

COGNITIVE NEUROSCIENCE

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

Adnan M. Niazi,^{1,2} Philip L. C. van den Broek,¹ Stefan Klanke,¹ Markus Barth,¹ Mannes Poel,² Peter Desain¹ and Marcel A. J. van Gerven¹

¹Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, NL 6500 HE, Nijmegen, The Netherlands

²Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, NL 7500 AE, Enschede, The Netherlands

Keywords: categorization, multivariate decoding, object-based attention, real-time fMRI

Abstract

Visual attention is used to selectively filter relevant information depending on current task demands and goals. Visual attention is called object-based attention when it is directed to coherent forms or objects in the visual field. This study used real-time functional magnetic resonance imaging for moment-to-moment decoding of attention to spatially overlapped objects belonging to two different object categories. First, a whole-brain classifier was trained on pictures of faces and places. Subjects then saw transparently overlapped pictures of a face and a place, and attended to only one of them while ignoring the other. The category of the attended object, face or place, was decoded on a scan-by-scan basis using the previously trained decoder. The decoder performed at 77.6% accuracy indicating that despite competing bottom-up sensory input, object-based visual attention biased neural patterns towards that of the attended object. Furthermore, a comparison between different classification approaches indicated that the representation of faces and places is distributed rather than focal. This implies that real-time decoding of object-based attention requires a multivariate decoding approach that can detect these distributed patterns of cortical activity.

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 our 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 *et al.* (2010), which employed single-unit recordings, neurons coding for Marilyn Monroe were identified. These neurons fired selectively when subjects were presented with a composite picture of Marilyn Monroe and Josh Brodin 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 Brodin). The study indicates that despite competing bottom-up sensory input, firing rates in medial temporal lobe neurons can be voluntarily regulated to reflect object-based selective attention. Studies using functional magnetic resonance imaging (fMRI), electroencephalography and magnetoencephalography 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; O'Craven *et al.*, 1999; 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. One of the first studies that employed MVPA to read subjective contents of the human brain using fMRI has nicely demonstrated this (Kamitani & Tong, 2005). The study showed that a classifier that 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

Correspondence: A. M. Niazi, as above.
E-mail: adnaniazi@gmail.com

Received 5 March 2013, accepted 4 October 2013

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 conjunctions 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 object category of an attended stimulus can be decoded non-invasively in real-time when stimuli from two different categories are presented simultaneously. More specifically, we examined whether a classifier trained on separately presented pictures of faces and places can be used to decode the attended object 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 tested if object-based attention can bias the neural patterns in face- and place-selective areas towards the attended category, and if these differentiating activity patterns can be picked up on a moment-to-moment basis 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 paralysed 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).

Previous studies have shown that pictures of faces and places invoke spatially distinct and dissociable cortical regions, namely, fusiform face area (FFA) for faces and parahippocampal place area 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 faces and places from these regions call for a multivariate decoding approach that can detect these overlapping and distributed neural patterns. 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; Anderson *et al.*, 2011; Hollmann *et al.*, 2011; Lee *et al.*, 2011; Xi *et al.*, 2011; DeBettencourt *et al.*, 2012). Moreover, the whole-brain decoder is highly suited for real-time fMRI because it automatically identifies sparse and distributed patterns of activity that are representation-specific. The employed method is also computationally fast, such that the entire experiment with both classifier training and testing can be conducted in a single non-stop session.

In order to examine which activity patterns were related to successful classification, we also assessed decoding performance when the feature space was restricted to only those voxels activated during a general linear model (GLM). For this purpose, we retrained the classifier *post hoc* on a restricted feature space of only those clusters activated in a GLM on the localizer task. Using this approach, we examined whether multivariate or average activity patterns within each cluster drove classifier performance. Finally, to assess if representation of object-based attention is distributed across multiple brain regions, we applied multivariate decoders to individual clusters activated in the GLM. If the object representation is distributed across various brain regions, then these individual clusters should yield poorer decoding performance compared with whole-brain or GLM-restricted decoders.

Because brain state predictions are available for every scan in real-time fMRI, these online detected brain states can be used as neurofeedback 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 impacted subjects' task performance, we conducted each attention trial twice, once with fMRI neurofeedback and once without it. However, due to the lack of statistically significant differences between feedback and non-feedback conditions, we will focus primarily on the non-feedback condition and refer the reader to the Supporting Information for a detailed analysis of the feedback condition. Results for both the feedback and non-feedback conditions showed that object-based attention can be successfully decoded within a real-time fMRI paradigm.

Materials and methods

Subjects

Seven subjects (six males, one female) with an average age of 23.4 years (SD = 4.6) participated in the study. All participants had normal vision, and received either monetary compensation or study credits for their participation. The study was approved by the local ethics committee (Commissie Mensgebonden Onderzoek Regio Arnhem-Nijmegen) and conformed with The Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the *British Medical Journal* (18 July 1964). Subjects gave written informed consent before the experiment. 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%.

Stimuli

The stimulus set consisted of color 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 with unfamiliar faces and places, respectively (Shah *et al.*, 2001; Pierce *et al.*, 2004; Rosenbaum *et al.*, 2004). All pictures were 450 × 450 pixels with a resolution of 95.987 pixels/inch and subtended a visual angle of 8°. The stimulus set was not corrected for luminance or spatial frequency.

Experimental protocol

Subjects were thoroughly briefed before the experiment to avoid any verbal communication during the real-time fMRI run. Video recordings of all experimental conditions were shown and the task was verbally explained by the experimenter with the help of these videos. No instructions were given to maintain a specific gaze direction. Subjects were allowed to close their eyes during the 12-s rest periods between blocks/trials, but were instructed to open their eyes a few seconds before this rest period was over.

Experimental design

The experiment consisted of two phases: a training phase (also called localizer) in which a classifier was trained on the cortical activity patterns induced by faces and places; and a test phase in which the classifier was used to decode the category of the attended picture in a hybrid of a simultaneously presented face and place.

The training phase consisted of 15 × 30-s blocks of face pictures interleaved with 15 × 30-s blocks of place pictures with 12 s rest intervals between consecutive blocks. Within each block, 15 pictures

were presented, and the first picture was repeated at a random position in the block. Subjects had 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 them actively engaged in the task throughout the training phase. Early repeats of the first picture were avoided by constraining it to repeat after three other pictures had been presented. Subjects were advised to attend to all pictures in a block regardless of when the first picture was repeated. Each picture within a block was presented for 1.5 s followed by a 0.5-s fixation period, as shown in Fig. 1A. All 14 pictures in each block were unique and used nowhere else in the experiment. The entire training phase took 22 min to complete.

