NeuroImage 83 (2013) 1063–1073

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

MEG-based decoding of the spatiotemporal dynamics of visual category perception

M.E. van de Nieuwenhuijzen *, A.R. Backus, A. Bahramisharif, C.F. Doeller, O. Jensen, M.A.J. van Gerven

Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Accepted 25 July 2013
Available online 6 August 2013

Keywords:
Spatiotemporal dynamics
Multivariate analysis
Perception
MEG

ABSTRACT

Visual processing is a complex task which is best investigated using sensitive multivariate analysis methods that can capture representation-specific brain activity over both time and space. In this study, we applied a multivariate decoding algorithm to MEG data of subjects engaged in passive viewing of images of faces, scenes, bodies and tools. We used reconstructed source-space time courses as input to the algorithm in order to localize brain regions involved in optimal image discrimination. Applying this method to the interval of 115 to 315 ms after stimulus onset, we show a focal localization of regression coefficients in the inferior occipital, middle occipital, and lingual gyrus that drive decoding of the different perceived image categories. Classifier accuracy was highest (over 90% correctly classified trials, compared to a chance level accuracy of 50%) when dissociating the perception of faces from perception of other object categories. Furthermore, we applied this method to each single time point to extract the temporal evolution of visual perception. This allowed for the detection of differences in visual category perception as early as 85 ms after stimulus onset. Furthermore, localizing the corresponding regression coefficients of each time point allowed us to capture the spatiotemporal dynamics of visual category perception. This revealed initial involvement of sources in the inferior occipital, inferior temporal and superior occipital gyrus. During sustained stimulation additional sources in the anterior inferior temporal gyrus and superior parietal gyrus became involved. We conclude that decoding of source-space MEG data provides a suitable method to investigate the spatiotemporal dynamics of ongoing cognitive processing.

1. Introduction

Visual perception is an intricate yet fast process. This is not only the case for isolated low-level image properties such as object location and luminance. We are also able to discern different image categories such as faces and tools very rapidly. Indeed, saccade studies show that as little as 120 ms is required before a saccade differentiating between animals and scenes is initiated (Kirchner and Thorpe, 2006), and only 100–110 ms before saccading to human faces (Crouzet et al., 2010). As these latencies include the process of saccade planning, differential brain activity for faces as compared to other semantic categories should already be present as early as 80 ms after stimulus onset (Crouzet et al., 2010). However, neuroimaging research has been inconclusive about this exact timing, demonstrating variable latencies ranging from 40 to 75 ms (Bacon-Macé et al., 2005; Kirchner et al., 2009; Ramkumar et al., 2013; Seeck et al., 1997; VanRullen and Thorpe, 2001) up to 100–150 ms (Amano et al., 2006; Bode et al., 2012; Carlson et al., 2011; Fuentemilla et al., 2010; Jafarpour et al., 2013; Liu et al., 2009; Simanova et al., 2010; Thorpe et al., 1996; Wokke et al., 2012). In addition to the fast onset of vision-related activity, the temporal evolution of this activity is also complex, consisting of multiple visual event-related potential components, including the P1, N1, and the face-related N170 (Bentin et al., 1996).

Next to this rapid and complex temporal evolution, a multitude of brain regions is employed in the process of visual perception. Areas along both the ventral and dorsal streams of visual processing are recruited, ranging from the striate cortex in the occipital lobe to the anterior temporal and parietal lobes (Mishkin et al., 1983). Furthermore, this activity is thought to proceed in rapid succession (Kirchner and Thorpe, 2006; Riesenhuber and Poggio, 1999; Serre et al., 2007; Thorpe and Fabre-Thorpe, 2001). Therefore, a method is required that can capture both the temporal and spatial aspects of visual processing.

Although electrophysiological methods do provide exquisite temporal resolution, they generally lack the spatial resolution to localize the aforementioned temporal evolution. We therefore aim to improve spatial localization by applying a multivariate classification algorithm to reconstructed source-space activity, calculated from magnetoencephalography (MEG) recordings. Multivariate methods can be more sensitive than univariate methods when discriminative information is distributed across multiple sources (Kriegeskorte, 2011; Kriegeskorte et al., 2006; Lange et al., 1999). Previous work with functional magnetic

* Corresponding author at: Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.
E-mail addresses: m.vandenieuwenhuijzen@donders.ru.nl (M.E. van de Nieuwenhuijzen), a.backus@donders.ru.nl (A.R. Backus), a.bahramisharif@donders.ru.nl (A. Bahramisharif), c.doeller@donders.ru.nl (C.F. Doeller), ole.jensen@donders.ru.nl (O. Jensen), m.vangerven@donders.ru.nl (M.A.J. van Gerven).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.07.075

resonance imaging (fMRI) data has shown that multivariate methods allow for the decoding of perceived stimuli from brain activity, both of low-level image properties (Haynes and Rees, 2005; Kamitani and Tong, 2005; Miyawaki et al., 2008; Thirion et al., 2006; van Gerven et al., 2010a) as well as high-level image categories (Carlson et al., 2003; Connolly et al., 2012; Cox and Savoy, 2003; Haxby et al., 2001; Kay et al., 2008; Naselaris et al., 2009; van Gerven et al., 2010b). Although multivariate pattern analysis has become a widely used method for analysis of fMRI data, application to MEG and electroencephalography (EEG) data is less common. Nonetheless, multivariate techniques have been used to discern both image features (Carlson et al., 2011; Ramkumar et al., 2013) and semantic categories of images and words from electrophysiological data (Bode et al., 2012; Carlson et al., 2011; Chan et al., 2011; Fuentemilla et al., 2010; Guimaraes et al., 2007; Jafarpour et al., 2013; Mitchell et al., 2008; Murphy et al., 2009, 2011; Simanova et al., 2010; Suppes et al., 1997, 1999; van Gerven et al., 2013).

Most of these previous electrophysiological studies have used sensor-space data as input to the decoding algorithm. However, as these signals are a mixture of the time courses of the actual underlying sources, reconstructing and analyzing these underlying source-space activity time-courses should improve localization of involved neuronal generators. Some studies have already shown that source-space activity can indeed be used successfully as input to a classification algorithm (Sandberg et al., 2013; Sudre et al., 2012; Wu and Gao, 2011). However, these studies often restrict themselves to predetermined regions of interest, or have used source-space activity to improve classification performance without drawing additional conclusions about localization. In contrast, we propose to apply the classification algorithm to all source-space activity time-courses to determine what sources drive the classification of representation-specific information, independent of any a priori assumptions about possible sources. In addition, we facilitate interpretability by applying an elastic net algorithm (Friedman et al., 2010). This algorithm enforces a sparsity constraint, resulting in focal and hence interpretable sources (Carroll et al., 2009). By using this method we aim to achieve a better identification of the neuronal generators involved in early visual natural image perception. Furthermore, by performing classification on these source-space data for each separate time point, we attempt to extract which sources are involved in category perception at any specific point in time. This allows for high-resolution source localization on a time-scale in the order of milliseconds.

Next to source activity reconstruction with the more traditional LCMV beamformer (Van Veen et al., 1997), we also used the dynamic beamformer to get an additional estimation of source-space activity (Bahramisharif et al., 2012). This novel method may further improve the signal-to-noise ratio of the reconstructed time courses by imposing temporal smoothness on the source reconstructions. In this way, we aim to improve localization performance even further by providing less noisy input to the employed classification algorithm.

Low-level image properties are important features based on which the visual system discriminates between different image categories. For example, the spatial frequency of an image has been shown to be an important feature based on which the visual system assigns category membership (Crouzet and Thorpe, 2011). Therefore, we attempted to minimize the effect of these low-level image properties by showing images that were corrected for both luminance and spatial frequency. As the correction for spatial frequency reduced the visibility of the images, we also showed the same images corrected solely for luminance.

We used a passive viewing paradigm to avoid interference of any task-related effects, such as encoding or attention. After all, specific attention towards the stimulus could boost the neuronal activity (Reynolds et al., 2000) or even speed up the neuronal response (Noguchi et al., 2007).

As will be shown, source-space decoding of MEG data provides a sensitive way to investigate the spatiotemporal dynamics of visual perception as brain activity proceeds along the visual hierarchy.

