
</tr>
<tr>
<td>Superior temporal sulcus</td>
<td>‘sts’</td>
<td>154</td>
<td>–56  16  4</td>
</tr>
<tr>
<td>Lateral fusiform gyrus</td>
<td>‘fusiform gyrus’</td>
<td>381</td>
<td>–38  46  14</td>
</tr>
<tr>
<td>Intraparietal sulcus</td>
<td>‘intraparietal sulcus’</td>
<td>388</td>
<td>–30  50  44</td>
</tr>
<tr>
<td>Amygdala</td>
<td>‘amygdala’</td>
<td>1050</td>
<td>–22  4  20</td>
</tr>
<tr>
<td>Insula</td>
<td>‘insular cortex’</td>
<td>159</td>
<td>36  –6  10</td>
</tr>
<tr>
<td>Anterior temporal</td>
<td>‘anterior temporal’</td>
<td>116</td>
<td>-48  16  26</td>
</tr>
</tbody>
</table>

Reverse Inference Maps (Keyword/Activity) were downloaded from www.neurosynth.org on 10 September 2014. The database consisted of 9721 studies as of that date. x y z coordinates refer to MNI space in 2 mm³ resolution; see Yarkoni et al., 2011 for further details on the Neurosynth database.
representing a single face event. Each of the four classes thus consisted of 48 beta images. The first-level masks from preceding GLM analyses were used to exclude non-brain voxels. Feature extraction was performed using each of the respective ROIs, without additional voxel selection. Features were mean-centred and a Multiclass Gaussian Process Classifier (MGPC; Rasmussen, 2006) was used to test if each single ROI could distinguish between the four facial expressions. Gaussian process models are probabilistic kernel methods, which can be used for multi-class problems (Rasmussen, 2006).

Data were cross-validated by split-half analysis, with the 48 images from each group separated into odd and even trials. Control analyses were employed to ensure that classifier performance was not driven by differences in behavioural performance, by including only trials with correct responses. To ensure that test and training sets were independent, the same cross-validation schemes were carried out for 10 random shuffles of class labels, each shuffle balanced to contain an equal number of each facial expression in each group (25%), thus expected to perform at chance level.

3. Results

3.1. Behavioural data

Overall, mean accuracy in the gender discrimination task was 89% (SD = 4.0). There were significant effects for expression category \( F(3,39) = 13.07; p < .001; \eta_p^2 = .50 \) and the face \( \times \) gender interaction \( F(3,39) = 19.12; p < .001; \eta_p^2 = .60 \). Fearful and angry expressions were less accurately responded to than both happy and neutral faces (all \( p < .008; \) all \( d > .81 \)). Only within angry expressions was there a difference in accuracy due to stimulus gender, male faces being recognized significantly more accurately than female faces (\( p < .001; \) \( d = 2.01 \)) (see Table 2 for descriptive statistics). Mean reaction times for correct responses were not significantly different between conditions, as evaluated by a repeated measures Analysis of Variance (ANOVA) \( F(3,39) = 1.78; p = .167; \eta_p^2 = .12 \) (see Table 2 for descriptive statistics).

### Table 2 – Descriptive statistics of behavioural data at the group level.

<table>
<thead>
<tr>
<th>Facial expression</th>
<th>Reaction time (msec)</th>
<th>Gender classification Accuracy (Pr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male faces</td>
<td>Female faces</td>
</tr>
<tr>
<td>Happy</td>
<td>681.4 (172.31)</td>
<td>94.0 (3.91)</td>
</tr>
<tr>
<td>Angry</td>
<td>704.1 (183.07)</td>
<td>97.6 (3.15)</td>
</tr>
<tr>
<td>Fearful</td>
<td>674.7 (167.49)</td>
<td>83.6 (8.08)</td>
</tr>
<tr>
<td>Neutral</td>
<td>683.6 (160.79)</td>
<td>92.3 (5.13)</td>
</tr>
</tbody>
</table>

Reaction times and classification accuracies in the gender classification task. Shown are means and standard deviations of correct responses only.
3.2. Univariate analysis

A mass-univariate analysis for the average effect of face stimuli against fixation cross revealed enhanced activity in fusiform gyrus and ventral visual areas, insula, cingulate gyrus, amygdala, thalamus, frontal gyri and cerebellum, among others (Fig. 3).

