Proposed journal section
European Journal of Neuroscience
Cognitive Psychology

Title
Online decoding of object-based attention using real-time fMRI

Authors
Adnan M Niazi[1, 2], Philip L C van den Broek[1], Stefan Klanke[1], Markus Barth[1], Mannes Poel[2], Peter Desain[1], Marcel A J van Gerven[1]

Author Affiliations
1. Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, NL 6500 HE, Nijmegen, The Netherlands
2. University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, NL 7500 AE, Enschede, The Netherlands

Corresponding author
Adnan Muhammad Niazi,
Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, NL 6500 HE, Nijmegen, The Netherlands.
Tel: +31 (0)24 36 55931
Fax: +31 (0)24 36 52728
Email: adnaniazi@gmail.com

Running Title
Real-time Decoding of Object-based Attention

Total Number of
1. Pages: 31
2. Figures: 5
3. Tables: 0
4. Equations: 2

The total number of words in:
1. the whole manuscript: 8076
2. the Abstract: 187
3. the Introduction: 1051

Keywords
Real-time fMRI, categorization, multivariate decoding, object-based attention

Abstract

Visual attention is used to selectively filter relevant information depending on current task demands and goals. This study used real-time functional magnetic resonance imaging (fMRI) for online decoding of attention to objects belonging to two different semantic categories. Superimposed pictures of a face and place were presented to subjects who had to attend to only one of the pictures. Decoding of the semantic category of the attended picture was performed on a TR-by-TR basis using a multivariate decoder and its performance was also compared offline to a univariate region-of-interest based approach. The multivariate decoder yielded significantly above chance-level decoding accuracy whereas the univariate approach failed to perform above chance-level. Moreover, multivariate decoding was based on sparsely distributed patterns of activity in multiple brain regions. This indicates that optimal decoding of object-based attention requires a whole brain multivariate decoding approach which can take distribution patterns of cortical activity into account. The real-time fMRI system presented in the study not only allows us to probe object-based attention in an online setting but may also form the basis of brain-computer interfaces that are driven by modulations of high-level cognitive states.

1. Introduction

In our daily life, we are continuously flooded with a multiplicity of stimuli, all competing for our attention. However, only a small amount of information can be assimilated at any given time due to limited information-processing capacity (Desimone & Duncan, 1995). To effectively cope with this influx of information, the brain must filter out task-relevant information from the environmental stimuli based on current task demands (Rissman & Wagner, 2012). Selective

attention drives this filtering by focusing processing resources on particular aspects of the environment or stimuli, whilst disregarding others. This selective attention can be deployed to a certain feature such as color or motion (feature-based attention), to a certain location in space (space-based attention) or to an organized chunk of information that corresponds to an object (object-based attention) (Serences *et al.*, 2004). Object-based attention uses top-down control to enhance the sensory representation of the attended object, resulting in its corresponding features being processed more efficiently. Evidence for this top-down control has emerged from numerous studies using a variety of measurement techniques. For instance, in a study by Cerf and colleagues (Cerf *et al.*, 2010), which employed single-unit recordings, neurons encoding the Marilyn Monroe face fired selectively when subjects were presented with a composite picture of Marilyn Monroe and Josh Brolin while being asked to attend only to the picture of Marilyn Monroe. Subjects were able to robustly regulate the firing rate of their neurons, increasing the rate for the target picture (Marilyn Monroe) while simultaneously decreasing the rate for the non-target picture (Josh Brolin). The study indicates that despite competing bottom-up sensory inputs, firing rates in medial temporal lobe neurons can be voluntarily regulated to reflect object-based selective attention. Studies using fMRI, EEG, and MEG have likewise shown that cortical representations for the task-relevant stimuli can be enhanced while at the same time suppressing the activations for task-irrelevant stimuli or features (Luck *et al.*, 1993; Eimer, 1996; Hopf *et al.*, 2000; Serences *et al.*, 2004; Gazzaley *et al.*, 2005; Yi *et al.*, 2006; Rahnev *et al.*, 2011).

Recently, with the introduction of multivoxel pattern analysis (MVPA), new insights have been gained in understanding the effect of goal-directed top-down control on cortical representations. In one of the first studies that employed MVPA to read subjective contents of the human brain

using fMRI, Kamitani and Tong (Kamitani & Tong, 2005) argued that if distinct neural patterns are associated with different grating orientations then these neural signatures can be used to decode attended orientation in two overlapping gratings. The study showed that a classifier, which was initially trained to differentiate activation patterns of individual grating orientations, was also able to decode the attended grating orientation when any two gratings were simultaneously presented. Furthermore, distributed information about the attended orientation was present even in V1, the earliest cortical level of visual processing (see also (Li *et al.*, 2004; Haynes & Rees, 2006)). This indicates that despite the presence of competing bottom-up sensory inputs, attentional signals biased neural patterns in favor of the task-relevant features. Further studies have reported that attention-driven top-down control can modulate the cortical representation of a range of different stimuli, from simultaneously presented motion fields to simultaneously presented visual objects (Reddy & Kanwisher, 2006; Macevoy & Epstein, 2009; Reddy & Tsuchiya, 2010) and even conjunction of features such as color and motion (Seymour *et al.*, 2009). See (Rissman & Wagner, 2012; Tong & Pratte, 2012) for more exhaustive reviews.

In this study, we investigated if the semantic category of an attended stimulus can be decoded non-invasively and in real-time when stimuli from two different categories are presented simultaneously. More specifically, we examined whether a classifier trained on pictures of faces and places presented separately, can be used to decode the attended category (face or place) when both a face and a place are presented simultaneously in the form of a composite picture. By presenting superimposed pictures of a face and a place, we investigated if object-based attention can bias the neural patterns towards the attended category and if these differentiating activity patterns can be picked up by multivariate pattern analysis in a real-time fMRI setting. Such an

attention-driven real-time decoding setup could form the basis for a brain-computer interface (BCI) for severely paralyzed and locked-in patients. Furthermore, such a system could be used to investigate if people can be trained to enhance their attention or prolong their attentional span (Jensen *et al.*, 2011).

Studies have shown that pictures of faces and places invoke spatially distinct and dissociable cortical regions, namely, fusiform face area (FFA) for face pictures and parahippocampal place area (PPA) for scenes (Puce *et al.*, 1995; Kanwisher *et al.*, 1997; Epstein *et al.*, 1999). More recently however, these regions have been shown to have a more overlapping and distributed representation than previously thought (Haxby *et al.*, 2001; Ewbank *et al.*, 2005; Hanson & Schmidt, 2011; Mur *et al.*, 2012; Weiner & Grill-Spector, 2012). In light of this new view, optimal decoding of the attended stimulus from these regions calls for a multivariate decoding approach which can detect these overlapping and distributed neural patterns. Thus, a traditional real-time fMRI approach, where localizer scans are used to define the regions of interest (ROIs) for deriving neurofeedback (deCharms *et al.*, 2004; Caria *et al.*, 2007; Haller, 2010; Zotev *et al.*, 2011; Chapin *et al.*, 2012), might be suboptimal in this setting. Therefore, in this study, we used whole-brain data to train a classifier to predict the mental state of a subject as this approach does not rely on any prior assumptions about functional localization (Laconte *et al.*, 2007; Hollmann *et al.*, 2011; Lee *et al.*, 2011; Xi *et al.*, 2011; Anderson *et al.*, 2011; DeBettencourt *et al.*, 2012). Additionally, we compared the performance of our multivariate decoding approach with a univariate decoding approach that does not take distributed activation patterns into account. Furthermore, because the predicted brain state is available on a moment-to-moment basis in real-time fMRI, these online detected brain states can be used to train subjects to modulate their ongoing brain activity. Such brain-state dependent stimulation provides a new avenue for investigating the neuronal substrate

of cognition (Hartmann *et al.*, 2011; Jensen *et al.*, 2011). To ascertain how this brain-state dependent stimulation will impact subjects' task performance, we conducted each attention trial twice, once with fMRI neurofeedback and once without it.