In the test phase, 15 hand-picked pairs of transparently overlapped faces and places were used (see Figs S1, S2 and Movie S1), and subjects had to attend to either face or place items depending on the cue. Thirty trials were collected in the non-feedback condition, half of which had face as target (attend-face trials) and the remaining half of which had place as target (attend-place trials). Every trial started with presentation of the target and non-target cue pictures for 1.75 s each, followed by a 0.5-s fixation period. Cue pictures were labeled with either of the words 'Target' and 'Non-target', and the order of presentation of these cues was counterbalanced across subjects. After cueing, a hybrid image of the target and non-target picture was shown for 12 repetition times (TRs), and subjects had to attend to the target picture while ignoring the non-target picture (Fig. 1C and D). The relative mix of face and place pictures in the hybrid was not changed in the non-feedback condition, but it

was changed for the feedback condition (for full details about the feedback condition please consult the Supporting Information).

MRI acquisition parameters

Experiments were performed at the Donders Institute for Brain, Cognition and Behaviour using a Siemens MAGNETOM Tim TRIO 3.0 Tesla 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 $1 \times 1 \times 1$ mm). Then a real-time fMRI run was initiated and functional images were acquired using a single-shot gradient echo planar imaging sequence (TR/TE = 2000/30 ms; flip angle = 75° ; voxel size = $3 \times 3 \times 3.3$ mm; distance factor = 10%) with prospective acquisition correction (PACE) to minimize effects of head motion during data acquisition (Thesen *et al.*, 2000). Twenty-eight ascending axial slices were acquired, oriented at about 30° relative to the anterior-posterior commissure.

Real-time data export and preprocessing

During the real-time fMRI run, all functional scans were acquired using a modified scanner sequence and in-house software that sent each acquired scan over Ethernet to another computer, which stored them 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, USA) preprocessing pipeline.

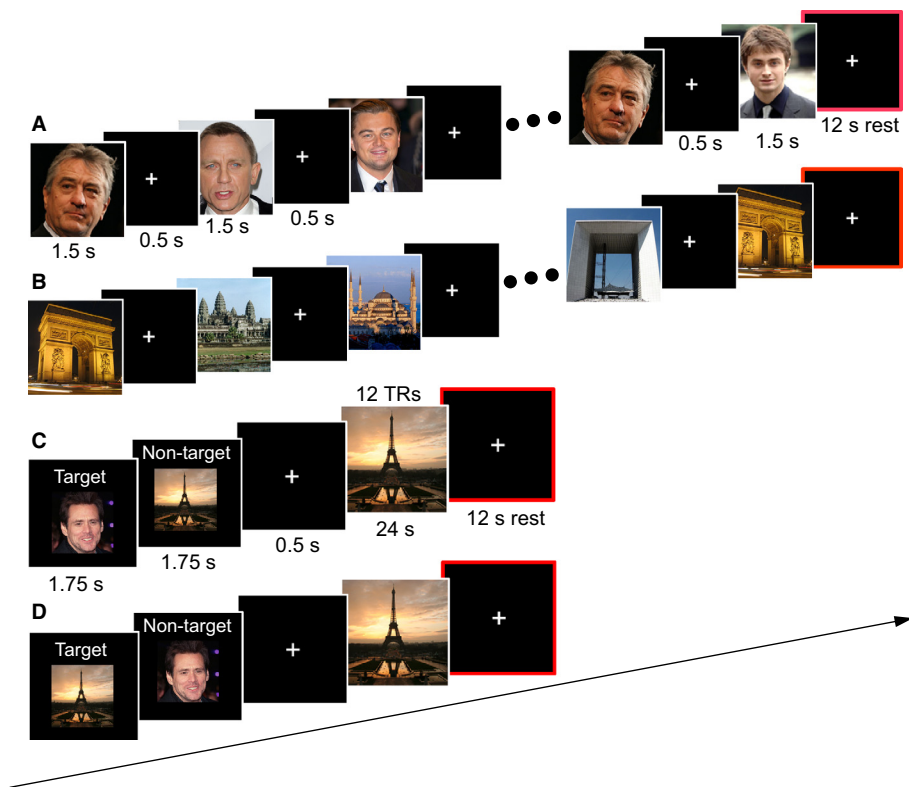


FIG. 1. Experimental design. (A) Face block in training phase in which pictures of famous people were presented. (B) Place block in training phase in which pictures of famous landmarks were presented. (C) Attend-face trial during decoding phase in the non-feedback condition. Each trial started with target and non-target cues. Participants then attended to the face object while ignoring the place object in a hybrid of transparently presented face and place pictures. (D) Attend-place trial in the non-feedback condition. Same as attend-face trials, except the subjects now had to attend to the place object while ignoring the face object. The order of presentation of target and non-target was counterbalanced across subjects. Note: copyrighted pictures used in the original experiment have been substituted in the above graphic by non-copyrighted lookalikes.

The first preprocessing step involved selecting one of the two image series generated by the scanner sequence: the PACE series of images that is only prospectively corrected and the MoCo (motion-corrected) series that is both prospectively and retrospectively corrected (Thesen *et al.*, 2000). We used the MoCo series of images as it contained the least residual motion. Then scans were slice-time corrected, followed by retrospective motion correction using an online rigid-body transformation algorithm with six degrees of freedom. This was done to remove any residual motion in the MoCo series. Then a recursive least-squares GLM 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. Next, we removed white matter and cerebral spinal fluid voxels from all scans using a gray matter mask, which was obtained from high-resolution anatomical images using SPM8s (Wellcome Department of Cognitive Neurology, Queens Square, London, UK) unified segmentation-normalization procedure (Ashburner & Friston, 2005). Volumes were resliced to the resolution of the functional scans using the first acquired functional scan as reference. After gray matter masking, top and bottom slices in each scan were masked to avoid using the bad voxels in these slices formed during online retrospective motion correction. Each scan, now fully preprocessed, was saved in a FieldTrip preprocessed data buffer. The entire real-time fMRI pipeline is shown in Fig. 2.

Feature extraction and classification

Once preprocessed, scans were then used for training and decoding. To train the classifier, scans collected in the training phase were shifted by 6 s to account for the hemodynamic delay. Then all scans corresponding to the 12-s rest periods between consecutive face and

place blocks were discarded. The remaining scans were labeled and used to train the decoder. We used logistic regression in conjunction with an elastic net regularizer. The elastic net regularization shrinks and selects 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:

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

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, yielding a sparse solution. For the regularization parameter λ , a regularization path was calculated with 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. For decoding object-based attention, each of the 12 scans in every trial was individually classified. The classification threshold was set to 0.5. A prediction probability below 0.5 indicated attention to the place object and above 0.5 indicated attention to the face object. During the actual real-time fMRI run, a whole-brain decoder (MVA-W) was used. That is, all gray matter voxels in every volume were used during training and decoding.