2. Methods

2.1. Subjects and stimuli

Three right-handed subjects (two males; two subjects were aged 27, one 25) were shown black-and-white photographs of faces, scenes, bodies and handheld objects. The choice for these categories was based on the spatial selectivity shown for these types of stimuli in the fusiform face area (faces), parahippocampal place area (PPA; scenes), extrastriate body area (bodies) and middle temporal gyrus (tools) (Downing et al., 2006). Images in the face category consisted of male and female faces with a neutral expression facing forward. Scene images were pictures from houses, forests and mountains. Images in the body category contained images of male and female whole bodies facing forward with the face grayed out. Tools and kitchen utensils made up the object category. Each category contained 30 different images, which were all shown twice per image correction condition (see below). Images were selected from different online sets. Face images were obtained from the Karolinska Directed Emotional Faces dataset (KDEF, images F1, F2, F6, F8, F9, F10, F11, F13, F17, F19, F20, F22, F24, F25, F27, M37, M39, M41, M42, M43, M44, M45, M47, M52, M56, M58, M63, M64, M65, M66; Goeleven et al., 2008; Lundqvist et al., 1998). The scene database was obtained from the Stanford Vision Lab (Fei-Fei and Perona, 2005). Images of bodies were selected from the bodily expressive action stimulus test set (BEAST; De Gelder and Van den Stock, 2011). Object images were derived from the Bank of standardized images (BOSS; Brodeur et al., 2010). All images were cropped to 300×300 pixels. A larger set of luminance and spatial frequency corrected images were rated prior to the study on their visibility and category membership by four naive viewers on a five-point scale. These four subjects did not take part in the actual experiment. The 30 images with the highest score were selected for the actual experiment. The included images all scored between 4 and 5 on visibility, indicating that they were very well recognizable, despite the blurriness of the images after correction for spatial frequency.

All stimuli were corrected for luminance only, as well as for both luminance and spatial frequency. Correction of the images was done with the SHINE toolbox for MATLAB (Willenbockel et al., 2010). Luminance correction was performed by scaling the mean luminance and standard deviation of the entire image. Luminance and spatial frequency correction was performed by matching the Fourier amplitude spectra of the images without optimization of the structural similarity index, and then equating the luminance histograms over the entire image. In addition, this correction step resulted in equal root mean square values of the contrast for all images. We will focus on the data of the luminance and spatial frequency corrected images, as this is the more stringent case. Results of the luminance only corrected data are comparable to those of the luminance and spatial frequency corrected data as shown in the Supplementary data.

Images were presented to the subjects according to the design depicted in Fig. 1.

During each trial, a natural image spanning the central 6° of the visual field was presented at the center of the screen for 2 s. Subjects had to focus on a central fixation dot with a diameter of 0.5° to maintain fixation and to prevent eye movements from being made when viewing the image. In 10% of the trials the fixation dot turned red in order to keep the subjects' attention fixed to the center of the screen. When this happened, the subject had to press a button with their right hand within 1.5 s, after which feedback was given and a new trial would start. These catch-trials were excluded from analysis. In this way passive viewing was enforced, as there was no need to attend the images and their categories or to encode them to perform the task successfully.

Stimuli were separated by an interstimulus interval of 2 s. During the first second, indicated by the absence of the fixation dot, the subjects were allowed to blink. Next, the fixation dot returned and subjects were again required to keep their gaze fixated steadily at the center of the screen without blinking.

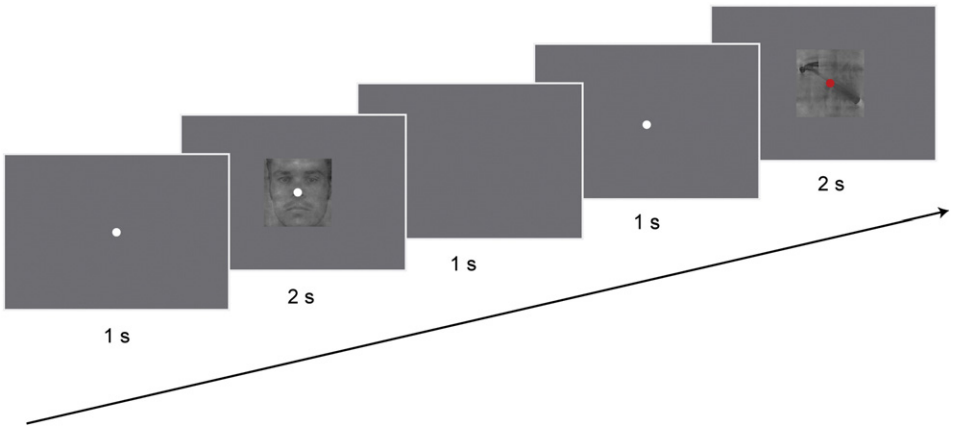

Fig. 1. Study design. Images were presented for 2 s while subjects fixated on the fixation dot, followed by an interstimulus interval of 2 s. Subjects had to press a button when the fixation dot turned red. During the first second of the interstimulus interval, indicated by the absence of the fixation dot, the subject was allowed to blink. Blinking was not allowed during the next second when the fixation dot was present again. The face image in this figure comes from the KDEF dataset (M64; Lundqvist et al., 1998), the tool image comes from the BOSS set (Brodeur et al., 2010).

The task consisted of 12 blocks of 40 trials separated by a subject-paced break. Every even block consisted of images that were only corrected for luminance, and every odd block consisted of images that were corrected for both luminance and spatial frequency. The study was approved by the local ethics review board (Commissie Mensgebonden Onderzoek Regio Arnhem–Nijmegen).

Stimuli were presented using Presentation software (Version 16.2, Neurobehavioral Systems, Inc.) via an LCD projector located outside the magnetically shielded room. Stimuli were back-projected onto a translucent screen via two front-silvered mirrors. The projector lag was measured using a photosensor placed on the screen while a black–white flickering stimulus was presented. Both the signal of the photosensor and the trigger of the stimulus were sent to an oscilloscope. The projector lag was defined as the lag between the trigger of the stimulus and the signal of the photosensor, and was found to be 35 ms. The preprocessing analyses assumed a projector lag of 50 ms. This was post-hoc corrected for by shifting the MEG data 15 ms backwards with respect to the stimulus triggers, resulting in slightly shifted reported time intervals. All time indications in this manuscript have been fully corrected for the projector lag.

2.2. MEG acquisition

MEG data were recorded using a 275-sensor whole-head system (CTF Systems Inc., Port Coquitlam, Canada) at a sampling frequency of 1200 Hz. Due to sensor malfunction, data from two sensors (MLT37 and MLF62) were not recorded. Subjects were seated in a dark magnetically shielded room. Three coils, one in both ears and one on the nasion, were used to determine head position relative to the sensors. Head motion was monitored during the measurement using a real-time head localizer (Stolk et al., 2013). When head motion exceeded 5 mm, subjects were asked to reposition their head to the original location, visually guided by the real-time head localizer. In this way, head motion has been kept below 5 mm over the entire run for two of the subjects, and below 7 mm for the third subject. No post-hoc correction of head movement was performed.

A continuous bipolar electrooculogram (EOG) was recorded for offline rejection of artifacts related to eye movement. This was done with four electrodes around the eyes — one below and above the left eye for vertical EOG, as well as one left of the left eye and right of the right eye for horizontal EOG. An electrocardiogram (ECG) was recorded with an electrode on the left collarbone and below the right rib. The ground electrode was located below the elbow of the left arm. Eye movements were also measured using an Eye Link SR Research Eye tracker. These data were used for additional rejection of eye movements that were not detected by the less sensitive EOG, and to ensure that the observed effects are not likely to be explained by ocular motion artifacts (see Multivariate analysis section).

2.3. Preprocessing

Data were analyzed using MATLAB version 7.9.0, R2009b (The Mathworks Inc., Natic, MA) and FieldTrip, an open source Matlab toolbox for the analysis of neuroimaging data (Oostenveld et al., 2011). Trials were defined as data ranging between 85 ms before stimulus onset and 2015 ms after stimulus onset. This interval was chosen as it spans the entire period during which the stimulus was presented (2 s), as well as a short baseline period before stimulus onset. MEG time courses of these trials were visually inspected, and trials that contained artifacts resulting from SQUID jumps and muscle contractions were rejected. Furthermore, EOG and eye tracker traces were inspected visually. Trials with evident eye movement in these signals were excluded from further analysis as well. After artifact rejection on average 49 ($sd = 2.58$) trials per category per condition remained for analysis. Data were low-pass filtered at 150 Hz, and 50 Hz line noise was removed from the data with a DFT notch filter. The period between 185 and 85 ms before stimulus onset was averaged per trial per channel and subtracted from the corresponding signal as baseline correction. The data were downsampled to 300 Hz to reduce memory and CPU load. Finally, environmental noise components measured by the third order synthetic gradiometers were subtracted from the sensor data.