2. Methods

2.1. Subjects

Seven subjects (6 males, 1 female) with an average age of 23.4 (SD = 4.6) years participated in the experiment. All participants had normal vision and received either monetary compensation or study credits for participating in the experiment. The study was approved by the local ethics committee and all subjects gave a written informed consent. To keep them motivated during the experiment, participants were promised a monetary reward if their task performance (i.e., average decoding accuracy) in the experiment exceeded 95%.

2.2. Stimuli

The stimulus set consisted of pictures of famous faces and famous places collected from the World Wide Web. Previous studies have shown larger activations for familiar faces and places compared to unfamiliar faces and places, respectively (Shah *et al.*, 2001; Pierce *et al.*, 2004; Rosenbaum *et al.*, 2004). All pictures had a height and width of 450 x 450 pixels with a resolution of 95.987 pixels/inch, subtending a visual angle of 8°. None of the pictures were corrected for luminosity or spatial frequency.

2.3. Experimental protocol

Before putting subjects in the scanner, they were thoroughly briefed about the experiment to avoid any verbal communication during the real-time fMRI run. Participants were shown video

recordings of all experimental conditions and the task was verbally explained by the experimenter with the help of these videos. They were specifically instructed that feedback would be delayed by about 4 seconds (pipeline delay) for technical reasons and that no matter what the feedback was, they were to always focus their attention on the target picture in the hybrid picture. No instructions were given to maintain a specific gaze position. Moreover, subjects were allowed to close their eyes during rest periods but were advised to open their eyes a few seconds before this rest period was over.

2.4. Experimental design

The experimental design consisted of two separate sessions. A training phase, where a classifier was trained on the cortical activity patterns induced by face and place stimuli, and a test phase, where the classifier was used to decode object-based attention to face or place stimuli in real time.

The training phase consisted of fifteen 30 second blocks of face pictures and fifteen 30 second blocks of place pictures. These face and place blocks were interleaved with each other, with 12 second rest intervals in between. In each block, fifteen pictures were presented and the first picture was repeated at a random position in the block. Subjects were instructed to press a button on a button box with their right index finger when they saw the first picture repeated in that block. This kept participants actively engaged in the task throughout the training phase. Each picture within a block was presented for 1.5 s followed by a 0.5 s fixation period as shown in Figure 1A. All 14 pictures in each block were unique and used nowhere else in the experiment. The entire training phase took 22 minutes to complete.

In the test phase, subjects were required to attend to either faces or places, presented as 15 hand-picked hybrid stimuli showing a superimposed face and place image (see supplementary figures S1 and S2 and supplementary movie M1). Stimuli were shown once with feedback to the subject (feedback condition) and once without it (non-feedback condition). Thirty trials were collected in each of the two conditions, where 15 trials had a face picture as the target and the other 15 trials had a place picture as target. We used 12 interleaved mini-blocks of feedback and non-feedback trials, with each mini-block containing 5 trials. The ordering of mini-blocks was counterbalanced across subjects. Every trial in the test phase started with presentation of the target and non-target cue pictures for 1.75 s each followed by a 0.5 s fixation. The order of presentation of target and non-target picture cues was counterbalanced across subjects. Then the hybrid image was shown and the subjects had to attend to the target picture while ignoring the non-target picture. Due to pipeline delay, the hybrid image was kept unchanged for the first two scans. Next, for trials in the feedback condition, the relative visibility of the face and place pictures in the hybrid image was enhanced or decreased by 5% in each TR, depending on the prediction made by the classifier. This gave a visual indication to the subjects of their task performance on a TR-by-TR basis (see Figure 1B). The trials timed out after 28 s (14 TRs).

2.5 MRI acquisition parameters

Experiments were performed at the Donders Institute for Brain, Cognition and Behaviour using a Siemens MAGNETOM Tim TRIO 3.0 T scanner with a 32-channel head coil. First, high-resolution anatomical images were acquired using an MPRAGE sequence (TE/TR = 3.03/2300 ms; 192 sagittal slices, isotropic voxel size of 1x1x1 mm). Then the real-time fMRI run was initiated and functional images were acquired using a single-shot gradient echo planar imaging

(EPI) sequence (TR/TE = 2000/30 ms; flip angle = 75°; voxel size = 3x3x3.3 mm; distance factor = 10%) with prospective acquisition correction (PACE) to minimize the effects of head motion during data acquisition (Thesen *et al.*, 2000). Twenty-eight ascending axial slices were acquired, oriented at about 30° to the AC-PC (Anterior-Posterior Commissure).

2.6. Real-time data export and preprocessing

All functional scans collected during the real-time fMRI run were acquired using a modified scanner sequence that dumped fMRI acquisition parameters in a text file at the start of each functional run. After the functional run was initiated, k-space data of each acquired functional image was inverse Fourier transformed and written to the hard-drive of the scanner host computer as a raw pixeldata file. As each new pixeldata file corresponding to every functional scan was generated, it was read using the protocol information generated earlier when the sequence was initiated. Pixeldata files were sent over Ethernet to another computer where the scans were stored in a FieldTrip (Oostenveld *et al.*, 2011) raw data buffer. Each newly buffered raw scan was then fed into a MATLAB-based (The Mathworks, Natick, MA) preprocessing pipeline.

The first preprocessing step involved selecting one of the two image series generated by the scanner sequence, the PACE series of images which is only prospectively corrected and the MoCo (Motion Corrected) series which is both prospectively and retrospectively corrected (Thesen *et al.*, 2000). We used the MoCo series of images as it had least residual motion in it. Then slice-time correction was applied followed by retrospective motion correction using an online rigid-body transformation algorithm with six degrees of freedom. This motion correction was carried out to remove any residual motion in the MoCo series. Then a recursive least-squares general linear

model was applied to each scan to remove nuisance signals (Bagarinao *et al.*, 2003). Five regressors, corresponding to DC offset, linear drift and three translational motion parameters, were used in the model. A grey matter mask was then applied to each scan to remove white matter and cerebral spinal fluid voxels. This mask was obtained from the high-resolution anatomical images with SPM8 (Wellcome Department of Cognitive Neurology, Queens Square, London, UK) using a unified segmentation-normalization procedure (Ashburner & Friston, 2005) and resliced to the resolution of the functional scans using the first acquired functional scan as a reference. After grey-matter masking, top and bottom slices in each scan were masked out. This is because retrospective motion correction always results in data loss at the boundary slices requiring their removal. Each scan, now fully preprocessed, was saved in a FieldTrip preprocessed data buffer. The entire real-time fMRI pipeline is shown in Figure 2.