Pattern analysis

To compare the whole-brain decoding approach to a GLM-based approach, we retrained the classifier offline on a restricted feature space of only those voxels that were detected in a GLM applied to the localizer. The GLM for this decoder was carried out on the training data and contained two regressors corresponding to the face and place blocks, and six rigid-body motion parameters as nuisance covariates. Two contrasts, faces > places and places > faces were formed to find voxels that responded strongly to faces and places, respectively. For each subject, these statistical images were assessed for cluster-wise significance using a cluster-defining threshold of $P = 0.01$. The 0.05 FWE-corrected critical cluster size was found using Newton–Raphson search (Nichols & Hayasaka, 2003) and ranged from 19 to 21 voxels across the group. We applied this GLM-based decoder in two ways. First, we used the voxels within all identified clusters as input to the elastic net classifier (GLM-restricted multivariate analysis; MVA-G). Second, we used the average time-series within each cluster as input the elastic net classifier (MVA-T). This allowed us to compare the impact of using multivariate vs. univariate patterns within each cluster. Additionally, we performed multivariate decoding on each individual cluster found in the GLM to examine if decoding of the attended object category is based on either

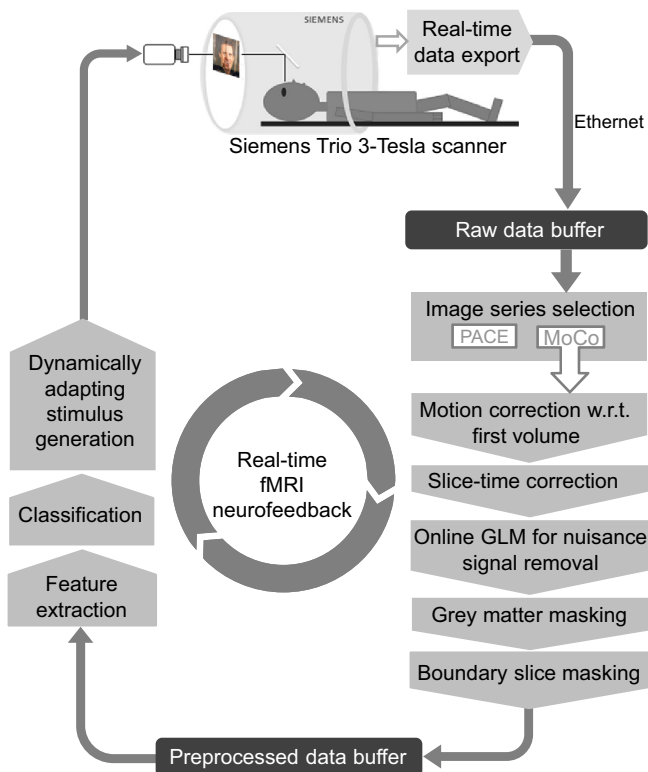


FIG. 2. Donders real-time fMRI pipeline. See main text for details.

localized or distributed patterns of cortical activation patterns. In cluster-wise decoding (MVA-C), time-series of all voxels in a cluster were averaged and then used for training and decoding. This analysis was repeated for each cluster found in each subject. Hence, a separate decoder was trained and tested for every cluster.

Furthermore, we also computed the anatomical label of voxels used by the decoders by grouping and labeling them using a subject-specific automatic anatomic labeling mask (Tzourio-Mazoyer *et al.*, 2002). We refer to these groups of classifier voxels with the same anatomical labels as regions. A region may contain one or more voxels that may or may not be spatially adjacent, but crucially each voxel in a region has the same anatomical label. The same procedure was then repeated for all subjects. Any region not activated in at least three subjects was dropped from further analysis. We then calculated average percent signal change for attend-face and attend-place trials in voxels in each of these groups.

Finally, to examine how the blood oxygen level-dependent (BOLD) signal evolved during an attention trial in MVA-W, we calculated percent signal change as a function of TR in face- and place-selective voxels for attend-face and attend-place trials. Face-selective voxels were defined as those voxels that were assigned positive weights by the classifier, whereas place-selective voxels were assigned negative weights.

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 summed log probability for the target picture ($\sum_{k=1}^{12} \log P(\text{Target}|\text{Scan}_k)$) exceeded the summed log probability of the non-target picture ($\sum_{k=1}^{12} \log P(\text{Non-Target}|\text{Scan}_k)$) for all 12 scans in a trial. Additionally, decoding accuracy was also calculated as a function of time (TR) within each trial to investigate how it evolved over the course of the trial duration. Decoding accuracy at a given TR was defined as the percentage of successfully decoded scans at that TR across the group. Furthermore, because the non-feedback condition contained attend-face and attend-place trials, performance for each of these trial types was calculated separately as well.

Behavioral testing

A behavioral test was conducted *post hoc* to assess the familiarity asymmetry of face and place pictures used in this study. In this web-based test, participants had to rank the familiarity of a picture on a five-point scale. In this way, all 589 pictures used in the study were ranked. In total, 97 participants (25 female) with an average age of 29.6 years ($SD = 7.1$) took part in this task. Thirty-two participants completed the test, while the remaining participants dropped out after ranking 96 pictures on average.

Results

In this study we tested if object-based attention to simultaneously presented faces and places could be decoded on a moment-to-moment basis using a whole-brain decoder (MVA-W) trained on pictures of separately presented faces and places. We also compared a whole-brain decoder with a GLM-restricted decoder (MVA-G). Furthermore, we studied if decoding is based on average time-series across clusters (MVA-T), or driven by multivariate activity patterns within individual clusters (MVA-C).

Comparison of decoding performance

We used a one-way ANOVA to test for differences in decoding performance among the four decoders. Decoding performance varied significantly (Fig. 3) across the four different decoders, $F_{3,24} = 9.04$, $P = 0.000346$. A Tukey test indicates that MVA-W ($M = 77.6$, $SD = 11.6$) was decoded significantly better than MVA-C ($M = 56.1$, $SD = 3.74$), $P = 0.001$. Similarly, MVA-G ($M = 79$, $SD = 9.75$) was decoded significantly better than MVA-C ($M = 56.1$, $SD = 3.74$), $P = 0.001$. No statistically significant difference was found between MVA-W, MVA-G and MVA-T ($M = 68.6$, $SD = 9.97$), though a trend towards significance could be observed. No statistically significant difference was found between MVA-C and MVA-T.