2.4. Beamforming

Source-space activity time courses of the whole trial were reconstructed with an LCMV beamformer (Van Veen et al., 1997). This method creates a spatial filter, which optimizes the signal coming from a given source while suppressing activity coming from other sources. No a priori selection of expected activated sources is required, making it a well-suited method for whole brain source activation reconstruction. An additional reconstruction was performed with the dynamic beamformer, which enforces a temporal smoothness on the reconstructed time courses by taking previous samples into account (Bahramisharif et al., 2012). This improves the signal-to-noise ratio of the activity time-courses which were used as input to the classification algorithm. The single shell model as described by Nolte (2003) was used as head model. Individual grids with a resolution of 10 mm were calculated based on T1 weighted MRI data acquired using a 1.5 T whole body scanner (Siemens Magnetom Avanto, Siemens, Erlangen, Germany). Vitamin E markers in the ears marked the same location as

the fiducial ear coils to allow for alignment of the MEG and MRI data. For both the LCMV and the dynamic beamformer the rank of the leadfield was reduced to two per voxel and the leadfield was normalized. Normalization was performed for each voxel by dividing the leadfield by the infinity norm.

Single trial covariance matrices were estimated from the sensor-space data in the time interval of 115 to 315 ms after stimulus onset for classification on that interval as a whole, as well as from 85 ms before stimulus onset to 2015 ms after stimulus onset for classification on each time point. Apart from the broadband filtering during preprocessing, the data were not additionally filtered before beamforming. The covariance matrices were regularized by adding 5% of the average eigenvalues of that matrix to its eigenvalues, to correct for the coarse estimation of the covariance matrix. The same filter was used for each trial in order to reconstruct time courses at virtual channels in the maximum power orientation. Activation of each virtual sensor is estimated as the first eigenvalue of the three dimensions. These source-space activation time-courses were used as input data to the classification algorithm.

2.5. Multivariate analysis

Multivariate analysis was performed with an elastic net logistic regression algorithm (Friedman et al., 2010). Given training data, the algorithm maximizes the log-likelihood, penalized by the elastic net penalty

$$P_\alpha(b) = \sum_{j=1}^{p}\left(\frac{1}{2}(1-\alpha)b_j^2 + \alpha\left|b_j\right|\right)$$

where b is the vector of regression coefficients. This penalty term combines L1 and L2 regularization through a mixing parameter α such that $\alpha = 0$ leads to L2-regularized logistic regression and $\alpha = 1$ leads to L1-regularized logistic regression. In all experiments, the mixing parameter was set to $\alpha = 0.01$, encouraging both sparseness and smoothness of the resulting vectors of regression coefficients. The influence of the elastic net penalty is controlled by a regularization parameter λ, which was optimized using a nested cross-validation procedure. The input data were standardized prior to this analysis. Mean and standard deviation were derived from the training set, and the data in this set were z-transformed using these values. In addition, data in the test set were also z-transformed using the mean and standard deviation derived from the training set.

Classifier performance was quantified in terms of accuracy (proportion of correctly classified trials). In addition, the absolute value of the estimated regression coefficient associated with each feature was indicative of the importance of that feature for classifier performance. These measures were computed by training a classifier on whole-brain data at the interval of 115 to 315 ms after stimulus onset. This time period was chosen because the peak of the visual event-related response roughly lies within this interval (e.g. Ales et al., 2012; Fellinger et al., 2012; Itier and Taylor, 2004; Liu et al., 2009; Petrov et al., 2012; Seeck et al., 1997; VanRullen and Thorpe, 2001). Classifier performance was validated using five-fold cross-validation. By using cross-validation, the classifier was always tested on data it was not trained on, to prevent double dipping. In sensor space, 163,800 features were used (600 time points × 273 MEG sensors), whereas in source space this number was inflated to 1,626,600 features (600 time points × 2711 grid points). No a priori feature selection was performed. However, the elastic net algorithm itself, by imposing a sparsity constraint, performed feature selection by setting the weights of a large set of features which were not necessary for classification to zero.

Computations were run on a distributed computing cluster with cores whose clock rate ranged between 2.0 and 3.6 GHz. Sensor level analyses as described above took about 5 min and required about 4 GB of RAM per contrast per subject. Source space analyses required about 26 GB of RAM and took about 21 min to complete.

Next to analyses over all grid points, we also ran the classification algorithm over 115 to 315 ms after stimulus onset on a selection of source space grid points, in order to assess the relative influences of different regions of interest. We selected grid points that belonged to non-overlapping cubes over respectively the bilateral occipital lobes, right temporal lobe and bilateral frontal lobes. The grid point cube was visually matched to the specific lobe in the underlying brain volume.

In addition to this overall measure, we applied this same classification algorithm to each source-space activity time-point separately in order to assess the temporal evolution of detectable information in the brain. This was validated using ten-fold cross-validation, and resulted in an accuracy trace with a temporal resolution of 3.3 ms.

Finally, we applied the aforementioned classification protocol with ten-fold cross-validation to eye tracker and EOG data, as well as to the input images, in order to test whether eye movement or low-level image properties could be driving classification performance. If classification on EOG or the eye tracker signal would be possible, classifier performance found in the MEG signal could be due to artificial fluctuations induced by eye motion. If this would not be the case, it is less likely that the signal used for classification is driven by fluctuations related to ocular motion. Similarly, if classification on the images shown to the subjects would be possible, low-level image properties could be driving the classifier performance on the MEG signals, whereas if classification on the images would not exceed chance level these low-level properties would be less likely to drive classification, suggesting the involvement of higher-order semantic properties instead.

2.6. Statistical testing

Individual accuracies were tested on their deviation from chance level with a binomial test. Multiple comparisons were corrected for using a Bonferroni correction at an alpha level of 0.05 for 54 multiple comparisons (3 subjects × 6 contrasts × 3 input data types). Differences between accuracy levels of classification on sensor-space data and accuracies based on the different source-space reconstructions were tested with a binomial test at single-subject level, as proposed by Salzberg (1997), using a Bonferroni correction at an alpha level of 0.05 for 36 multiple comparisons (3 subjects × 6 contrasts × 2 comparisons between source and sensor space data). For the accuracy values per time point multiple comparisons were corrected for by calculating the false discovery rate (FDR) and corresponding threshold value at an alpha of 0.05 for all three subjects at each contrast and image correction condition. The most conservative FDR threshold value for each contrast and condition was then used as the threshold for the average time course over subjects.

3. Results

3.1. Overview of classification performance

When applying the classification algorithm to whole brain data between 115 and 315 ms after stimulus onset, accuracies often rose well above chance level (0.5; Table 1). Classifier accuracies based on the EOG and eye tracker traces were not significantly better than chance, except when contrasting bodies and scenes with tools for one subject (Bonferroni corrected $p < 0.05$, accuracy 0.66–0.67), implying that the results are in general based on genuine brain activity instead of eye motion artifacts. Classification accuracy based on the input images was in all cases well above chance level, with accuracies between 0.93 and 1 ($p < 0.001$). Hence, it could well be that low-level properties of the perceived images at least partially drive classification.

The highest accuracies were obtained when a distinction was made between the perception of faces and any other category. Furthermore, when performing classification on source-space time courses, differentiation between images showing bodies and scenes, as well as between bodies and tools became possible in all subjects. The only contrast

Table 1
Mean accuracy when performing classification based on sensor-space data, on source-space activity reconstructed with the LCMV beamformer, and on source-space activity reconstructed with the dynamic beamformer. Grand mean accuracies of the different data representations (sensor space and source space) are shown in the last row. Grand mean accuracies of the different contrasts are shown in the last column. The perceived images were corrected for both luminance and spatial frequency. Standard deviations are given between brackets. Asterisks indicate that the classification accuracy was significantly higher than chance level for all subjects (Bonferroni corrected $p < 0.05$). Values in bold font indicate a significant increase in accuracy for at least one subject compared to sensor-space accuracy (Bonferroni corrected $p < 0.05$). Values in italic font indicate a trend towards increase in accuracy for at least one subject compared to sensor-space accuracy (uncorrected $p < 0.05$).

Contrast	Sensor level (*sd*)	LCMV beamformer (*sd*)	Dynamic beamformer (*sd*)	Contrast average
Face–tool	0.80 (0.06)*	*0.88 (0.03)**	**0.90 (0.06)***	0.86
Face–scene	0.86 (0.05)*	0.87 (0.05)*	*0.94 (0.06)**	0.89
Face–body	0.83 (0.11)*	*0.83 (0.10)**	**0.93 (0.05)***	0.86
Scene–body	0.69 (0.09)	*0.77 (0.07)**	**0.86 (0.12)***	0.77
Body–tool	0.69 (0.08)	0.74 (0.06)*	*0.83 (0.07)**	0.75
Scene–tool	0.58 (0.09)	*0.67 (0.08)*	*0.75 (0.10)*	0.67
Average	0.74	0.79	0.87	

without a stable accuracy above chance level for individual subjects and also with the lowest average accuracy in both sensor and source space was the contrast between scenes and tools. Here accuracies only rose above chance level for only one subject in sensor space, and for two subjects in source space.