2.7. Feature extraction and classification

From the preprocessed data buffer, data was sent to another MATLAB script which extracted features. First, scans in the training phase were shifted by 6 s to account for the hemodynamic delay. Then scans corresponding to 12 s rest periods between every consecutive face and place block were dropped. Then all scans within each of the face and place blocks were averaged and used to train a classifier. We used logistic regression in conjunction with an elastic net regularizer. Elastic net regularization will shrink and select regression coefficients, identifying relevant features (voxels) while performing well in the presence of correlated variables, making it a good choice for fMRI decoding.

Given a training set $\{(x_i, y_i)\}_{i=1}^{N}$, where N is the total number of observations, x_i is the i^{th} observation and y_i the corresponding response, the elastic net logistic regression model is fitted by maximizing the penalized log likelihood:

$$(\hat{\alpha}, \hat{\beta}) = \max_{\alpha,\beta} \left[\frac{1}{N}\sum_{i=1}^{N}\{y_i(\alpha + x_i^T\beta) - \log(1 + \exp(\alpha + x_i^T\beta))\} - \lambda P_\gamma(\beta)\right]$$

where λ is the regularization parameter, α is an offset term, β is a vector of regression coefficients and $P_\gamma(\beta) = \sum_{j=1}^{p}\left[\frac{1}{2}(1 - \gamma)\beta_j^2 + \gamma|\beta_j|\right]$ is the elastic net regularizer with mixing parameter γ. For this study, the value of γ was fixed to 0.99 for a sparse solution. For the regularization parameter λ, the whole regularization path was calculated with the maximum number of allowed iterations set to 100. The optimal setting of λ was then computed using nested cross-validation on 75% of the training data. Using a coordinate gradient descent algorithm (Friedman *et al.*, 2010), classifier training took only a few minutes to complete, after which the decoding phase was initiated. The preprocessed data, along with the trained classifier and all its predictions during the actual real-time run, were saved to disk for use in subsequent offline analyses.

During the decoding phase, each scan was classified individually using the trained classifier. An important aspect of this study is that scans in the first 6 seconds of each trial (transition period of the hemodynamic response) were also classified and used to generate neurofeedback. A study by LaConte and colleagues showed that hemodynamic activity in the transition period contains reliable information that can be decoded (Laconte *et al.*, 2007). These scans have mostly been ignored in real-time fMRI studies as only scans in the stable period are used where the hemodynamic response has reached a stable state. Because a timely feedback is a crucial prerequisite for operant conditioning (Mulholland *et al.*, 1979), our approach of classifying scans

in transition period mitigates the effect of hemodynamic inertia thereby making feedback more intuitive.

2.8. Neurofeedback generation

In trials with neurofeedback, the relative contribution of face versus place pictures in the hybrid image was dynamically updated at every TR based on classifier output. This dynamically adapted stimulus was generated using StimBox, a custom designed MATLAB-based layer built on top of Psychtoolbox (Brainard, 1997), designed specifically for generation of dynamic stimulus content for real-time neurofeedback studies. The timing of the stimulus and scheduling of the whole experiment was controlled in BrainStream (www.brainstream.nu), a MATLAB-based software package for real-time processing of continuous data streams.

2.9. Performance evaluation

Decoding performance was quantified in terms of accuracy, defined as the percentage of successfully predicted trials. A trial was regarded successful if the sum of log probability for target pictures ($\sum_{k=1}^{12} \log P(Target|Scan_k)$) was greater than the sum of log probability of non-target pictures ($\sum_{k=1}^{12} \log P(Non-Target|Scan_k)$) for all 12 scans. In this way decoding performance for the feedback and non-feedback condition was calculated. Because each condition contained trials in which subject had to attend either face (face trails) or place (place trials), performance for each of these trial types was calculated separately as well. A one-tailed t-test was then carried out to find if the performance was significantly above the 50% chance level. Furthermore, accuracy was calculated on a TR-by-TR basis to investigate how it evolved over the course of the 12 TRs.

We also compared the performance of our multivariate approach with a traditional univariate ROI based approach. In this method, ROI's are identified using the training data and then blood-oxygen-level-dependent (BOLD) activity from these ROIs are used drive the feedback during the test phase. To select the face-selective and place-selective ROIs, an SPM univariate analysis was performed on the training data. Two regressors, corresponding to the face and place blocks, were used in the general linear model. No other regressors were added to the GLM because the data had already been preprocessed by the real-time fMRI pipeline. Two contrasts (faces > places and places > faces) were then constructed and thresholded at 0.05 (family-wise error corrected). Two seed voxels were selected in each of these contrasts using the voxel with the highest t-value in both left and right hemispheres. A rectangular area of 5 x 5 voxels in the x-y plane and 3 voxels in the z-direction was generated around each of the seed voxels. In this way, face and place-selective regions were obtained in each of the two hemispheres. Using the data in the test phase, the detrended averaged BOLD activity in the face and place selective ROIs was first normalized. Then, the time series of face-selective ROI was divided (point-by-point in time) by the time series place-selective ROIs to yield the category of the attended class. That is, a ratio above one indicated a face whereas a ratio below one indicated a place. Decoding accuracy for each subject was then calculated by taking the proportion of correctly classified TRs relative to the total number of TRs.

2.10. Pattern analysis

In order to disentangle which features drove classification performance, an ROI analysis was carried out to investigate which brain regions were used by the classifier for training. This also allowed us to examine whether a particular region responded more strongly to one or the other class. Voxel selection was done using Analyze4D (www.analyze4d.com). ROIs for individual

subjects were defined using the locations of the non-zero weights of the classifier trained in the actual real-time run. These ROI were labeled using a subject-specific automatic anatomic labeling (AAL) mask (Tzourio-Mazoyer *et al.*, 2002) by warping the standard AAL mask from Montreal Neurological Institute (MNI) space to native space (Alemán-Gómez *et al.*, 2006) using the inverse of the spatial transformation matrix obtained from the SPM8 unified segmentation-normalization procedure (Ashburner & Friston, 2005). Before averaging the time courses of all voxels in each ROI, time courses of all voxels in each of these ROIs were high-pass filtered (0.01 Hz) with a discrete-cosine least-squares linear regression using the NIAK toolbox (Bellec *et al.*, 2011). Subsequently, percent signal change for each ROI was calculated using the mean of the entire time series as the baseline. The same procedure was repeated for all subjects and then the results across all ROIs and conditions were averaged together across the group. If the percent signal change was greater in the face block of the training session than the place blocks, then the region was categorized as a face-selective region. Otherwise, the region was categorized as a place-selective region.

Lastly, we also analyzed the classifier weights as a sanity check. The binary logistic regression assigned positive weights to face-responsive voxels and negative weights to place-selective voxels. For each subject, the classifier weights within each ROI were averaged together and then summed together across the ROIs. Regions showing increased percent signal change for face blocks during training in the ROI analysis should have been assigned positive weights and negative weights should have been assigned to the regions showing increased percent signal for the place blocks in the training session.