3.3. MVP analysis

Two analyses were carried out. One where all trials were included (main analysis) and one where only correctly responded to trials were used (control analysis). This analysis should compensate for any potential differences in response patterns and task difficulty between the four expression categories, which were suggested by the behavioural data. Moreover, both analyses were compared against random performance by shuffling all condition labels.

3.3.1. Main analysis

Accuracies in all regions were tested against 25% chance performance with one-sample $t$-tests. Accuracies in all regions were significantly ($p < .05$) above chance-level, in 46 of 50 cases (7 ROIs x 7 sizes + whole brain ROI), suggesting successful facial expression discrimination in all target brain regions.

Regarding the random shuffles, 48 of the 50 comparisons against 25% chance were not significant ($p > .05$), suggesting that the used classification scheme is valid. Using $p = .05$ as a significance threshold, on average 2.5 tests out of 50 can be expected to be significant purely by chance.

As evident from Fig. 4, accuracy of the four regions increased overall with ROI size (ANOVA for main effect of ROI size: $F(6,78) = 41.1, p < .001$). On a descriptive level, the increase in accuracy seems to be less pronounced for the largest ROIs, suggesting that a plateau may be reached. Since ROIs cannot be infinitely enlarged due to anatomical constraints, it is unlikely that inclusion of even larger ROIs would provide additional information, at least in most cases. The 31.6% accuracy of the whole brain ROI implies that further increases in size would ultimately lead to a convergence of all ROIs to this value.

Across the tested ROI sizes, a stable rank order of discrimination accuracy emerged, as evident from visual inspection and from the fact that there was no interaction of ROI by size [$F(36,468) = 1.2, p > .1$]. Therefore, in addition to testing for a significant main effect of region [$F(6,78) = 4.4, p < .001$], accuracy values for each region were averaged across all 7 ROI sizes and directly compared with each other using paired $t$-tests (Fig. 5).

The paired comparisons revealed that fusiform gyrus, STS and anterior temporal gyrus all exhibited significantly higher accuracies than the amygdala and inferior occipital gyrus (Fig. 5b). The fusiform gyrus showed significantly higher accuracies in 4 out of the 6 pairwise comparisons, including comparisons with intraparietal sulcus and insula, while being numerically superior but not significantly different from either STS or anterior temporal regions.

3.3.2. Control analysis

For the control analysis, only correctly responded to trials were used. These trials were matched to include an equal number of male and female faces by random exclusion of surplus cases, leading to a mean of 35 trials per category (SD = 5.0, range 22–40 trials per participant). In spite of the reduced trial number, results closely resembled the original patterns (Figs. 6 and 7).

Fig. 3 – Results of the univariate whole-brain analysis for an all conditions > baseline contrast; blue-green colour scale shows results thresholded at $p < .001$ uncorrected with a $p < .05$ cluster-level FWE-correction; red-yellow colour scale shows results thresholded at $p < .05$ FWE-corrected; colour-coding as scaled in the upper-hand bars reflects size of t-values; structural image represents mean normalized MPRAGE image of the study sample; left of image is left of brain for coronal view; figure created using MRicron.
All 50 cases in the control analysis were significantly above chance ($p < .05$), and 48 of the 50 random shuffles were not significantly above or below chance ($p > .05$).

Regarding the pairwise comparisons, fusiform gyrus, STS and anterior temporal regions still showed significantly higher accuracies than the amygdala. Accuracies in the fusiform gyrus were still overall highest at a descriptive level and exhibited significantly higher accuracies as compared to amygdala, inferior occipital gyrus and insula (Fig. 7b). STS was also superior to inferior occipital gyrus.