In general, the results seem to indicate that accuracy increased when applying the classifier to activity time-courses reconstructed with the LCMV beamformer compared to sensor-space time courses. This trend seems even stronger when the classifier was applied to time courses reconstructed with the dynamic beamformer. A single-subject binomial test revealed that, for one subject, accuracy increased significantly when discerning faces and scenes from bodies, as well as faces from tools, based on activity time-courses reconstructed with the dynamic beamformer as compared to sensor space ($p < 0.0005$). In addition, there was a trend towards increased accuracies for at least one subject in all contrasts when applying the classifier to time courses reconstructed by the dynamic beamformer compared to sensor-space time courses (uncorrected $p < 0.05$). This trend was also observed when applying the algorithm to time courses reconstructed with the LCMV beamformer compared to sensor-space time courses (uncorrected $p < 0.05$), albeit not for the contrast between faces and scenes, and between bodies and tools. For images that were corrected for luminance only, results were comparable, although the individual differences were more pronounced and variable (see Supplementary Table 1).

For the sake of conciseness the main text of this paper will focus on results obtained for the luminance and spatial frequency corrected faces versus tools contrast, because of the high and stable accuracy in all conditions and subjects. Also, other contrasts with similar classification scores may have their drawbacks: scene-related activity could be expected to originate from the PPA (Epstein and Kanwisher, 1998), which may be located too medially for MEG to pick up. Hence, the performance of faces versus scenes may be explained as a face signal versus noise. In addition, the performance of the classifier in discriminating faces from bodies was more variable over subjects.

3.2. Localization based on sensor level data

The regression coefficients estimated for classification on sensor-space data over the interval of 115 to 315 ms after stimulus onset, averaged over time, are plotted in Fig. 2. Additionally, the event related fields corresponding to sensors with the highest absolute feature weights are shown. As expected, the regression coefficients are strongest when the difference between the event related fields of the different categories is largest, implying that it is this difference that the classifier is likely to pick up. The features with the highest absolute coefficients are located towards the occipital cortex. The localization pattern seems to show some individual variation. However, this could also be explained by a dipole in the same location but with a different orientation. As sensor-space analysis is sub-optimal in providing more information about localization, the same classification algorithm was applied to source-space activity to further analyze the neuronal generators involved in visual perception.

3.3. Source localization

As can be seen in Fig. 3, applying the classification algorithm to source-space activation time-courses between 115 and 315 ms after

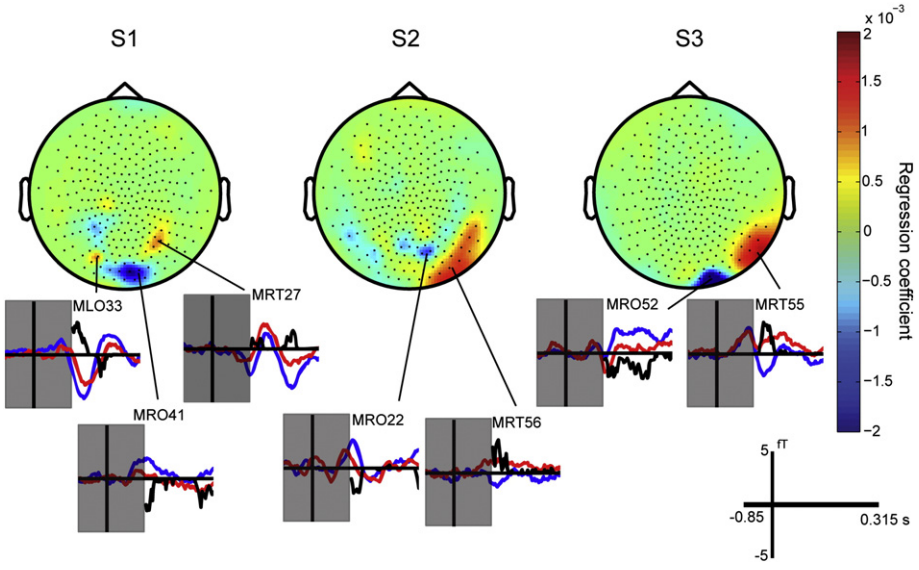

Fig. 2. Top panel: Individual maps per subject of time averaged regression coefficients for the discrimination between faces and tools based on sensor level data 115 to 315 ms after stimulus onset. Red colors indicate positive regression coefficients; blue colors indicate negative regression coefficients. The respective accuracies corresponding to these maps are 0.90, 0.77 and 0.86. Lower panel: event-related fields (ERFs) over the interval starting at 85 ms before stimulus onset until 315 ms after stimulus onset for selected sensors with high absolute regression coefficients. Red traces are the ERFs for tools; blue traces are the ERFs for faces. The black lines denote the regression coefficients per time point at a fixed yet arbitrary scale. The gray box indicates the part of the ERF which was not included in the classification.

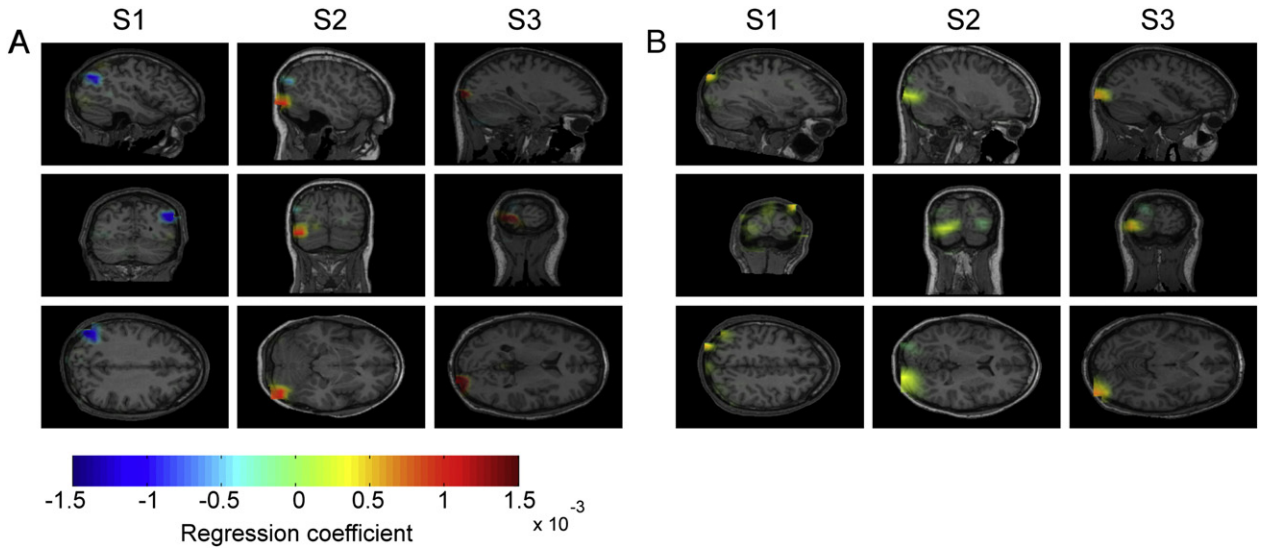

Fig. 3. Individual maps of time averaged regression coefficients for the discrimination between faces and tools based on source-space activity time-courses 115 to 315 ms after stimulus onset. Red colors indicate positive regression coefficients; blue colors indicate negative regression coefficients. A) Regression coefficients of classification based on source-space activity reconstructed with the LCMV beamformer. The respective accuracies corresponding to these maps are 0.87, 0.85 and 0.91. B) Regression coefficients of classification based on source-space activity reconstructed using the dynamic beamformer. The respective accuracies corresponding to these maps are 0.89, 0.85 and 0.96.

stimulus onset resulted in more focal averaged feature maps than when running this same analysis on sensor-space data. The non-zero regression coefficients are still localized in biologically plausible areas in visual cortex. Again there is some individual variation, as the largest source for the first subject lies in the left middle occipital gyrus, whereas for subjects two and three the largest source is positioned in the right inferior occipital gyrus, or even the lingual gyrus. When using the dynamic beamformer to reconstruct the activity time courses the features involved were localized similarly. Very similar localization results were obtained for images that were corrected for luminance only (see Supplementary Fig. 1).

3.4. Classification per time point

We assessed the temporal evolution of image category perception by training and testing classifiers on each source-space activity timepoint separately. This way, accuracy values were obtained for each 3.3 ms. Classification accuracy peaked during the first 100–200 ms after stimulus onset. After this initial peak, accuracy decreased yet remained around a significant deviation from chance level (see Fig. 4). For luminance corrected stimuli the results were again similar (see Supplementary Fig. 2).

This sustained activity in the average traces was also seen when contrasting faces and bodies. However, in all other contrasts the average accuracy trace dropped to chance level after the initial peak. Still, individual traces often showed a sustained effect around the significance threshold when the average trace did not (see Supplementary Fig. 3).