3. Results

The aim of this experiment was to investigate if the category of the attended target picture could be decoded in real time when a hybrid of the target and non-target picture is presented. Furthermore, the impact of real-time fMRI neurofeedback on decoding performance was examined. The results (see Figure 3A) show that the average accuracy for the feedback and non-feedback conditions was 79% (SD = 11.0%) and 78% (SD = 11.6%) respectively, which is significantly above chance level ($p < 0.002$). Face trials in feedback and non-feedback conditions were decoded with accuracies of 87% (SD = 15.6%) and 84% (SD = 14.3%) respectively (significant at $p < 0.002$). Similarly, place trials in feedback and non-feedback conditions were decoded with accuracies of 70% (SD = 16.7%) and 71% (SD = 15.3%) respectively (significant at $p < 0.05$). Paired samples t-tests failed to reveal any statistically significant difference between feedback and non-feedback conditions and trials. Figure 3B shows the results for the univariate ROI-based approach. Feedback and non-feedback conditions were decoded with an average accuracy of 52% (SD= 2.1%) and 52.1% (SD= 7.1%) respectively none of which was significantly above chance level ($p < 0.05$).

We expected that decoding performance would follow the hemodynamic response, increasing for scans in the transition period and levelling off in the stable period. By analyzing the decoding performance as a function of TR (see Figure 3C), one may observe that accuracy increased in the first six seconds of hemodynamic activity (TRs 1-3), and then leveled off in the following eighteen seconds of hemodynamic activity (TRs 4-12). Above chance-level decoding performance was achieved for all TRs after the transition period and even for some TRs within the transition period, indicating that the transition period does contain reliable information that can be decoded and used for feedback. At the level of individual TRs, significantly more TRs were classified as a face than place. A paired samples t-test shows a statistically significant difference ($t(6) = 2.898$, $p = .027$, α

= .05) between scans classified as face (M = 412.71, SD = 48.124) and place (M = 307.29, SD = 48.124).

The region-of-interest analysis on the classifier weights indicated that the classification algorithm selected 31 distinct brain regions across the group (see supplementary Figure S3 for a list of all these regions). Regions not activated in more than two subjects were excluded from further analysis. This left only nine brain regions as shown in Figures 4A and C. These included bilateral fusiform and lingual gyri, right parahippocampal gyrus, left and right inferior occipital lobes, and right middle and superior temporal lobes. Right fusiform gyrus, left and right inferior occipital lobes, and right middle and superior temporal lobes responded strongly to faces and, hence, were labeled as face-selective regions. Left fusiform gyrus, bilateral lingual gyri and right parahippocampal gyrus were more responsive to place stimuli and labeled as place-selective regions. The details of the results of the ROI analysis can be found in Figures 5A and B.

We anticipated that classifier weights would be positive for face-selective areas and negative for place-selective areas. The classifier weight analysis confirms that this was the case. Face-selective areas such as the right fusiform gyrus, left and right inferior occipital lobes, and right middle and superior temporal lobes were all assigned positive weights. Place-selective areas such as the left fusiform gyrus, bilateral lingual gyri and right parahippocampal gurus were all assigned negative weights. The classifier weights averaged across all subjects for all the brain regions are shown in Figure 4B.

To visually confirm that the experimental manipulation was having the desired effect, time courses of voxels selected by the classifier were analyzed. Figure 4E shows the time courses for two such voxels, one in the face-selective region in right superior temporal gyrus (MNI coordinates 46, -44,

14) and the other in the place-selective region of left fusiform gyrus (MNI coordinates -32, -68, -11). The anatomical labels for these two voxels were derived from the nearest grey-matter region reported by the Talaraich daemon (Lancaster *et al.*, 2000). The figure confirms that the voxel in the face-selective region is responding more strongly to face conditions and the voxel in the place-selective region is responding more strongly to place conditions.

4. Discussion

We studied whether object-based attention can be decoded in a real-time fMRI setting when two competing objects are presented simultaneously. Results show that the attended stimulus could be decoded in real-time with a significant accuracy of 78.5%. We observed that classification was biased towards faces as significantly more scans were classified as a face than as a place. Although we scrutinized saliency maps of the 50/50 hybrid image to ensure that none of the two pictures in each pair was more salient than the other, differences in salience cannot be completely ruled out and hence the observed bias towards one category. Another possible reason for the observed bias could be that faces are just easier to focus on due to our innate predispositions and the extensive experience with faces (Greenough & Black, 1992; Nelson, 2003).

We expected that above-chance accuracies would be obtained for scans in the transition period and our analysis confirms this. This supports the finding of LaConte et al., where an offline analysis showed that the transition period of the hemodynamic response contains reliable information that can be decoded with above-chance accuracy (LaConte *et al.*, 2007). We have therefore shown that predictions using scans in the transition period can be used online to reduce real-time fMRI neurofeedback delay by as much as six seconds.

Our multivariate decoding approach was also compared with a traditional univariate approach. Whereas the multivariate approach achieved a very high decoding performance of 78.5%, the univariate approach failed to reveal any significant decoding performance. This suggests that face and place representations are expressed in terms of distributed patterns of brain activity, requiring a multivariate decoding approach. This finding is corroborated by the fact that the classifier used multiple voxels distributed across various brain regions for decoding (please see supplementary Figure S4).

We investigated if fMRI neurofeedback of scan-by-scan brain state classification outcome can improve task performance. Neurofeedback did not significantly influence decoding performance. This is in contrast to our expectation that neurofeedback will help drive attention towards the target category, thereby resulting in higher decoding accuracies for trials with feedback compared to trials without feedback. A possible explanation for this might be that feedback and non-feedback trials were conducted in interleaved mini-blocks. This interleaving might have diminished any learning effects, as subjects would not have been able to discover any consistent strategy due to frequent switching between the feedback and non-feedback trials. In future studies, rather than using a within-subject design for feedback and non-feedback conditions, a between-subject design should be used. Moreover, the duration of feedback was chosen to be 12 TRs (24 seconds) as a compromise between the number of trials and the experiment duration. This might have been too short for any significant reinforcement learning. Previous real-time studies have used trial durations ranging from 15 s to 60 s conducted over the course of multiple days (see (Weiskopf *et al.*, 2004) for a review). Finally, the feedback was updated every TR which might have resulted in cognitive overload, thereby resulting in suboptimal learning in the feedback condition. Future studies should investigate the use of slower feedback update rates.

Our analysis revealed nine regions that were consistently used by the classifier to derive the predictions. Among these regions is the left fusiform gyrus which is usually associated with reading and word processing (McCandliss *et al.*, 2003; Hillis *et al.*, 2005; Dehaene & Cohen, 2011). However, this area has also been suggested to be sensitive to the conjunction of object and background scene information (Goh *et al.*, 2004). This view is strengthened by invasive studies in primates that also pointed to the presence of neurons in this area which are responsive to the conjunction of object features (Baker *et al.*, 2002; Brincat & Connor, 2004). Left fusiform gyrus may be showing more activity for place blocks than for face blocks since pictures of famous places in the stimulus set contained not only objects but also a wide variety of backgrounds. Pictures used in the face blocks rarely had objects in them. Only a few face pictures had necklaces, earrings, glasses, et cetera. Right fusiform gyrus showed a preference for face blocks whereas the left parahippocampal gyrus showed a preference for place blocks. These two regions have been implicated in many studies to be responsible for the processing of faces and place, respectively (Aguirre *et al.*, 1996, 1998; Kanwisher *et al.*, 1997; McCarthy *et al.*, 1997; Epstein & Kanwisher, 1998).