Taken together, these results suggest that whole-brain multivariate decoding and GLM-restricted decoding perform comparably. Furthermore, because MVA-W and MVA-G both performed significantly higher than MVA-C, it indicates that decoding depends on distributed patterns of cortical activity. Finally, lower decoding performance for MVA-T compared with MVA-W and MVA-G suggests that multivariate patterns of activity distributed across clusters drive decoding performance.

Whole-brain multivariate decoding

To further examine online decoding results using MVA-W, we tested how its decoding performance evolved during the trials. The results of a TR-by-TR analysis in the non-feedback condition (Fig. 4A) showed that decoding accuracy followed BOLD activity, increasing in the initial 6 s and leveling off afterwards. Moreover, attend-face trials were decoded with an accuracy of 84% ($SD = 14.3$), whereas attend-place trials were decoded with an accuracy of 71% ($SD = 15.3$), respectively. A paired-samples *t*-test failed to reveal a statistically significant ($t_6 = 1.8117$, $P = 0.12$) difference between attend-face and attend-place trials (Fig. 4B). However, a statistically significant asymmetry was found for the familiarity of face and place stimuli in the *post hoc*

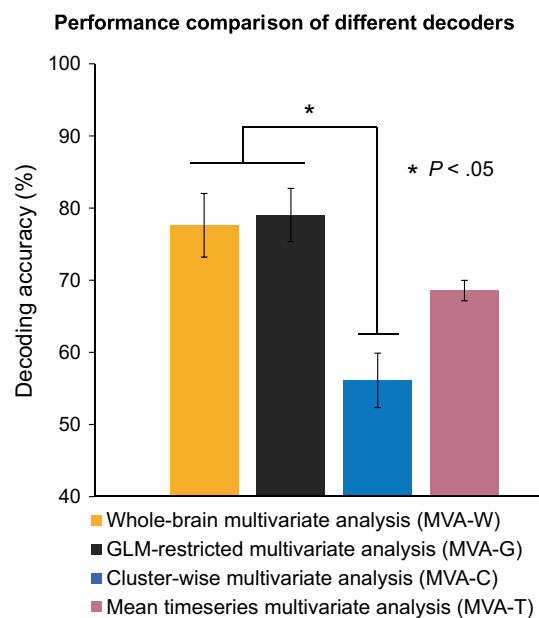


FIG. 3. Comparison of the decoding performance for four different decoding techniques in the non-feedback condition. Both the whole-brain and GLM-restricted decoders performed significantly higher than the cluster-wise multivariate decoder ($P < .05$). Error bars represent standard error of the mean. Note: only MVA-W analysis was performed online, while the remaining analyses were done offline.

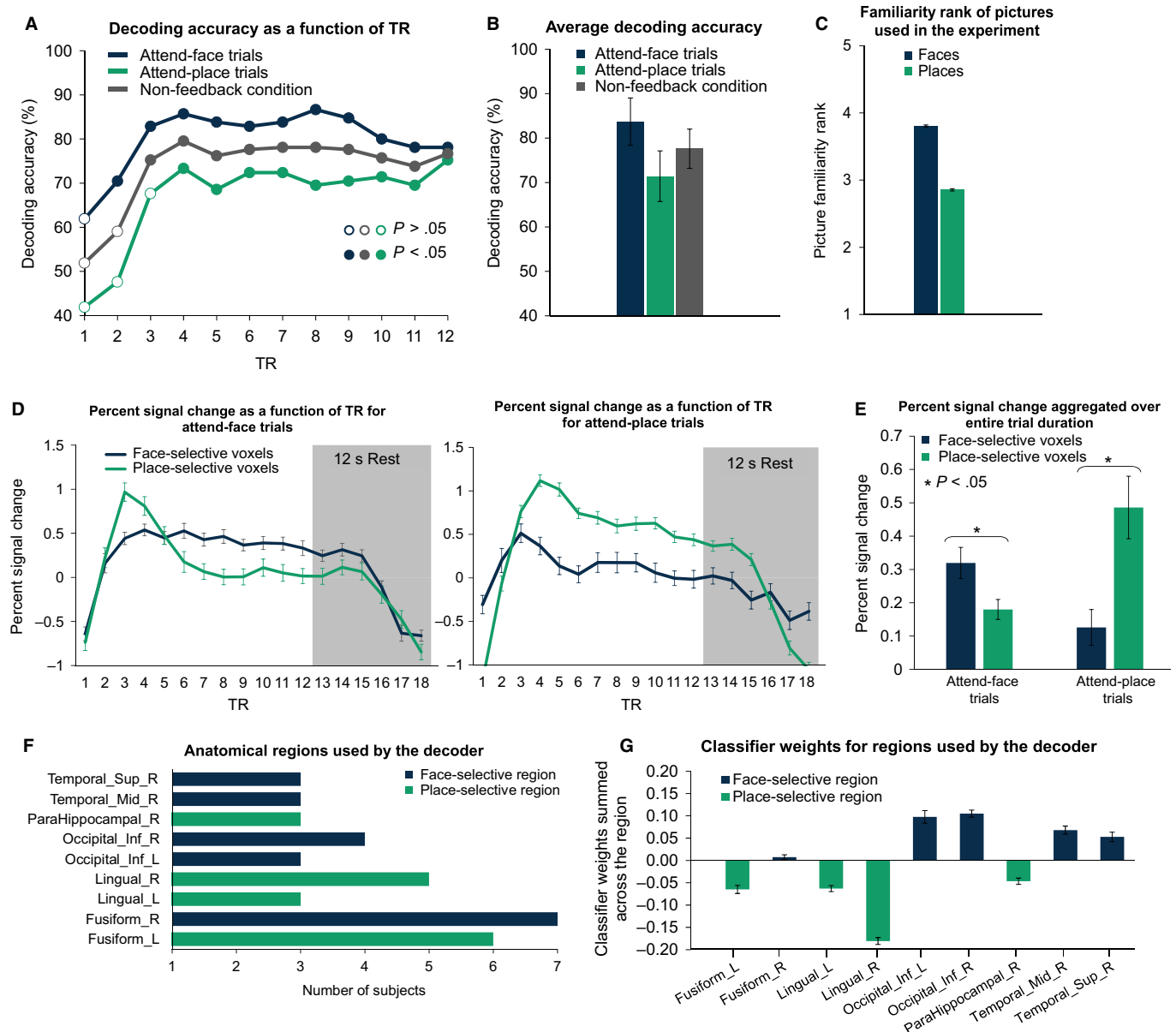


FIG. 4. Whole-brain multivariate analysis (MVA-W). (A) Decoding performance as a function of TR for the non-feedback condition and the two trial types constituting it. Above chance-level decoding is indicated by filled circles, whereas below chance-level decoding is marked by empty circles. (B) Average decoding accuracy. (C) Subjective familiarity rank of the face and place pictures used in the experiment. (D) Percent signal change at every TR for attend-face and -place trials in face-selective and place-selective regions used by the classifier. (E) Average percent signal change in face- and place-selective regions for attend-face and -place trials. (F) Anatomical grouping of face- and place-selective voxels recruited by the decoder, and number of subjects these voxels were activated in. (G) Classifier weights. The plotted weights represent the sum of weights of voxels used by the classifier in each region averaged across the group. Error bars represent standard error of the mean.

behavioral test. A paired-samples *t*-test showed that subjects ranked faces ($M = 3.805$, $SD = 0.015$) more familiar than places ($M = 2.85$, $SD = 0.016$), $t_{10668} = 43.19$, $P = 0$.