The accuracy traces based on the dynamic beamformer behaved similar to those based on the LCMV beamformer. However, the peaks became more pronounced and the traces were smoother in case of the dynamic beamformer. Indeed, the lag-one autocorrelation for all traces as a measure of smoothness revealed that for eight out of twelve contrasts the accuracy trace based on the dynamic beamformer was smoother ($4.1 < t(2) < 17.5$, FDR corrected $p < 0.05$). The other contrasts did not show a significant effect, but did show a trend in the same direction ($2.1 < t(2) < 3.4$, FDR corrected $p < 0.1$).

Focusing on the onset of the initial peak, we observed that the average classification accuracy rises above chance level as early as 85 ms after stimulus onset (see Table 2). For contrasts between images that were only corrected for luminance, latencies as short as 65 ms were observed (see Supplementary Table 2). Peak average classification accuracy occurred around 130 ms. Note that the longest latencies may not be valid. Because onset latency is defined as the first time point at which the average accuracy trace rises above the FDR-corrected threshold, latencies may not be extracted properly when the peak is very low or absent. After all, in these cases the onset may only be detected at the peak, or around later spurious peaks, lengthening the latency. This is the case for those contrasts where the onset latency is the same as the peak latency, such as with scenes versus bodies and tools.

3.5. Spatiotemporal localization

When classifying on each single time point, a specific vector of regression coefficients is derived for each corresponding accuracy value. Therefore, it is possible to localize the features of importance for classification at a specific time point. This allows identification of which brain regions are discriminative at each single point in time. An example of this analysis for the best performing subject (S3) is shown in Fig. 5. Similar plots are derived for the other subjects (see Supplementary Fig. 4). This analysis reveals that during the initial accuracy peak, 125 to 225 ms after stimulus onset, three main clusters are used by the classifier: one in the right inferior occipital gyrus/lingual gyrus, one in the superior occipital gyrus/cuneus, and a final cluster in the inferior temporal gyrus. During sustained visual stimulation signals in these areas are still used by the classifier. In addition, more anterior regions along the dorsal and ventral streams become incorporated in the classifier model after the initial peak, specifically the inferior parietal/postcentral gyrus and a more anterior region of the inferior temporal gyrus.

To assess the influence of specific regions on classification performance, we applied the classification algorithm to all time points in the interval between 115 and 315 ms after stimulus onset for selected regions of interest. Classifying faces and tools based on only occipital activity time-courses was possible with a high accuracy (mean = 0.89, $sd = 0.06$). Classification on the right temporal lobe was also possible, albeit with a less high accuracy (mean = 0.73, $sd = 0.09$). Finally, applying the classifier algorithm to frontal lobe activity time-courses, where stimulus

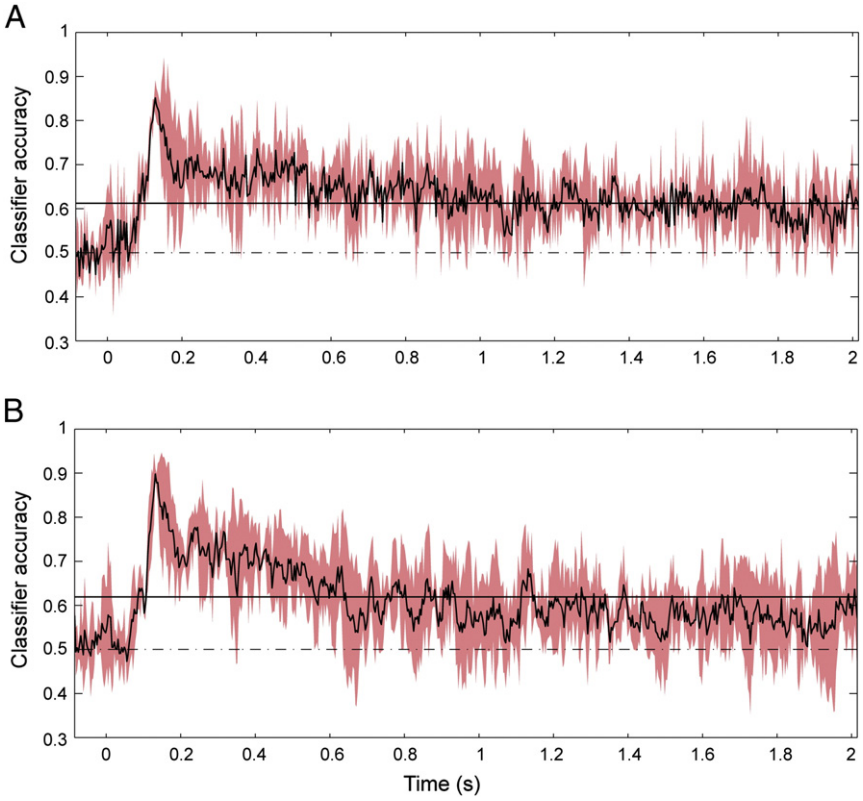

Fig. 4. Average accuracy traces for the contrast faces versus tools. The red areas around the traces indicate the 95% confidence interval. Stimulus onset is at 0 s. The dashed horizontal lines indicate chance level performance. The solid horizontal lines signify the FDR-corrected threshold for deviation from chance level. After the initial peak, classification performance remained sustained around the FDR-corrected threshold for this contrast. A) Average accuracy trace based on source-space activity reconstructed with the LCMV beamformer. The latency for which the trace starts to rise significantly above the FDR-corrected threshold is 85 ms. B) Average accuracy trace based on source-space activity reconstructed with the dynamic beamformer. Again, the latency for which the trace starts to rise significantly above the FDR-corrected threshold is 85 ms.

specific information was not expected, was indeed not possible (mean = 0.55, *sd* = 0.05).

4. Discussion

In this paper, we showed that multivariate classification algorithms can be successfully applied to reconstructed MEG source-space activity time-courses in order to track the spatiotemporal dynamics of visual perception with a high resolution, both in the spatial and in the temporal domain. Overall, classification of different stimulus categories was possible with high accuracy, and resulted from a focal source in the inferior occipital lobe or nearby downstream areas. Moreover, investigating the temporal evolution of classifier accuracy per single time point revealed that the distinction between image categories could be detected in the brain as early as 85 ms after stimulus onset. After an initial peak, this distinction was still detectable for some of the contrasts. Localizing the non-zero regression coefficients per single time point allowed us to capture the fast temporal dynamics of visual category perception in detail. Finally, using source-space activity time-courses reconstructed with the dynamic beamformer as input to the algorithm boosted classifier accuracy as compared to using source-space activity time-courses reconstructed with the LCMV beamformer. The dynamic beamformer also gave smoother time-resolved accuracies as obtained from the single time point analysis. In addition to the analysis of visual perception, this approach to decode source-space MEG activity can also provide a suitable method to investigate high-resolution spatiotemporal dynamics of other ongoing cognitive processes. For example, Sudre et al. (2012) have revealed spatiotemporal dynamics of semantic processing in a manner comparable to ours.

Table 2
Overview of the onset latency at which the average accuracy trace of different contrasts first rises significantly above the FDR-corrected chance level, and the peak latency at which the maximum classification accuracy is reached.

Contrast	LCMV beamformer onset/peak (ms)	Dynamic beamformer onset/peak (ms)
Face–tool	85.0/128.3	85.0/131.7
Face–scene	105/128.3	98.3/128.3
Face–body	88.3/125.0	78.3/138.3
Scene–body	131.7/131.7	125.0/195.0
Body–tool	91.7/198.3	78.3/215.0
Scene–tool	421.7/421.7	171.7/178.3

4.1. Overall classifier performance

A range of different natural image categories could be decoded successfully with our approach. Classification accuracy often exceeded 0.8, and ranged up to over 0.9. Especially faces were strongly dissociable from other image categories, with classification performance consistently exceeding chance level for all subjects and analysis pipelines. This is in line with the idea that faces are processed faster and more efficiently than other objects (Farah et al., 1998). On the other hand, scenes and tools were much harder to dissociate from each other. For this contrast, classification performance did not consistently rise above chance level. This could be explained by scene-related activity likely arising from the parahippocampal place area (Epstein and Kanwisher, 1998). As this area is localized medially in the brain, activity arising from this deep region may be harder to detect with MEG than activity originating in cortical sources. Therefore, scene-related activity may have an appreciably lower signal-to-noise ratio. An attempt to