Two other regions selected by the classifier are the right medial temporal lobe and the right superior temporal lobe. Their involvement could be related to activity modulations induced by famous as opposed to non-famous stimuli. A study by Tempini and colleagues (Gorno-Tempini & Price, 2001) showed an effect of fame in anterior medial temporal gyrus (aMTG) that is common to faces and buildings, though this was stronger in right than in left aMTG. In our study the right temporal gyrus shows a preference for faces but not for places. This could be because many of the

supposedly famous landmarks used in the stimulus set were not as familiar to the subjects as expected. Most participants reported that although they could recognize many of the famous faces, they could not identify many of the pictures of the famous places. This asymmetry could drive classifier responses that dissociate between faces and places.

Finally, both left and right inferior occipital gyri were activated in the experiment, showing more activation for the face blocks. These regions contain the occipital face area (OFA). The OFA is spatially adjacent to the FFA and preferentially represents parts of the face, such as eyes, nose, and mouth (Liu *et al.*, 2002; Pitcher *et al.*, 2007, 2008). OFA is an essential component of the cortical face perception network and it represents face parts prior to subsequent processing of more complex facial aspects in higher face-selective cortical regions.

Concluding, we have shown that real-time fMRI allows for online prediction of attention to objects belonging to different semantic categories. Prediction is based on distributed patterns of activity in multiple brain regions. The outlined developments not only allow us to probe object-based attention in an online setting but may also form the basis of brain-computer interfaces that are driven by modulations of high-level cognitive states.

Acknowledgments

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. The first author was supported by a UTS grant from the University of Twente. We thank Paul Gaalman for his technical support during experimental set up and development of the real-time fMRI pipeline.

Abbreviations

AAL
Automatic anatomic labeling
BOLD
Blood oxygenation level dependent
EPI
Echo planar imaging
fMRI
Functional magnetic resonance imaging
HRF
Hemodynamic response function
MNI
Montreal Neurological Institute
MVPA
Multivoxel pattern analysis
ROI
Region of interest
SPM
Statistical parametric mapping
TR
Repetition time

References

Aguirre, G.K., Detre, J.A., Alsop, D.C., & D'Esposito, M. (1996) The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, **6**, 823–829.

Aguirre, G.K., Zarahn, E., & D'Esposito, M. (1998) An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, **21**, 373–383.

Alemán-Gómez, Y., Melie-García, L., & Valdés-Hernandez, P. (2006) IBASPM: Toolbox for automatic parcellation of brain structures. *Neuroimage*, **31**, 29-185.

Anderson, A., Douglas, P.K., & Cohen, M.S. (2011) Real-time functional MRI Classification of Brain States using Markov-SVM Hybrid Models: Peering inside the rt-fMRI black box. In Langs, G., Rish, I., Grosse-Wentrup, M., & Murphy, B. (eds.), Pro*ceedings of Neural Information Processing Syst*ems. Sierra Nevada, Spain, p. 242-255.

Ashburner, J. & Friston, K.J. (2005) Unified segmentation. *NeuroImage*, **26**, 839–851.

Bagarinao, E., Matsuo, K., Nakai, T., & Sato, S. (2003) Estimation of general linear model coefficients for real-time application. *NeuroImage*, **19**, 422–429.

Baker, C.I., Behrmann, M., & Olson, C.R. (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, **5**, 1210–1216.

Bellec, P., Carbonell, F., Perlbarg, V., Lepage, C., Lyttelton, O., Fonov, V., Janke, A., Tohka, J., & Evans, A. (2011) A neuroimaging analysis kit for Matlab and Octave. *Proceedings of the 17th International Conference on Functional Mapping of the Human Brain*.

Brainard, D.H. (1997) The psychophysics toolbox. *Spatial Vision*, **10**, 433–436.

Brincat, S.L. & Connor, C.E. (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, **7**, 880–886.

Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., & Birbaumer, N. (2007) Regulation of anterior insular cortex activity using real-time fMRI. *NeuroImage*, **35**, 1238–1246.

Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R.Q., Koch, C., & Fried, I. (2010) On-line, voluntary control of human temporal lobe neurons. *Nature*, **467**, 1104–1108.

Chapin, H., Bagarinao, E., & Mackey, S. (2012) Real-time fMRI applied to pain management. *Neuroscience Letters*, **520**, 1–8.

deCharms, C. R., Christoff, K., Glover, G.H., Pauly, J.M., Whitfield, S., & Gabrieli, J.D.E. (2004) Learned regulation of spatially localized brain activation using real-time fMRI. *NeuroImage*, **21**, 436 – 443.

DeBettencourt, M.T., Lee, R.F., Cohen, J.D., Norman, K.A., & Turk-Browne, N.B. (2012) Real-time decoding and training of attention. *Journal of Vision*, **12**, 377.

Dehaene, S. & Cohen, L. (2011) The unique role of the visual word form area in reading. *Trends in Cognitive Science*, **15**, 254–262.

Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193-222.

Eimer, M. (1996) The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, **99**, 225–234.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, **23**, 115–125.

Epstein, R. & Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature*, **392**, 598–601.

Ewbank, M.P., Schluppeck, D., & Andrews, T.J. (2005) fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *NeuroImage*, **28**, 268–279.

Friedman, J., Hastie, T., & Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.

Gazzaley, A., Cooney, J.W., McEvoy, K., Knight, R.T., & D'Esposito, M. (2005) Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience*, **17**, 507–517.

Goh, J.O.S., Siong, S.C., Park, D., Gutchess, A., Hebrank, A., & Chee, M.W.L. (2004) Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *Journal of Neuroscience*, **24**, 10223–10228.

Gorno-Tempini, M.L. & Price, C.J. (2001) Identification of famous faces and buildings: a functional neuroimaging study of semantically unique items. *Brain*, **124**, 2087–2097.

Greenough, W.T. & Black, J.E. (1992) Induction of brain structure by experience: Substrates for cognitive development. In Gunnar, M.R. & Nelson, C.E. (eds.) *Developmental Behavioral Neuroscience*, Vol 24, The Minnesota Symposium on Child Psychology, Lawrence Erlbaum Associates, Hillsdale, p. 155–200.

Haller, S. (2010) Real-time fMRI feedback training may improve chronic tinnitus. *European Radiology*, **20**, 696–703.

Hanson, S.J. & Schmidt, A. (2011) High-resolution imaging of the fusiform face area (FFA) using multivariate non-linear classifiers shows diagnosticity for non-face categories. *NeuroImage*, **54**, 1715–1734.

Hartmann, T., Schulz, H., & Weisz, N. (2011) Probing of brain states in real-time: Introducing the ConSole environment. *Frontiers in Psychology*, **2**, 36.