Additionally, we tested how BOLD signal varied for attend-face and attend-place trials in voxels used by the decoder (Fig. 4D and E). A two-tailed paired-samples *t*-test on percent signal change showed that face-selective voxels responded more strongly to attend-face trials ($M = 0.319$, $SD = 0.123$) than to attend-place trials ($M = 0.179$, $SD = 0.142$), $t_6 = 2.468$, $P = 0.048$. Likewise, place-selective voxels responded strongly to attend-place trials ($M = 0.125$, $SD = 0.079$) compared with attend-face trials ($M = 0.485$, $SD = 0.248$), $t_6 = -4.84$, $P = 0.0028$. This shows that

category-specific voxels responded strongly to the preferred category than to the non-preferred category.

Anatomical grouping of voxels used by the decoder showed that the selected voxels were distributed across 31 distinct brain regions across the subjects (see Fig. S4 for a list of all these regions). Regions not activated in at least three subjects were excluded from further analysis. This left only nine brain regions, as shown in Fig. 4F. 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 were assigned positive weights and responded strongly

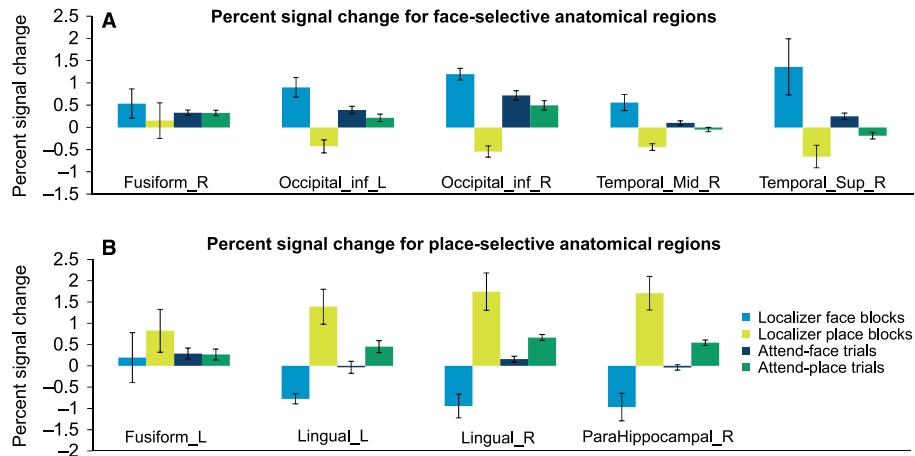


FIG. 5. Percent signal change for voxels selected by the MVA-W classifier in different anatomical regions. (A) Percent signal change for face-selective regions. Voxels in these regions responded more strongly to faces than places in the localizer. (B) Percent signal change for place-selective regions that responded more strongly to places than faces in the localizer. Error bars represent standard error of the mean.

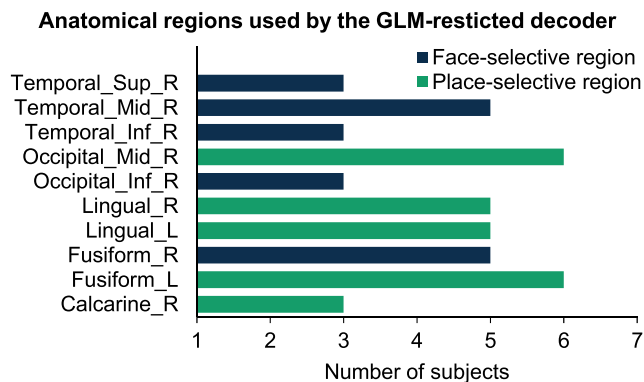


FIG. 6. Anatomical regions used by the GLM-restricted decoder (MVA-G) and the number of subjects for which these regions were activated.

to faces during the localizer task (Fig. 5A). Hence, these were labeled as face-selective regions. Left fusiform gyrus, bilateral lingual gyri and right parahippocampal gyrus were assigned negative weights and were more responsive to place stimuli in the localizer task (Fig. 5B), and therefore labeled as place-selective regions. The classifier weights summed across all subjects for all these regions are shown in Fig. 4G.

GLM-restricted multivariate decoding

The MVA-G model not only gave decoding performance similar to that of MVA-W, but also recruited voxels from the same regions as were used in the MVA-W model. While nine regions were used in the MVA-W decoding model, 10 regions were recruited in the MVA-G model (Fig. 6), out of which six were the same as that in the MVA-W decoder. Percent signal change across these regions is shown in Fig. 7. The fact that MVA-G identified a number of different regions compared with MVA-W may be explained by the fact that these regions contain redundant information that is ignored by MVA-W due to the sparseness constraint imposed by the elastic net classifier. MVA-T also gave above-chance classification performance, though the observed trend was that it was generally lower than MVA-G.

Cluster-wise multivariate decoding

Thirty-four distinct clusters were found across the group in the individual GLM. Those clusters that were not activated in three or more subjects were removed from further analysis. Decoding performance for the remaining 12 clusters is summarized in Fig. 8. As stated earlier, the average decoding performance for MVA-C was found to be significantly lower than MVA-W and MVA-G. These results suggest that within each small cluster not much discriminable information is present about the attended category. However, if decoding is extended to multiple brain regions such as that in MVA-W or MVA-G, then distributed patterns of cortical activation can help increase the decoding performance dramatically.

Discussion

In this study, we decoded object-based attention to transparently overlapped faces and places in real time on a single TR basis using a whole-brain decoder trained on separately presented pictures of faces and places. A decoding accuracy of 77.6% was obtained for the non-feedback condition, which is high considering that decoding was performed on a single TR without averaging multiple scans.