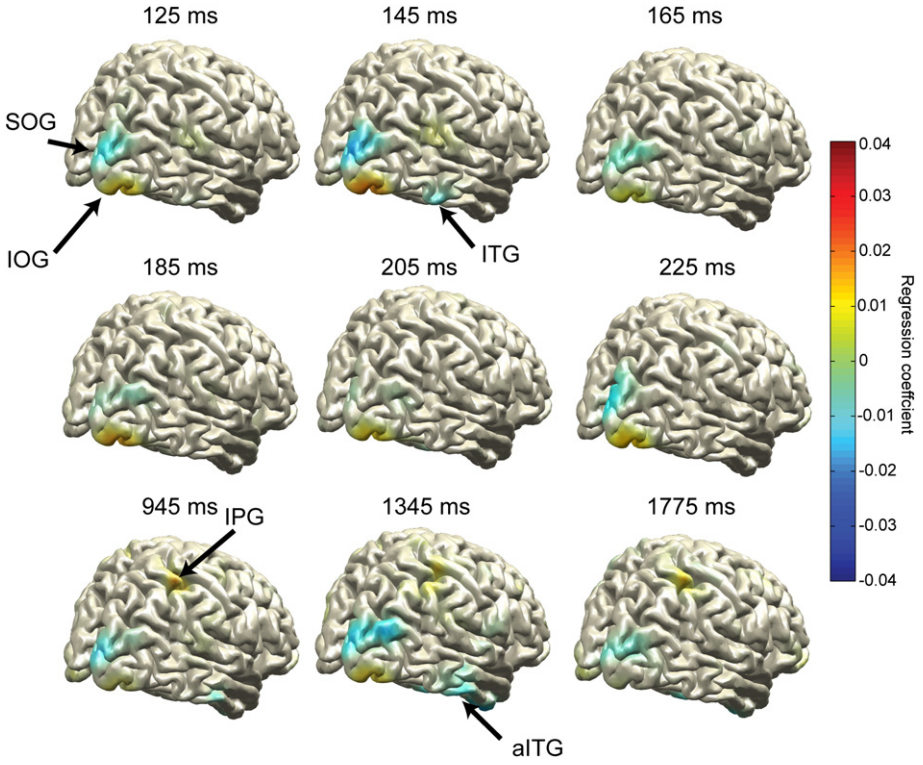

Fig. 5. Localization of regression coefficients for single time samples during the initial accuracy peak and the sustained period for the best subject (S3). Data are shown for the contrast faces versus tools. Source-space activity time-courses were reconstructed with the dynamic beamformer. Note the initial involvement of three clusters, one in the inferior occipital/lingual gyrus (IOG), one in the superior occipital gyrus/cuneus (SOG), and one in the inferior temporal gyrus (ITG). During the sustained period regions more anterior in the dorsal (inferior parietal/postcentral gyrus (IPG)) and ventral streams (more anterior inferior temporal gyrus (aITG)) become of importance to the classifier. Red colors are indicative of positive regression coefficients; blue colors indicate negative coefficients.

discriminate between scenes and any other category would then actually constitute a distinction between category-specific activity and aspecific activity for the scene trials. Hence, classification would rely on information of only one category. If the other category has strong category-specific activity, this distinction may be easily made. However, if this category-specific activity is not so pronounced, or if it resembles the aspecific activity pattern of scene trials, classification accuracy may not always reach significance.

Another possible explanation for contrasts with faces having the highest accuracies could be the correction for spatial frequency. This correction decreased image visibility, and it could be that faces were better recognizable than other categories after correction. This is, however, unlikely. First, when images were only corrected for luminance, which does not affect image visibility, classifier accuracies were still highest when discriminating between faces and other objects. Second, from a larger set only the most visible images were selected for the actual experiment, and all selected images were rated as well recognizable. The difference between faces and the other stimulus categories may also arise from a difference in encoding. However, as the task was unrelated to the images, subjects were not required to actively process the images. A difference in encoding strategy is therefore unlikely.

4.2. Temporal evolution of classifier accuracy

Previous studies have used multivariate methods over multiple short time windows of electrophysiological data to get a handle on the temporal evolution of neuronal activity involved in the perception of stimulus categories (Bode et al., 2012; Carlson et al., 2003; Liu et al., 2009; Ramkumar et al., 2013; Simanova et al., 2010; van Gerven et al., 2013). These studies have shown a very early onset of brain activity related to category perception, starting as early as about 100 ms after stimulus onset. As pointed out by Crouzet et al. (2010), category-specific information can be detected in the brain as early as about 80 ms after stimulus onset. In line with this, we observed above-chance classification as early as 85 ms after stimulus onset. These latencies are plausible in the light of monkey studies that have shown that visual information is detectable in the early occipital cortex around 50 ms after stimulus onset, and about 90 ms after stimulus onset in the temporal lobe (Thorpe and Fabre-Thorpe, 2001).

Next to onset latency, we also inferred how category-specific information evolved during image perception. Accuracy peaked in most contrasts around 130 ms. This is in line with observations in the aforementioned studies, though the peak we found was narrower and occurred somewhat earlier. This may be due to the temporal resolution, which in this study was as high as 3.3 ms and may therefore have resulted in a less smoothed and hence narrower peak. The peak occurred around the P1 visual event-related field, before the face-specific N170 (Bentin et al., 1996). Some studies have shown category-specificity during this P1 component (Itier and Taylor, 2002; Taylor, 2002) while other studies did not show this specificity earlier than the N170 component (Rossion et al., 2003; Rousselet et al., 2007). This study provides further evidence for detectable category-related information already being present as early as the P1 component.

After the initial peak, accuracies were found to decrease again. In some cases (contrasting faces with tools or bodies) accuracy remained significantly above chance level. In other cases, however, average accuracy decreased to chance level during the remainder of stimulus presentation. Still, at the individual level, many contrasts did stay around their individual FDR threshold during this sustained period. It could well be that, as the individual sustained effects fluctuated around the significance threshold, this effect averaged out to values below the conservatively chosen group level threshold. Alternatively, the lack of consistent

sustained activity could be an effect of inattentional blindness (Rees et al., 1999), as subjects solely had to focus on the fixation dot, and therefore paid no attention to the image itself.

It should be noted that the onset latencies differed between the different contrasts. However, peak latency is often stable. This may be a signal-to-noise issue. After all, onset latency is defined as the time between stimulus onset and the first time point at which the average accuracy trace passes the FDR threshold. Contrasts with lower peaks cross this threshold later than contrasts with more pronounced peaks, even when the peak occurs at the same latency. This may also explain the longer latencies observed for contrasts where scenes are involved, as scene-related activity may actually be noise (see above). However, this observation only holds for images corrected for both luminance and spatial frequency. It could be argued that this is an effect of a general decrease in signal-to-noise ratio for images that were corrected for spatial frequency as well.

4.3. Localization of temporal dynamics

In addition to the temporal dynamics that this method allowed us to extract, it also enabled more precise pinpointing of the neuronal sources underlying visual perception. When classifying on all data measured between 115 and 315 ms after stimulus onset, features used by the classifier seemed to specifically lie in the inferior occipital gyrus and lingual gyrus, as well as the middle occipital gyrus for one of the subjects. Compared to sensor-space analysis, using source-space data as classifier input improved localization by rendering more focal, spatially unmixed sources.

Moreover, when applying the classification algorithm exclusively to activity time-courses originating from the occipital lobe, classification accuracy was high, indicating a strong influence of signals from this specific region on the classifier. It is likely that classification in this region of interest was driven by a complex mixture of low-level image properties that make up a natural image. Indeed, because classification was possible with high accuracy on the input images, regardless of applied correction for low-level image properties, it is not possible to rule out the influence of these features. At the same time, when restricting the algorithm to activity originating from sources over the right temporal lobe classification was still possible. The involvement of these sources may be indicative of at least some semantic category-related information being used to drive classification.

More interesting, however, is to assess the spatial features used by the classifier at individual time points in order to determine how neuronal sources evolve over time. This showed that, while classifying on all data in the 115 to 315 ms interval after stimulus onset binned together revealed only one main occipital source, this source was segregated into multiple clusters when assessing this interval per single time point. Indeed, between 125 and 225 ms after stimulus onset three clusters were found to alternately be used to a more or lesser extent by the classifier in the distinction between faces and tools: the inferior occipital/lingual gyrus, the superior occipital gyrus/cuneus, and the inferior temporal gyrus. These areas are biologically plausible regions for category perception. For example, the occipital face area is located in the inferior occipital gyrus and has been shown to be related to face processing (Gauthier et al., 2000; Nichols et al., 2010). Also, the lingual gyrus has been implicated in face-specific ERP responses (Allison et al., 1999; McCarthy et al., 1999). The superior occipital gyrus and cuneus belong to the dorsal stream, which has been shown to be activated by tools (Almeida et al., 2008; Almeida et al., 2010) or at least tool-shaped objects (Sakuraba et al., 2012). The inferior temporal gyrus, on the other hand, is part of the ventral stream and as such related to object identity (Mishkin et al., 1983).

Involvement of these areas was not limited to the initial peak. Also during sustained visual stimulation the classifier used information from these areas. In addition, during the sustained period areas more anterior along the ventral (anterior inferior temporal gyrus) and dorsal (inferior parietal/postcentral gyrus) stream became involved, again being suggestive of the additional influence of semantic category-related information. However, sensitivity of these areas to low-level visual features cannot be ruled out.