4. Discussion

This study compared the relative importance of seven brain regions from the Haxby model for discriminating facial expressions, using MVPA of fMRI data. As could be expected by their theoretically postulated general involvement in face processing, consistent above-chance accuracies were found for all target regions, as well as for a whole brain mask. When directly comparing regions, fusiform gyrus, STS and anterior temporal regions showed highest accuracy values. Of note, the fusiform gyrus exhibited significantly higher accuracies than amygdala, inferior occipital gyrus, intraparietal sulcus and insula.

Control analyses closely followed these results in numerical terms, while also showing that the reliability of some of the pairwise comparisons requires further confirmation. Still, the main comparisons provide a consistent picture of superior performance of fusiform gyrus and STS voxels for discrimination between multiple facial expressions. Random shuffles guarding against spurious results were at chance level.

Fig. 4 – Mean accuracy of the main analysis for all ROIs as a function of size; a) Real labels; b) Shuffled labels (expected to perform at chance); STS, superior temporal sulcus.

Fig. 5 – a) Mean accuracy values of the main analysis for all ROIs, averaged over all sizes; b) Results of pairwise $t$-tests for the comparison of averaged ROI accuracies from the main analysis; $t$-values colour-coded according to the right-hand bar; positive $t$-values denote higher values in row compared to column region; asterisk denotes significant result at $p < .05$; FG, fusiform gyrus, STS, superior temporal sulcus; AT, anterior temporal; INS, insula; IS, intraparietal sulcus; IO, inferior occipital gyri; AMY, amygdala.
In summary, all brain regions from the Haxby model could be used robustly for above-chance prediction of the facial expression a participant has seen. However, they differed in their relative contribution. The overall size of accuracies is comparable to previous studies on facial expressions (Petro et al., 2013; Said et al., 2010) or face gender (Kaul, Rees, & Ishai, 2011). It should be taken into account that no pre-selection of most diagnostic voxels was performed, thereby preserving noisy data. Thus, the present results represent a conservative estimate of the classification capacity of the selected brain regions.

The present results suggest the need for re-evaluation of the relative importance of the fusiform gyrus for processing facial expressions. They also help to assess the degree to which regions in the neural face processing system can take over different tasks.

The STS has been conceptualised as the sole core face area in charge of processing facial expressions. The present results, much as meta-analytical data (Fusar-Poli et al., 2009; Sabatinelli et al., 2011), other experimental findings (Fox et al., 2009; Ganel et al., 2005; Kawasaki et al., 2012; Monroe et al., 2013), or re-entrant models of emotion processing (Vuilleumier, 2005) suggest that the fusiform gyrus may also play a critical role in coding facial expressions. In addition, a recent MVPA study (Harry et al., 2013) also reported above-chance classification in fusiform gyrus when trying to classify six different emotion categories, providing further evidence for its role in processing facial expressions. The present study further corroborates these findings and provides critical supplementary information by directly comparing fusiform gyrus with STS and other brain regions from the Haxby model.

In their seminal paper, Haxby et al. (2000) already discussed that the fusiform gyrus may play at least a supportive role in the perception of emotional expressions. Present data indicate a role that seems to go well beyond a purely

---

**Fig. 6** – Mean accuracy of the control analysis for all ROIs as a function of size; a) Real labels; b) Shuffled labels (expected to perform at chance); STS, superior temporal sulcus.