We also tested if neurofeedback of scan-by-scan brain state classification results can improve decoding performance by using a feedback condition in which the relative mix of the face and place picture was adjusted depending on classification results. However, neurofeedback did not significantly improve decoding performance (see Supporting Information). When we analysed the results of TR-by-TR decoding performance, we did not observe an improvement in accuracy over time for feedback trials. This contradicted our expectation that neurofeedback of the attended stimulus in the form of its enhancement in the hybrid picture would result in higher decoding performance. From a purely perceptual point of view, enhancement of the target picture should make classification easier as an enhanced target picture would resemble more closely the neural patterns that the classifier was originally trained on. To examine why no improvement in decoding accuracy was observed in the feedback condition, we computed classifier prediction probability as a function of TR (see Supporting Information). We indeed observed an increase in the prediction probability of attended stimuli for successful feedback trials. However, we also observed a decrease in the

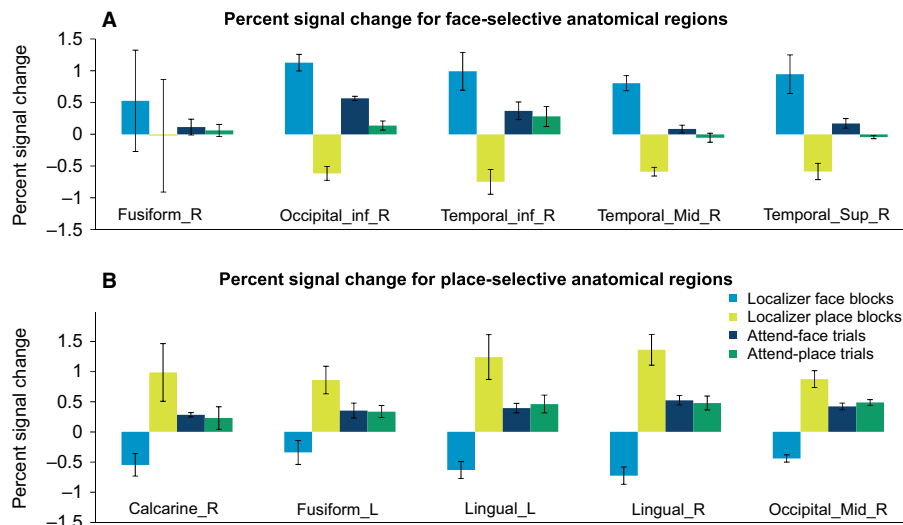


FIG. 7. Percent signal change for voxels selected by the MVA-G classifier in different anatomical regions. (A) Percent signal change for face-selective regions. Voxels in these regions responded more strongly to faces than places in localizer. (B) Percent signal change for place-selective regions that responded more strongly to places than faces in the localizer. Error bars represent standard error of the mean.

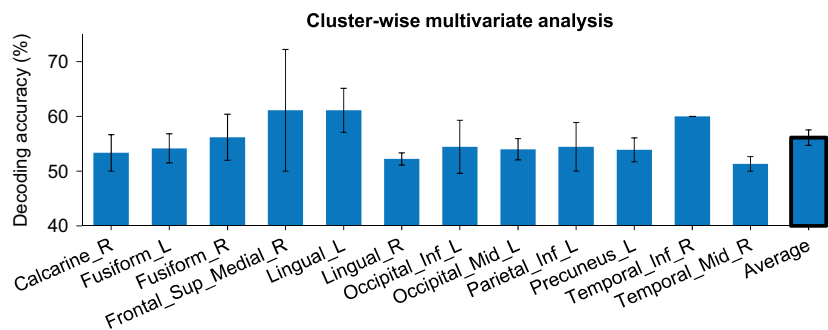


FIG. 8. Cluster-wise multivariate analysis (MVA-C). These clusters were detected in a GLM on the localizer. Any cluster not activated in more than two subjects is not shown in this graphic. Error bars represent standard error of the mean.

prediction probability for unsuccessful feedback trials. This is because in the feedback condition, visibility of the attended picture increased as a trial progressed, irrespective of whether it was the target or distractor picture. As a result, when both successful and unsuccessful trials were combined and the TR-by-TR prediction probabilities were computed again, we did not observe any difference between feedback and non-feedback conditions. Hence, no significant difference between feedback and non-feedback conditions was observed.

A number of other design choices may have affected performance in the feedback condition. First, because feedback and non-feedback trials were conducted in interleaved mini-blocks, it might have weakened any learning effect as subjects would not have been able to discover a 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. Second, the duration of feedback was chosen to be 12 TRs (24 s) as a compromise between the number of trials and the experiment duration. This might have been too short for any significant strategy learning. Previous real-time studies have used trial durations ranging from 15 to 60 s conducted over the course of multiple days (see Weiskopf *et al.*, 2005 for a review). Third, 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. Fourth, adjusting the relative contribution of attended and unattended pictures based on decoder output did not allow us to dissociate between the effect of neurofeedback and the effect of change in BOLD signal due to change in the perceptual input. Future neurofeedback designs should avoid changing object properties by using a more abstract neurofeedback such as adjusting the color of the background surrounding the hybrid picture depending on the results of the decoding. Finally, a decoder trained on separately presented pictures of faces and places might not be the optimal way of investigating the effects of neurofeedback. This is because a decoder trained on faces and places will recruit only those regions that it finds useful for distinguishing between face and place pictures. Presenting decoder output as neurofeedback to the subjects may have little impact on their task performance because the regions that respond to neurofeedback may not be incorporated in the decoding model trained on just faces and places. Hence, even if the subject's brain is responding to neurofeedback, the decoder may be unable to detect it. Therefore, it is necessary that future studies using MVPA-generated neurofeedback could aim to incorporate the brain regions responsible for processing feedback into the model.

In case of whole-brain decoding, nine regions were consistently used by the classifier to drive the predictions. Among these regions was 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). The left fusiform gyrus may be showing more activity for place blocks than for face blocks because 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. The 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 places, respectively (Aguirre *et al.*, 1996, 1998; Kanwisher *et al.*, 1997; McCarthy *et al.*, 1997; Epstein & Kanwisher, 1998). Furthermore, bilateral lingual gyri were also activated for place pictures. The lingual gyrus performs bottom-up perceptual analysis of a scene in order to recognize it. Lesions in this area are known to result in topographical disorientation, which refers to the inability to orient oneself in a scene or surrounding (Aguirre *et al.*, 1996; Sulpizio *et al.*, 2013).

Two other regions selected by the classifier were 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 the anterior medial temporal gyrus (amTG) that is common to faces and buildings, though this was stronger in the right than in the 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 famous landmarks used in the stimulus set were less familiar to subjects compared with famous people.

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). The 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.