There could be an additional effect of field spread. However, during whole-brain analysis multiple separated sources were extracted that were of influence to the classifier. These different sources were moreover observed at separate time scales, indicative of being different underlying sources. Furthermore, classification on frontal sources only is not possible, making the influence of a general effect less likely.

A certain amount of blurring or shifting of sources is possible, as there was some limited head motion in all subjects. However, although the clusters involved in early and sustained visual perception were located close to each other, they could be clearly distinguished using our approach by being assigned regression coefficients with different signs. Therefore it is unlikely that there has been mixing of these sources due to a blurring effect, or even field spread.

4.4. Comparison of source reconstruction methods

Next to classification on source-space activity time-courses reconstructed with the LCMV beamformer, we performed the same analyses on source-space activity time-courses reconstructed using the dynamic beamformer (Bahramisharif et al., 2012). This beamformer method, by taking previous samples into account, results in smoother source-space activity time-courses. Applying the classification algorithm to these activity time-courses resulted in smoother accuracy traces, as well as boosted classification accuracy while maintaining an equal localization performance. This effect can likely be explained by the increased signal-to-noise ratio of the input activity.

We observed a significant increase in classifier performance when applying the algorithm to activity time-courses reconstructed by the dynamic beamformer compared to sensor-space time courses for all contrasts, as well as a trend for the other contrasts. We also observed a trend towards increase in performance when using source-space activity time-courses reconstructed by the LCMV beamformer instead of sensor-space time courses. The lack of strong significant results despite these trends could be a ceiling effect, as classifier accuracies were already quite high for sensor-level data. In addition, with about 49 trials per category an increase in performance of 10% would still be small in the absolute number of additional correctly classified trials, making significant results hard to obtain. Finally, there is a loss in power due to the multiple comparisons resulting from the different contrasts, subjects and data representations. It should be noted that, for one subject, classifier performance sometimes worsened when using source-space instead of sensor-space time courses. This could be caused by a less optimal source space reconstruction. After all, anything that can influence the quality of the source-space signal in its turn influences the accuracy of the classification algorithm.

Observed differences between sensor-space and both source-space accuracies can be explained by differences in regularization. After all, the regularization may be differentially influenced by the various data representations. The average regularization parameter λ was slightly different for contrasts based on sensor-space data (8.0), source-space activity reconstructed with the LCMV beamformer (10.3) and source-space activity reconstructed with the dynamic beamformer (12.3). In addition, the beamformer may suppress noise from unrelated sources, which would then lead to an input with a higher signal-to-noise ratio for the classifier.

4.5. Interpretation of classifier parameters

The magnitude of the regression coefficient for a given feature is indicative of the contribution of that feature to the current classification problem. This, however, does not imply that there is a direct relationship between features of importance to the classifier, and the underlying

neuronal substrates driving category perception (Bießmann et al., 2012). Furthermore, it is possible that features of importance for the classifier are actually noise components (Yamashita et al., 2008). This makes a straightforward interpretation of regression coefficients difficult. However, as the elastic net method imposes a sparsity constraint, only a small number of features have coefficients set to non-zero values. This small set of features is sufficient nonetheless to reach the classifier accuracies described. All other features are not involved in that specific classification, and hence not necessary to obtain high classifier accuracies.

Not only the magnitude, but also the sign of the regression coefficients has to be interpreted with care. For example, the classification algorithm may decide in favor of tools when the sum of feature values multiplied by their regression coefficients is positive, whereas the decision is in favor of faces when this weighted sum has a negative value. However, this does not mean that negative regression coefficients are indicative of 'face-features' and positive regression coefficients signify 'tool-features'. For example, a negative regression coefficient at a negative ERF component results in a net positive coefficient, hence pushing the classifier decision to the opposite category.

The focal nature of the obtained sources can partially be explained in terms of properties of the employed classification algorithm. The elastic net algorithm is regularized by an L1 and L2 norm, which respectively favor few focal sources and multiple smoother sources with lower corresponding regression coefficients. We set the mixing parameter α such that the model incorporated strong L2 regularization and weak L1 regularization, in order to be able to detect distributed, yet focal sources. Indeed, we often found multiple focal sources involved during the different stages of perception. Adjusting the mixing parameter α towards stronger L1 regularization resulted in a sparser model in which sources were indeed more focal (see Supplementary Fig. 5).

We observed minor individual differences in localization. For example, in the overall classification results, one subject showed a contribution of the middle occipital gyrus to the classifier, whereas for the other two subjects this contribution was located more inferiorly, as well as in the opposite hemisphere. These differences could be explained in terms of variation in anatomy, resulting in different dipole configurations and therefore differences in signal quality. The algorithm may then favor different clusters that have a more distinct signal. Still, this does not make the localizations we find arbitrary. After all, classifier accuracies for all these models were high, and hence the selected sources can be regarded as highly relevant for visual perception. In addition, adjusting the regularization parameter α does not change the location of the source, only the sparsity of that source. Also, intrasubject localization was similar for source-space activity reconstructed with the LCMV beamformer and reconstructed with the dynamic beamformer, and different amounts of correction of low-level image properties also showed similar plausible sources. Therefore, localization seemed to be robust in the light of changes in parameter settings and representation of input data.

To summarize, we showed that by applying a multivariate classification algorithm to source-space activity time-courses, we were able to investigate the spatiotemporal dynamics of visual perception. Not only did we use the temporal resolution of MEG data to pinpoint the early onset and temporal evolution of perceptual information, we also localized this temporal pattern to focal biologically plausible sources. By optimally combining the spatial and temporal domain, this method allowed us to study the spatiotemporal dynamics of perception, in order to assess the evolution of neural sources involved over time.

Acknowledgments

This research was funded by the Netherlands Organisation for Scientific Research (NWO-MaGW) under grant number 404.10.500.

Conflict of interest

The authors declare that there are no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.07.075.

References

Ales, J.M., Appelbaum, L.G., Cottereau, B.R., Norcia, A.M., 2012. The time course of shape discrimination in the human brain. NeuroImage 67, 77–88.
Allison, T., Puce, A., Spencer, D.D., McCarthy, G., 1999. Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. Cereb. Cortex 9 (5), 415–430.
Almeida, J., Mahon, B.Z., Nakayama, K., Caramazza, A., 2008. Unconscious processing dissociates along categorical lines. Proc. Natl. Acad. Sci. U. S. A. 105 (39), 15214–152148.
Almeida, J., Mahon, B.Z., Caramazza, A., 2010. The role of the dorsal visual processing stream in tool identification. Psychol. Sci. 21 (6), 772–778.
Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., Ohtani, Y., 2006. Estimation of the timing of human visual perception from magnetoencephalography. J. Neurosci. 26 (15), 3981–3991.
Bacon-Macé, N., Macé, M.J.-M., Fabre-Thorpe, M., Thorpe, S.J., 2005. The time course of visual processing: backward masking and natural scene categorisation. Vis. Res. 45 (11), 1459–1469.
Bahramisharif, A., Van Gerven, M., Schoffelen, J.-M., Ghahramani, Z., Heskes, T., 2012. The dynamic beamformer. Machine learning and interpretation in neuroimaging. Lect. Notes Comput. Sci. 7263, 148–155.
Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. 8 (6), 551–565.
Bießmann, F., Dähne, S., Meinecke, F., 2012. On the interpretability of linear multivariate neuroimaging analyses: filters, patterns and their relationship. NIPS Workshop on Machine Learning and Inference in Neuroimaging, pp. 1–6.
Bode, S., Sewell, D.K., Lilburn, S., Forte, J.D., Smith, P.L., Stahl, J., 2012. Predicting perceptual decision biases from early brain activity. J. Neurosci. 32 (36), 12488–12498.
Brodeur, M.B., Dionne-Dostie, E., Montreuil, T., Lepage, M., 2010. The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PLoS One 5 (5), e10773.
Carlson, T., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15 (5), 704–717.
Carlson, T., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. J. Vis. 11 (10), 1–17.
Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., Rao, A.R., 2009. Prediction and interpretation of distributed neural activity with sparse models. NeuroImage 44, 112–122.
Chan, A.M., Halgren, E., Marinkovic, K., Cash, S.S., 2011. Decoding word and category-specific spatiotemporal representations from MEG and EEG. NeuroImage 54 (4), 3028–3039.
Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618.
Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19 (2), 261–270.
Crouzet, S.M., Thorpe, S.J., 2011. Low-level cues and ultra-fast face detection. Front. Psychol. 2 (342), 1–9.
Crouzet, S.M., Kirchner, H., Thorpe, S., 2010. Fast saccades towards faces: face detection in just 100 ms. J. Vis. 10 (4), 1–17.
De Gelder, B., Van den Stock, J., 2011. The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. Front. Psychol. 2 (181), 1–6.
Downing, P.E., Chan, A.W.-Y., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. Cereb. Cortex 16 (10), 1453–1461.
Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.
Farah, M.J., Wilson, K.D., Drain, M., Tanaka, J.N., 1998. What is "special" about face perception? Psychol. Rev. 105 (3), 482–498.
Fei-Fei, L., Perona, P., 2005. A Bayesian hierarchical model for learning natural scene categories. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, pp. 524–531.
Fellinger, R., Gruber, W., Zauner, A., Freunberger, R., Klimesch, W., 2012. Evoked traveling alpha waves predict visual-semantic categorization-speed. NeuroImage 59 (4), 3379–3388.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.
Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., Düzel, E., 2010. Theta-coupled periodic replay in working memory. Curr. Biol. 20 (7), 606–612.
Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. The fusiform "face area" is part of a network that processes faces at the individual level. J. Cogn. Neurosci. 12 (3), 495–504.
Goeleven, E., De Raedt, R., Leyman, L., Verschuere, B., 2008. The Karolinska directed emotional faces: a validation study. Cogn. Emot. 22 (6), 1094–1118.
Guimaraes, M.P., Wong, D.K., Uy, E.T., Grosenick, L., Suppes, P., 2007. Single-trial classification of MEG recordings. IEEE Trans. Biomed. Eng. 54 (3), 436–443.