Haxby, J. V, Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, **293**, 2425–2430.

Haynes, J.-D. & Rees, G. (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, **7**, 523–534.

Hillis, A.E., Newhart, M., Heidler, J., Barker, P., Herskovits, E., & Degaonkar, M. (2005) The roles of the "visual word form area" in reading. *NeuroImage*, **24**, 548–559.

Hollmann, M., Rieger, J.W., Baecke, S., Lützkendorf, R., Müller, C., Adolf, D., & Bernarding, J. (2011) Predicting decisions in human social interactions using real-time fMRI and pattern classification. *PloS ONE*, **6**, e25304.

Hopf, J.M., Luck, S.J., Girelli, M., Hagner, T., Mangun, G.R., Scheich, H., & Heinze, H.J. (2000) Neural sources of focused attention in visual search. *Cerebral Cortex*, **10**, 1233–1241.

Jensen, O., Bahramisharif, A., Oostenveld, R., Klanke, S., Hadjipapas, A., Okazaki, Y.O., & Van Gerven, M. a J. (2011) Using brain-computer interfaces and brain-state dependent stimulation as tools in cognitive neuroscience. *Frontiers in Psychology*, **2**, 100.

Kamitani, Y. & Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, **8**, 679–685.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, **17**, 4302–4311.

LaConte, S.M., Peltier, S.J., & Hu, X.P. (2007) Real-time fMRI using brain-state classification. *Human Brain Mapping*, **28**, 1033–1044.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P. V, Nickerson, D., Mikiten, S.A., & Fox, P.T. (2000) Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, **10**, 120–131.

Lee, S., Ruiz, S., Caria, A., Veit, R., Birbaumer, N., & Sitaram, R. (2011) Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation: a multivariate investigation. *Neurorehabilitation and Neural Repair*, **25**, 259–267.

Li, W., Piëch, V., & Gilbert, C.D. (2004) Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, **7**, 651–657.

Liu, J., Harris, A., & Kanwisher, N. (2002) Stages of processing in face perception: an MEG study. *Nature Neuroscience*, **5**, 910–916.

Luck, S.J., Fan, S., & Hillyard, S.A. (1993) Attention-related modulation of sensory-evoked brain activity in a visual search task. *Journal of Cognitive Neuroscience*, **5**, 188–195.

Macevoy, S.P. & Epstein, R. A. (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, **19**, 943–947.

McCandliss, B.D., Cohen, L., & Dehaene, S. (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, **7**, 293–299.

McCarthy, G., Puce, A., Gore, J.C., & Allison, T. (1997) Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, **9**, 605–610.

Mulholland, T., Boudrot, R., & Davidson, A. (1979) Feedback delay and amplitude threshold and control of the occipital EEG. *Biofeedback and Self-Regulation*, **4**, 93–102.

Mur, M., Ruff, D. A, Bodurka, J., De Weerd, P., Bandettini, P. A, & Kriegeskorte, N. (2012) Categorical, yet graded – Single-image activation profiles of human category-selective cortical regions. *The Journal of Neuroscience*, **32**, 8649–8662.

Nelson, C. (2003) The development of face recognition reflects an experience-expectant and activity-dependent process. In Pascalis, O. & Slater, A. (eds), *The Development of Face Processing in Infancy and Early Childhood: Current Perspectives*, Nova Science Publishers, Hauppauge, pp. 79–97.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, **2011**, 156869.

Pierce, K., Haist, F., Sedaghat, F., & Courchesne, E. (2004) The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. *Brain*, **127**, 2703–2716.

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. (2008) TMS disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience*, 2**8**, 8929-8933.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, **17**, 1568–1573.

Puce, A., Allison, T., Gore, J.C., & McCarthy, G. (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, **74**, 1192–1199.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F.P., & Lau, H. (2011) Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, **14**, 1513–1515.

Reddy, L. & Kanwisher, N. (2006) Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology*, **16**, 408–414.

Reddy, L. & Tsuchiya, N. (2010) Reading the mind's eye: Decoding category information during mental imagery. *Neuroimage*, **50**, 818–825.

Rissman, J. & Wagner, A.D. (2012) Distributed representations in memory: insights from functional brain imaging. *Annual Review of Psychology*, **63**, 101–128.

Rosenbaum, R.S., Ziegler, M., Winocur, G., Grady, C.L., & Moscovitch, M. (2004) "I have often walked down this street before": fMRI studies on the hippocampus and other structures during mental navigation of an old environment. *Hippocampus*, **14**, 826–835.

Serences, J.T., Schwarzbach, J., Courtney, S.M., Golay, X., & Yantis, S. (2004) Control of object-based attention in human cortex. *Cerebral Cortex*, **14**, 1346–1357.

Seymour, K., Clifford, C.W.G., Logothetis, N.K., & Bartels, A. (2009) The coding of color, motion, and their conjunction in the human visual cortex. *Current Biology*, **19**, 177–183.

Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001) The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain*, **124**, 804–815.

Thesen, S., Heid, O., Mueller, E., & Schad, L.R. (2000) Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine*, **44**, 457– 465.

Tong, F. & Pratte, M.S. (2012) Decoding patterns of human brain activity. *Annual Review of Psychology*, **63**, 483–509.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**, 273–289.

Weiner, K.S. & Grill-Spector, K. (2012) The improbable simplicity of the fusiform face area. *Trends in Cognitive Science*, **16**, 251–254.

Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., & Mathiak, K. (2004) Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *Journal of Physiology – Paris*, **98**, 357–373.

Xi, Y.T., Xu, H., Lee, R., & Ramadge, P.J. (2011) Online kernel SVM for real-time fMRI brain state prediction, *2011 IEEE International Conference on Acoustics Speech and Signal Processing*, 2040–2043.

Yi, D.-J., Kelley, T. A., Marois, R., & Chun, M.M. (2006) Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. *Brain Research*, **1080**, 53–62.

Zotev, V., Krueger, F., Phillips, R., Alvarez, R.P., Simmons, W.K., Bellgowan, P., Drevets, W.C., & Bodurka, J. (2011) Self-regulation of amygdala activation using real-time fMRI neurofeedback. *PLoS ONE*, **6**, e24522.