We also found that above-chance accuracies were obtained for some scans in the transition period, i.e. the first 6 s of the BOLD activity after stimulus onset. This supports the finding of Laconte *et al.* (2007), where an offline analysis showed that the transition period of the hemodynamic response contains reliable information that can be decoded with above-chance accuracy. We have therefore shown that predictions for scans in the transition period, if required, can be used in real-time fMRI to reduce neurofeedback delay by as much as 6 s.

Additionally, we tested how a whole-brain classifier compared with a GLM-restricted classifier. In whole-brain decoding, the input features to the classifier included all voxels in the entire volume. This classifier could therefore include any voxels in the model that it considered useful for separating the two classes. On the other hand, in the GLM-restricted approach, the input features to the classifier were univariately reduced to only those voxels that responded to the experimental manipulation. We found that both these classifiers yielded the same decoding performance. The whole-brain multivariate approach is potentially a more sensitive approach as it can not only detect voxels that respond to the experimental manipulation but can also take interactions between the voxels into account that are ignored by a

massively univariate approach such as a GLM. Moreover, using a whole-brain elastic net logistic regression classifier in real-time fMRI decoding experiments results in a simpler and computationally more efficient experimental design. Averaging of time-series within each cluster resulted in a performance loss, though this was not significant given the low number of subjects used. At the same time, one should be aware of the fact that multivariate approaches may also be sensitive to confounds that systematically covary with the conditions of interest. The fact that the GLM identified regions that overlapped with those found by the multivariate approach provides support that the multivariate approach is also driven by neural correlates of shifts in object-based attention.

Furthermore, we analysed if the decoding was driven by highly localized activity patterns or by distributed cortical activations by training and testing decoders on individual clusters detected in the GLM. Because decoding on these small individual clusters yielded poor decoding performance compared with the whole-brain or GLM-restricted decoders, it suggests that faces and places are encoded in the brain using distributed patterns cortical activations, and as such detection of these patterns requires a multivariate decoder with input features spread across the brain.

Finally, because the MVA-W classifier - trained only on pictures of separately presented faces and places - could not recruit any regions related to attention, we conducted a reverse MVPA to find regions associated with attention. We trained two classifiers: one on the feedback condition; and the other on the non-feedback condition. Subsequently, these classifiers were tested on the localizer. We not only found activations in the same brain regions responsible for processing faces and places as we found in MVA-W, but also detected additional brain regions associated with attention and cognitive control. We found activation in superior frontal, middle frontal and superior medial frontal gyri. These are part of the frontal-parietal network that have been known to become active in top-down attentional control paradigms (Li *et al.*, 2010) and during bistable perception in which the observer's perception can fluctuate between competing stimuli (Knäpen *et al.*, 2011). We also found activation in crus I of the left cerebellum. The cerebellum not only plays an important role in motor coordination, but has also been shown to be involved in higher cognitive functions such as selective visual attention (Allen, 1997). Moreover, activations in middle and anterior cingulate were also detected. Previous studies have shown that these regions play a crucial role in attention-demanding tasks by competition monitoring and goal-directed selective attention (Danckert *et al.*, 2000; Davis *et al.*, 2000). Activation in bilateral precuneus was also found, but only in the classifier trained on the non-feedback condition. Activation in this region has been shown in a previous study (Hahn *et al.*, 2006) during engagement of top-down spatial selective attention. This may imply that subjects were engaged in both object-based and space-based visual attention during the non-feedback condition. Apart from activation of these additional brain regions in the reverse MVPA, we also observed that the classifier trained on the feedback condition performed significantly higher than the classifier trained on the non-feedback condition. This indicates that a classifier trained only on pictures of separately presented faces and places may not be the most optimal way of decoding object-based visual attention.

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

Supporting Information

Additional supporting information can be found in the online version of this article:

Fig. S1. A basis set of 15 face-place pairs used in decoding phase. Each pair was used twice in each condition, once with the face picture set as target and the other time with the place picture set as target. Note: Copyrighted pictures used in the original experiment have been replaced in the above graphic by their non-copyrighted look-alikes.

Fig. S2. Graph-based visual saliency algorithm was used to select the face-place pairs. Saliency of the 50/50 hybrid and each of its constituents were observed and only those pairs were selected for which the 50/50 hybrid had an equal number of salient points for the face and place picture.

Fig. S3. Stimulus timeline. (A) Example of an attend-face trial in non-feedback condition. (B) Example of an attend-place trial in feedback condition. After cues have been presented, the face-place hybrid image was updated every TR depending on classification result of the preceding TR.

Fig. S4. List of all brain regions from which voxels were selected by the MVA-W classifier for training. Only regions that were activated across three or more subjects were used for further analyses.

Fig. S5. (A) Absolute number of voxels selected in the regions used by classifier for training averaged across the group. (B) Percentage of voxels used per region averaged across the group. Error bars show standard error of the mean.

Fig. S6. (A) Decoding accuracy as a function of TR for feedback and non-feedback condition, and attend-face and attend-place trials that constitute these two conditions. The filled round markers represent significantly above-chance decoding ($P < 0.05$) whereas the empty markers represent below-chance decoding ($P > 0.05$). (B) Mean decoding accuracy. Error bars indicate standard error of the mean.

Fig. S7. Comparison of percent signal change in feedback and non-feedback conditions. (A) Percent signal change for attend-face trials in feedback and non-feedback condition. The top plots show percent signal change at every TR during a trial (including the 12 s rest period). The bottom plot shows the percent signal change aggregated over the 12 TRs. (B) Percent signal change for attend-place trials in feedback and non-feedback conditions. Error bars represent standard error of the mean.

Fig. S8. Comparison of prediction probabilities of the decoder for feedback and non-feedback conditions. (A) Prediction probability for feedback and non-feedback conditions containing both successful and failed trials. No significant difference was found. (B) Prediction probability for only successful trials in feedback and non-feedback conditions. The prediction probability for feedback trials was significantly higher than non-feedback trials (C) Prediction probability for only failed trials in feedback and non-feedback conditions. The prediction probability for failed trials was significantly stronger (lower) for feedback trials compared to non-feedback trials. Error bars represent standard error of the mean.

Fig. S9. (A) Average decoding performance for classifiers trained on feedback and non-feedback conditions. The classifier trained on the feedback condition was decoded with significantly higher accuracy than the classifier trained on the non-feedback condition. (B). Anatomical regions recruited by the classifiers trained on feedback and non-feedback conditions

Movie S1. The movie demonstrates an example of a trial in feedback and non-feedback conditions. Furthermore it shows the actual performance of one particular subject for all attend-face and attend-place trials in the feedback condition.