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 (5539), 2425–2430.
Haynes, J.-D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8 (5), 686–691.
Itier, R.J., Taylor, M.J., 2002. Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. NeuroImage 15 (2), 353–372.
Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb. Cortex 14 (2), 132–142.
Jafarpour, A., Horner, A.J., Fuentemilla, L., Penny, W.D., Düzel, E., 2013. Decoding oscillatory representations and mechanisms in memory. Neuropsychologia 51 (4), 772–780.
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8 (5), 679–685.
Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. Nature 452 (7185), 352–355.
Kirchner, H., Thorpe, S.J., 2006. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vis. Res. 46 (11), 1762–1776.
Kirchner, H., Barbeau, E.J., Thorpe, S.J., Régis, J., Liégeois-Chauvel, C., 2009. Ultra-rapid sensory responses in the human frontal eye field region. J. Neurosci. 29 (23), 7599–7606.
Kriegeskorte, N., 2011. Pattern-information analysis: from stimulus decoding to computational-model testing. NeuroImage 56, 411–421.
Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103 (10).
Lange, N., Strother, S.C., Anderson, J.R., Nielsen, F.A., Holmes, A.P., Kolenda, T., Savoy, R., Hansen, L.K., 1999. Plurality and resemblance in fMRI data analysis. NeuroImage 10, 282–303.
Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62 (2), 281–290.
Lundqvist, D., Flykt, A., Öhman, A., 1998. The Karolinska Directed Emotional Faces — KDEF. Department of Neurosciences Karolinska Hospital, Stockholm.
McCarthy, G., Puce, A., Belger, A., Allison, T., 1999. Electrophysiological studies of human face perception. II: response properties of face-specific potentials generated in occipitotemporal cortex. Cereb. Cortex 9 (5), 431–444.
Mishkin, M., Ungerleider, L., Macko, K., 1983. Object vision and spatial vision: two cortical pathways. Trends Neurosci. 6, 414–417.
Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320 (5880), 1191–1195.
Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H.C., Sadato, N., et al., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60 (5), 915–929.
Murphy, B., Baroni, M., Poesio, M., 2009. EEG responds to conceptual stimuli and corpus semantics. Proc. ACL/EMNLP, pp. 619–627.
Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., Lakany, H., 2011. EEG decoding of semantic category reveals distributed representations for single concepts. Brain Lang. 117 (1), 12–22.
Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L., 2009. Bayesian reconstruction of natural images from human brain activity. Neuron 63 (6), 902–915.
Nichols, D.F., Betts, L.R., Wilson, H.R., 2010. Decoding of faces and face components in face-sensitive human visual cortex. Front. Psychol. 1 (28), 1–13.
Noguchi, Y., Tanabe, H.C., Sadato, N., Hoshiyama, M., Kakigi, R., 2007. Voluntary attention changes the speed of perceptual neural processing. Eur. J. Neurosci. 25, 3163–3172.
Nolte, G., 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys. Med. Biol. 48 (22), 3637–3652.
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, 1–9.
Petrov, Y., Nador, J., Qian, J., 2012. VEP correlates of feedback in human cortex. PLoS One 7 (12), e51791.
Ramkumar, P., Jas, M., Pannasch, S., Hari, R., Parkkonen, L., 2013. Feature-specific information processing precedes concerted activation in human visual cortex. J. Neurosci. 33 (18), 7691–7699.
Rees, G., Russell, C., Frith, C., Driver, J., 1999. Inattentional blindness versus inattentional amnesia for fixated but ignored words. Science 286 (5449), 2504–2507.
Reynolds, J.H., Pasternak, T., Desimone, R., 2000. Attention increases sensitivity of V4 neurons. Neuron 26, 703–714.
Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2 (11), 1019–1025.
Rossion, B., Joyce, C.A., Cottrell, G.W., Tarr, M.J., 2003. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. NeuroImage 20 (3), 1609–1624.
Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2007. Single-trial EEG dynamics of object and face visual processing. NeuroImage 36 (3), 843–862.
Sakuraba, S., Sakai, S., Yamanaka, M., Yokosawa, K., Hirayama, K., 2012. Does the human dorsal stream really process a category for tools? J. Neurosci. 32 (11), 3949–3953.
Salzberg, S., 1997. On comparing classifiers: pitfalls to avoid and a recommended approach. Data Min. Knowl. Discov. 1, 317–328.
Sandberg, K., Bahrami, B., Kanai, R., Barnes, G.R., Overgaard, M., Rees, G., 2013. Early visual responses predict conscious face perception within and between subjects during binocular rivalry. J. Cogn. Neurosci. 25 (6), 969–985.
Seeck, M., Michel, C.M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., Schomer, D.L., 1997. Evidence for rapid face recognition from human scalp and intracranial electrodes. Neuroreport 8 (12), 2749–2754.
Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U. S. A. 104 (15), 6424–6429.
Simanova, I., Van Gerven, M., Oostenveld, R., Hagoort, P., 2010. Identifying object categories from event-related EEG: toward decoding of conceptual representations. PLoS One 5 (12), e14465.
Stolk, A., Todorovic, A., Schoffelen, J.-M., Oostenveld, R., 2013. Online and offline tools for head movement compensation in MEG. NeuroImage 68, 39–48.
Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., Mitchell, T., 2012. Tracking neural coding of perceptual and semantic features of concrete nouns. NeuroImage 62, 451–463.
Suppes, P., Lu, Z.L., Han, B., 1997. Brain wave recognition of words. Proc. Natl. Acad. Sci. U. S. A. 94 (26), 14965–14969.
Suppes, P., Han, B., Epelboim, J., Lu, Z.L., 1999. Invariance of brain-wave representations of simple visual images and their names. Proc. Natl. Acad. Sci. U. S. A. 96 (25), 14658–14663.
Taylor, M.J., 2002. Non-spatial attentional effects on P1. Clin. Neurophysiol. 113 (12), 1903–1908.
Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., Dehaene, S., 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. NeuroImage 33 (4), 1104–1116.
Thorpe, S.J., Fabre-Thorpe, M., 2001. Seeking categories in the brain. Science 291, 260–262.
Thorpe, S.J., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.
Van Gerven, M.A.J., Cseke, B., De Lange, F.P., Heskes, T., 2010a. Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior. NeuroImage 50 (1), 150–161.
Van Gerven, M.A.J., De Lange, F.P., Heskes, T., 2010b. Neural decoding with hierarchical generative models. Neural Comput. 22 (12), 3127–3142.
Van Gerven, M.A.J., Maris, E., Sperling, M., Sharan, A., Litt, B., Anderson, C., Baltuch, G., et al., 2013. Decoding the memorization of individual stimuli with direct human brain recordings. NeuroImage 70, 223–232.
Van Veen, B.D., Van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44 (9), 867–880.
VanRullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. J. Cogn. Neurosci. 13 (4), 454–461.
Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., Tanaka, J.W., 2010. Controlling low-level image properties: the SHINE toolbox. Behav. Res. Methods 42 (3), 671–684.
Wokke, M.E., Sligte, I.G., Steven Scholte, H., Lamme, V.A.F., 2012. Two critical periods in early visual cortex during figure-ground segregation. Brain Behav. 2 (6), 763–777.
Wu, W., Gao, S., 2011. Learning event-related potentials (ERPs) from multichannel EEG recordings: a spatio-temporal modeling framework with a fast estimation algorithm. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2011, 6959–6962.
Yamashita, O., Sato, M., Yoshioka, T., Tong, F., Kamitani, Y., 2008. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. NeuroImage 42, 1414–1429.