---

**Fig. 7** – a) Mean accuracy values of the control analysis for all ROIs, averaged over all sizes; b) Results of pairwise t-tests for the comparison of averaged ROI accuracies from the control analysis; t-values colour-coded according to the right-hand bar; positive t-values denote higher values in row compared to column region; asterisk denotes significant result at \( p < .05 \); FG, fusiform gyrus; STS, superior temporal sulcus; AT, anterior temporal; INS, insula; IS, intraparietal sulcus; IO, inferior occipital gyri; AMY, amygdala.
The role of subcortical regions such as the amygdala may not be crucial, as reflected in the amygdala's overall poor accuracy in the present study. This is striking especially since studies with non-human primates show that it can code for facial expressions as well as face identities (Gothard, Battaglia, Erickson, Spitler, & Amaral, 2007). Maybe differential recruitment of the amygdala only takes place when perception is accompanied by a critical difference in the amount of arousal induced by the employed stimuli, which may not have been present in the current design. In this study the comparison of task-induced activity against fixation cross does show overall amygdala activation and argues against the present results being due to signal loss in this area.

Therefore, this result may also reflect a true lack of primary discriminative ability, in line with some GLM findings showing no difference in mean activation of the amygdala for different facial expressions, compared with neutral faces (Fitzgerald, Angstädtt, Jelsone, Nathan, & Phan, 2006; van der Gaag, Minderaa, & Keysera, 2007). Still, it needs to be acknowledged that reduced signal-to-noise ratio in the amygdala is a commonly occurring problem in fMRI, due to the region's location near the nasal cavities, and cannot be completely ruled out in the present study.

Likewise, the insula did not exhibit a remarkable role in expression classification. This may be due to the choice of expressions which did not include disgust, which has been shown to activate the insula frequently (Sambataro et al. 2006). This issue could be addressed in future studies.

The present results indicate STS and fusiform gyrus as most important for processing emotion expressions. Therefore, a vital question to be addressed by lesion studies or intracranial recordings concerns the temporal dynamics of interactions between these regions. Especially, it should be examined to what degree high accuracies in the fusiform gyrus are due to input it receives from STS and whether activity in STS is a necessary condition for the engagement of fusiform gyrus in expression processing.

The similar performance of fusiform gyrus and STS might also be partly due to the use of static photographs in the present study. It is conceivable that with video clips showing the movements of facial muscles (e.g., Said et al., 2010), STS would outperform the fusiform gyrus. However, the original Haxby model also emphasised that STS is well-able to interpret static configurations of faces, using knowledge about how they typically move to produce different expressions (Haxby et al., 2000, p. 288). Nevertheless, the present results should be extended using animated stimulus material.

It also needs to be investigated to what degree the performance in the selected regions is due to coding for low-level features, since on average each facial expression produces a different retinal image, irrespective of whether genuine recognition of an expression is taking place. Therefore, expression decoding from stimuli with different viewing angles, different emotion intensities (cf. Fox et al. 2009) or of upside-down faces, which do not produce the same emotional effects as correctly oriented faces (Phelps, Ling, & Carrasco, 2006), may be used in future studies to further our understanding of the processing of emotional expressions in the brain. If however the present results were mainly driven by differences in low-level features, highest performance in occipital areas would have to be expected, which is not the case. Future studies could also take advantage of using functional localizers, to precisely map the face-selective subregions of the fusiform gyrus and other areas. This would allow to delineate the precise overlap between face-selective and expression-selective regions within this structure.

To summarize, our study compared the predictive accuracy of seven key regions from the Haxby model for coding expressions of emotion. We found consistent evidence for a prominent role of the fusiform gyrus, as well as STS. While our findings are in principle consistent with the model by Haxby et al. (2000), they also suggest that the relative importance of these regions may deviate from the initial proposal and that the role of the fusiform gyrus may need to be extended to processing emotion expressions, at least when viewed from static photographs.

Acknowledgements

Research was funded by DFG grant KI1286/4-1 and the DFG, Cluster of Excellence 277 “Cognitive Interaction Technology”. MRI facilities used for the project are supported by the “Gesellschaft für Epilepsieforschung e.V.”. We would like to thank Sebastian Schindler for helpful discussion.

References


Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. Cerebral Cortex, 15, 1234–1242.