Acknowledgements

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 the experimental setup and development of the real-time fMRI pipeline. We are very grateful to the editors and the anonymous reviewers for their encouraging and constructive comments on our manuscript.

Abbreviations

aMTG, anterior medial temporal gyrus; BCI, brain-computer interface; BOLD, blood oxygen level-dependent; FFA, fusiform face area; fMRI, functional magnetic resonance imaging; GLM, general linear model; MoCo, motion-corrected; MVA-C, cluster-wise multivariate analysis; MVA-G, GLM-restricted multivariate analysis; MVA-T, mean timeseries multivariate analysis; MVA-W, whole-brain multivariate analysis; MVPA, multivoxel pattern analysis; OFA, occipital face area; PACE, prospective acquisition correction; TR, repetition time.

References

- Aguirre, G.K., Detre, J.A., Alsup, D.C. & D'Esposito, M. (1996) The parahippocampus subserves topographical learning in man. *Cereb. 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.
- Allen, G. (1997) Attentional activation of the cerebellum independent of motor involvement. *Science*, **275**, 1940–1943.
- Anderson, A., Han, D., Douglas, P.K., Bramen, J. & 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), *Machine Learning and Interpretation in Neuroimaging*. Springer Berlin Heidelberg, Spain, pp. 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. *Nat. Neurosci.*, **5**, 1210–1216.
- Brincat, S.L. & Connor, C.E. (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.*, **7**, 880–886.
- 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.
- Danckert, J., Maruff, P., Ymer, C., Kinsella, G., Yucel, M., de Graaff, S. & Currie, J. (2000) Goal-directed selective attention and response competition monitoring: evidence from unilateral parietal and anterior cingulate lesions. *Neuropsychology*, **14**, 16–28.
- Davis, K.D., Hutchison, W.D., Lozano, A.M., Tasker, R.R. & Dostrovsky, J.O. (2000) Human anterior cingulate cortex neurons modulated by attention-demanding tasks. *J. Neurophysiol.*, **83**, 3575–3577.
- DeBettencourt, M.T., Lee, R.F., Cohen, J.D., Norman, K.A. & Turk-Browne, N.B. (2012) Real-time decoding and training of attention. *J. Vision*, **12**, 377.
- Dehaene, S. & Cohen, L. (2011) The unique role of the visual word form area in reading. *Trends Cogn. Sci.*, **15**, 254–262.
- Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, **18**, 193–222.
- Eimer, M. (1996) The N2pc component as an indicator of attentional selectivity. *Electroen. Clin. Neuro.*, **99**, 225–234.
- Epstein, R. & Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature*, **392**, 598–601.
- Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, **23**, 115–125.
- 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. *J. Stat. Softw.*, **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. *J. Cognitive Neurosci.*, **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. *J. Neurosci.*, **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.
- Hahn, B., Ross, T.J. & Stein, E.A. (2006) Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. *NeuroImage*, **32**, 842–853.
- 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. *Front. Psychol.*, **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. *Nat. Rev. Neurosci.*, **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. *Cereb. 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. *Front. Psychol.*, **2**, 100.
- Kamitani, Y. & Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.*, **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. *J. Neurosci.*, **17**, 4302–4311.
- Knäpen, T., Brascamp, J., Pearson, J., van Ee, R. & Blake, R. (2011) The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.*, **31**, 10293–10301.
- Laconte, S.M., Peltier, S.J. & Hu, X.P. (2007) Real-Time fMRI using brain-state classification. *Hum. Brain Mapp.*, **1044**, 1033–1044.
- 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. *Neurorehab. Neural Re.*, **25**, 259–267.
- Li, W., Piëch, V. & Gilbert, C.D. (2004) Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.*, **7**, 651–657.
- Li, L., Gratton, C., Yao, D. & Knight, R.T. (2010) Role of frontal and parietal cortices in the control of bottom-up and top-down attention in humans. *Brain Res.*, **1344**, 173–184.
- Liu, J., Harris, A. & Kanwisher, N. (2002) Stages of processing in face perception: an MEG study. *Nat. Neurosci.*, **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. *J. Cognitive Neurosci.*, **5**, 188–195.
- Macevoy, S.P. & Epstein, R.A. (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.*, **19**, 943–947.
- McCandliss, B.D., Cohen, L. & Dehaene, S. (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.*, **7**, 293–299.
- McCarthy, G., Puce, A., Gore, J.C. & Allison, T. (1997) Face-specific processing in the human fusiform gyrus. *J. Cognitive Neurosci.*, **9**, 605–610.
- 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. *J. Neurosci.*, **32**, 8649–8662.
- Nichols, T. & Hayasaka, S. (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.*, **12**, 419–446.
- O'Craven, K.M., Downing, P.E. & Kanwisher, N. (1999) fMRI evidence for objects as the units of attentional selection. *Nature*, **401**, 584–587.
- Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.*, **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., Walsh, V., Yovel, G. & Duchaine, B. (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Curr. Biol.*, **17**, 1568–1573.
- Pitcher, D., Garrido, L., Walsh, V. & Duchaine, B. (2008) TMS disrupts the perception and embodiment of facial expressions. *J. Neurosci.*, **8**, 700.
- Puce, A., Allison, T., Gore, J.C. & McCarthy, G. (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.*, **74**, 1192–1199.
- Rahnev, D., Maniscalco, B., Graves, T., de Huang, E., Lange, F.P. & Lau, H. (2011) Attention induces conservative subjective biases in visual perception. *Nat. Neurosci.*, **14**, 1513–1515.
- Reddy, L. & Kanwisher, N. (2006) Coding of visual objects in the ventral stream. *Curr. Opin. Neurobiol.*, **16**, 408–414.
- Reddy, L. & Tsuchiya, N. (2010) Reading the minds 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. *Annu. Rev. Psychol.*, **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. *Cereb. 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. *Curr. Biol.*, **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.
- Sulpizio, V., Comitteri, G., Lambrey, S., Berthoz, A. & Galati, G. (2013) Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behav. Brain Res.*, **242**, 62–75.
- Thesen, S., Heid, O., Mueller, E. & Schad, L.R. (2000) Prospective acquisition correction for head motion with image-based tracking for Real-Time fMRI. *Magn. Reson. Med.*, **465**, 457–465.
- Tong, F. & Pratte, M.S. (2012) Decoding patterns of human brain activity. *Annu. Rev. Psychol.*, **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 Cogn. Sci.*, **16**, 251–254.
- Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N. & Mathiak, K. (2005) Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *J. Physiol. 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. *IEEE International Conference on Acoustics Speech and Signal Processing ICASSP*, 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 Res.*, **1080**, 53–62.