Figures

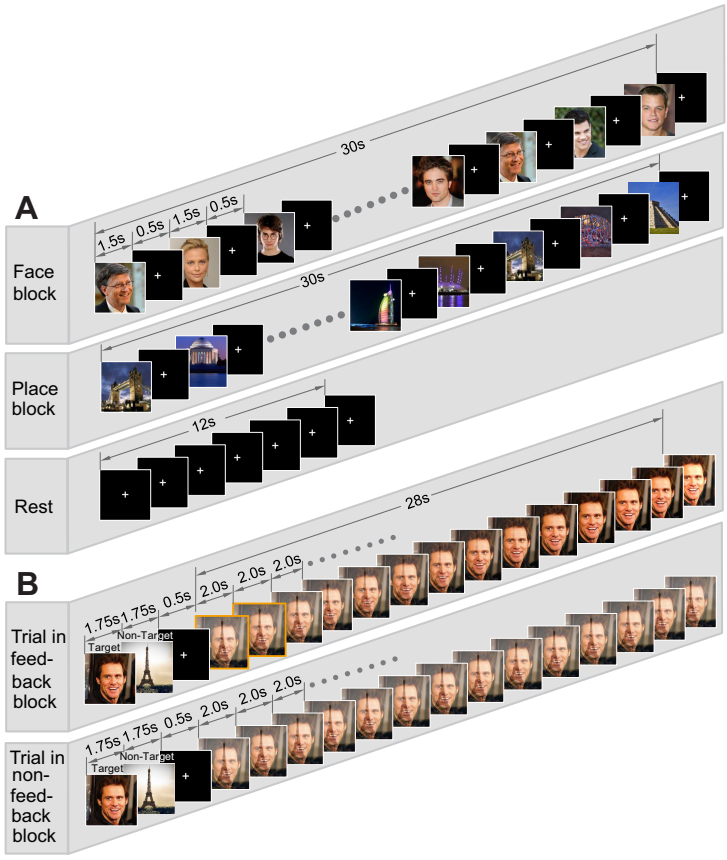

Figure 1. Experimental design. (A) Face, place and rest blocks in the training session. The first picture in each block was repeated at a random position in the block and subjects were required to press a button when they detected this repetition. (B) Example of a trial in a feedback and non-feedback block during the test session. In feedback trials, the hybrid image was continuously updated. However, in the first 2 TRs highlighted in yellow, the hybrid picture remains at 50/50 contrast due to pipeline delay. In non-feedback trials, the hybrid remained at 50/50 contrast at all times.

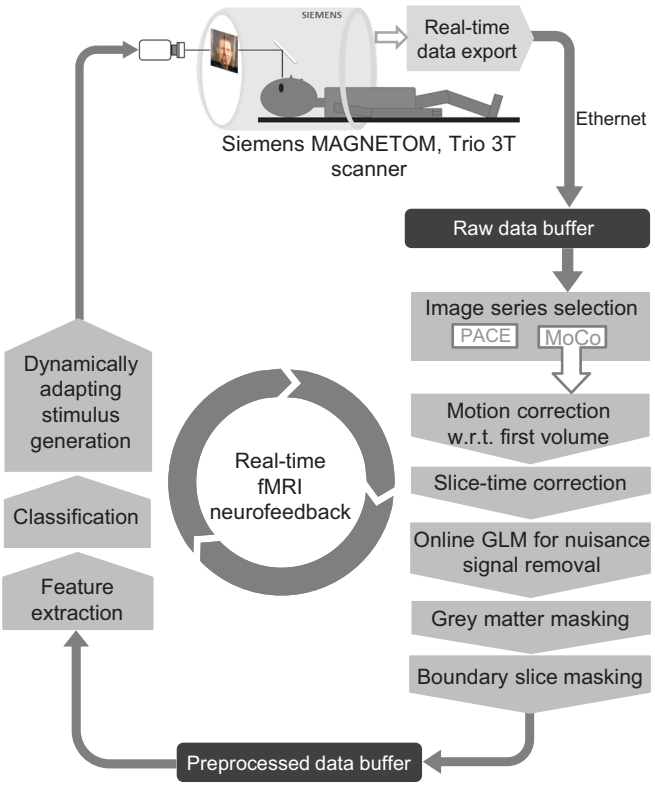

Figure 2. Donders real-time fMRI pipeline. See main text for details.

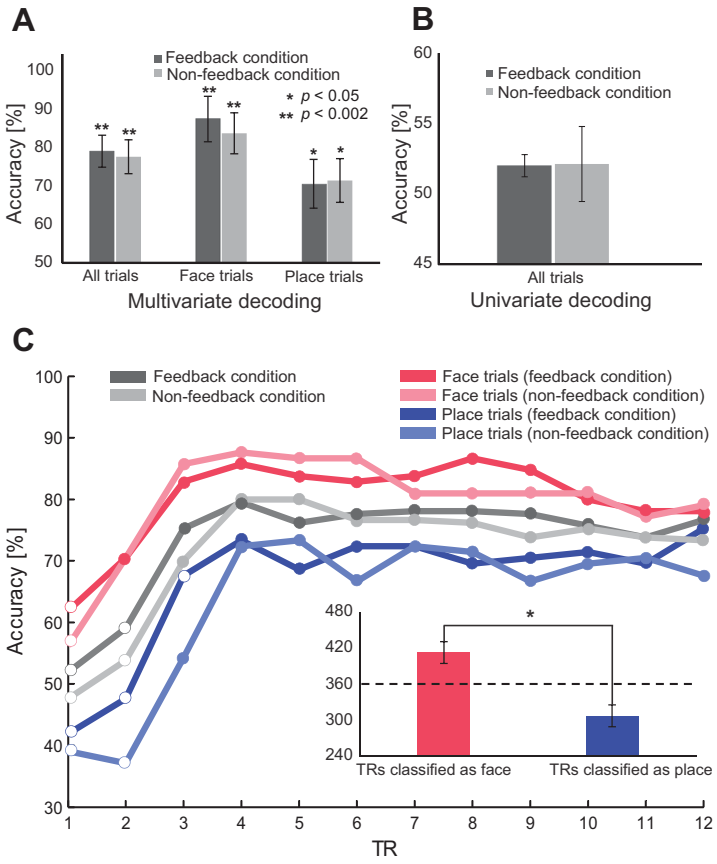

Figure 3. (A) Multivariate decoding performance for feedback and non-feedback conditions and different trial types within each of these conditions. The error bars represent standard error of the mean. (B) Univariate decoding performance for feedback and non-feedback conditions. The error bars represent standard error of the mean. (C) Decoding performance as a function of TR. The open circles show that the decoding performance was found to be statistically insignificant at that point whereas the filled circles shows that the decoding performance was found to be statistically significant ($p < 0.05$). The inset shows the relative number of TRs classified as face or place across the group. The dashed line represents the ideal unbiased distribution of face and place TRs.

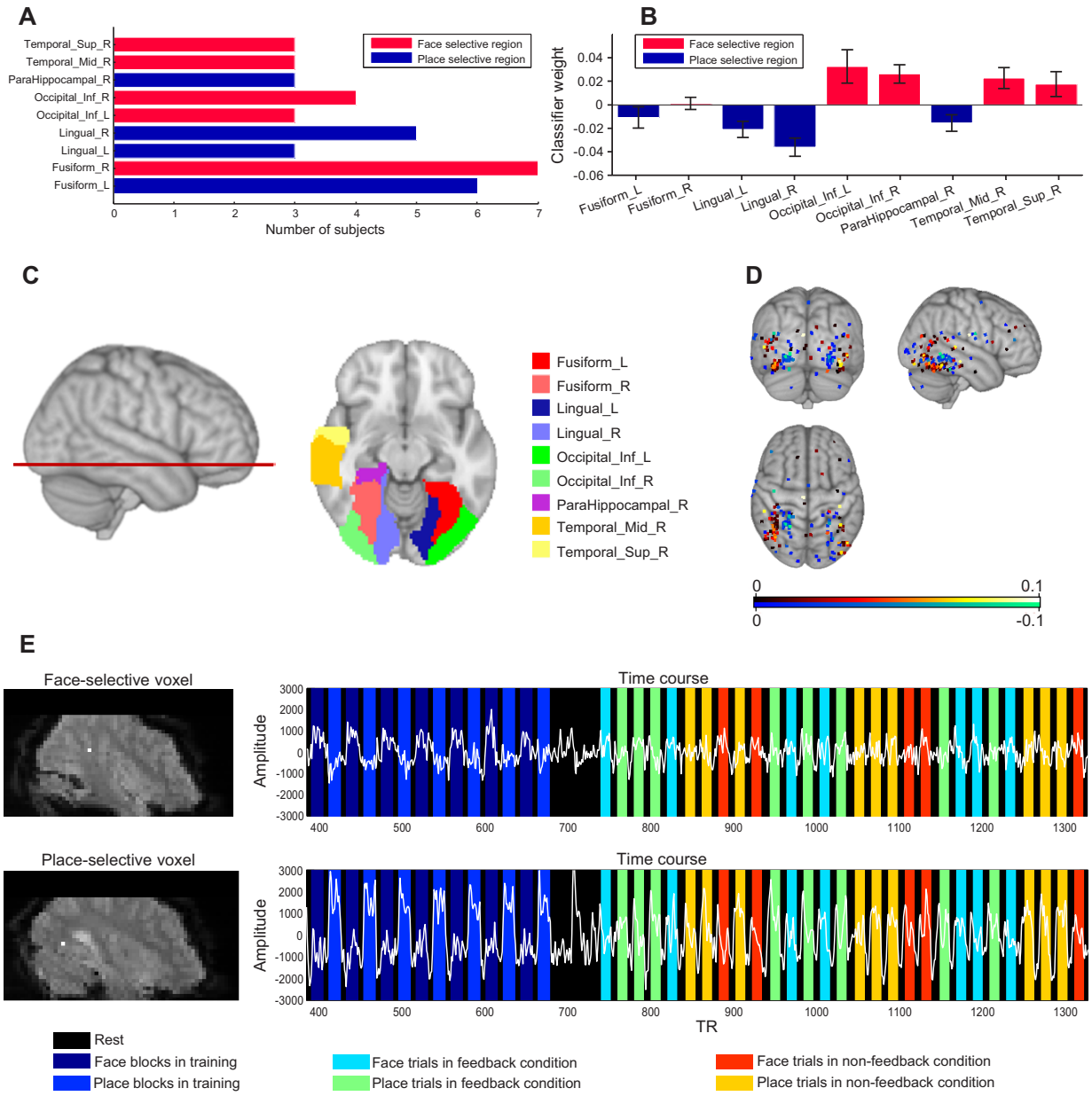

Figure 4. (A) Nine anatomical regions used by the classifier during training. These nine regions were activated in three or more subjects. (B) Classifier weight for all nine regions. Positive weights are assigned to the face-selective regions and negative to the place-selective regions. The error bars represent standard error of the mean. (C) An anatomical view of the nine regions mask overlaid on the MNI152 template. (D) Sparsely distributed classifier weights across the group overlaid on the MNI152 template. (E) The top plot shows the time course of one face-selective voxel in right superior temporal cortex (MNI coordinates 46, -44, 14 and highlighted in white) as selected by the classifier during training. The bottom plot shows the time course of one place-selective voxel in left fusiform gyrus (MNI coordinates -32, -68, -11 and highlighted in white) as selected by the classifier during training.

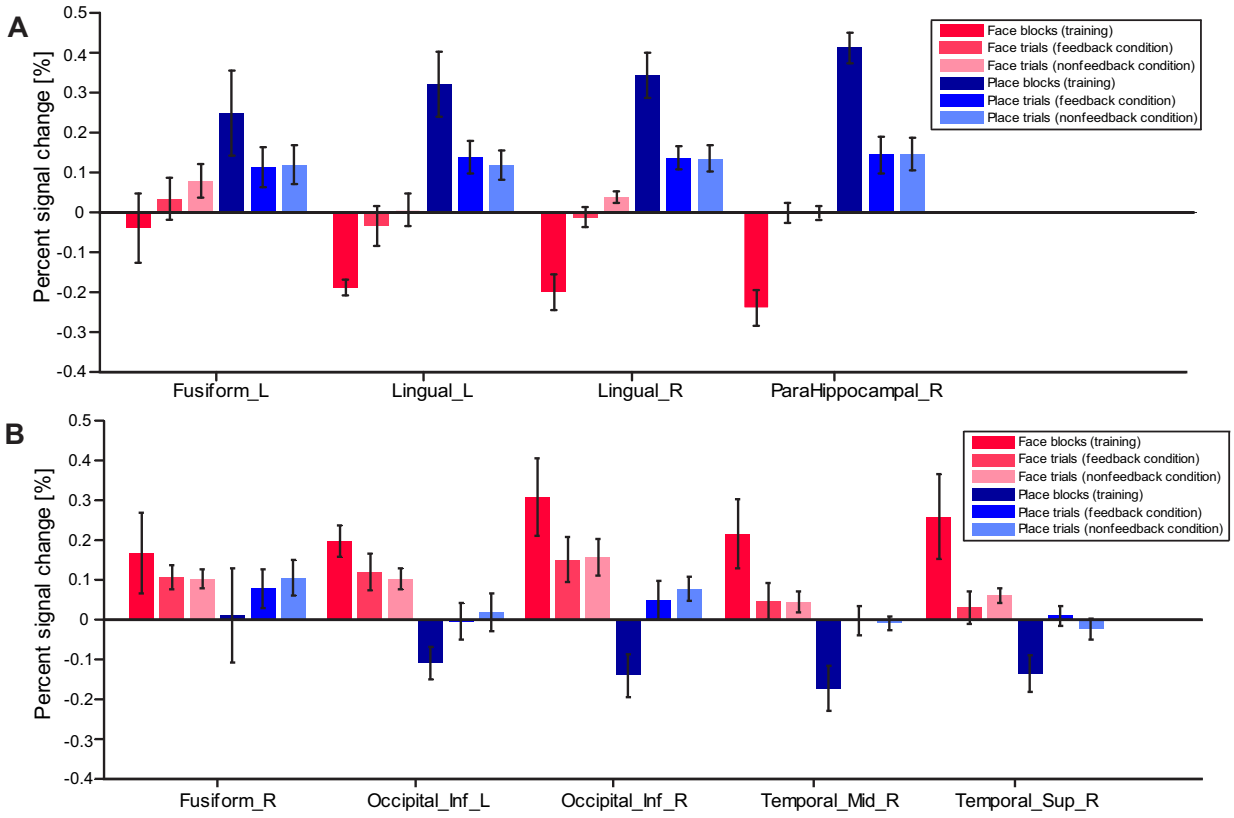

Figure 5. ROI analysis results for all nine regions activated across the group. A region was regarded as face-selective if the percent signal change in face blocks was greater than the percent signal change in place blocks during training. The opposite criterion was applied for place-selective regions. (A) ROI analysis for all place-selective regions. (B) ROI analysis for all face-selective regions. The error bars indicate standard error of the mean.