Multiclass fMRI data decoding and visualization using supervised self-organizing maps

Citation for published version (APA):

Hausfeld, L., Valente, G., & Formisano, E. (2014). Multiclass fMRI data decoding and visualization using supervised self-organizing maps. *Neuroimage*, 96, 54-66.
<https://doi.org/10.1016/j.neuroimage.2014.02.006>

Document status and date:

Published: 01/08/2014

DOI:

[10.1016/j.neuroimage.2014.02.006](https://doi.org/10.1016/j.neuroimage.2014.02.006)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Multiclass fMRI data decoding and visualization using supervised self-organizing maps

Lars Hausfeld^{*}, Giancarlo Valente, Elia Formisano

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands

ARTICLE INFO

Article history:

Accepted 3 February 2014

Available online 12 February 2014

Keywords:

fMRI
Decoding
Multiclass classification
Self-organizing maps

ABSTRACT

When multivariate pattern decoding is applied to fMRI studies entailing more than two experimental conditions, a most common approach is to transform the multiclass classification problem into a series of binary problems. Furthermore, for decoding analyses, classification accuracy is often the only outcome reported although the topology of activation patterns in the high-dimensional features space may provide additional insights into underlying brain representations. Here we propose to decode and visualize voxel patterns of fMRI datasets consisting of multiple conditions with a *supervised* variant of self-organizing maps (SSOMs). Using simulations and real fMRI data, we evaluated the performance of our SSOM-based approach. Specifically, the analysis of simulated fMRI data with varying signal-to-noise and contrast-to-noise ratio suggested that SSOMs perform better than a *k*-nearest-neighbor classifier for medium and large numbers of features (i.e. 250 to 1000 or more voxels) and similar to support vector machines (SVMs) for small and medium numbers of features (i.e. 100 to 600 voxels). However, for a larger number of features (>800 voxels), SSOMs performed worse than SVMs. When applied to a challenging 3-class fMRI classification problem with datasets collected to examine the neural representation of three human voices at individual speaker level, the SSOM-based algorithm was able to decode speaker identity from auditory cortical activation patterns. Classification performances were similar between SSOMs and other decoding algorithms; however, the ability to visualize decoding models and underlying data topology of SSOMs promotes a more comprehensive understanding of classification outcomes. We further illustrated this visualization ability of SSOMs with a re-analysis of a dataset examining the representation of visual categories in the ventral visual cortex (Haxby et al., 2001). This analysis showed that SSOMs could retrieve and visualize topography and neighborhood relations of the brain representation of eight visual categories. We conclude that SSOMs are particularly suited for decoding datasets consisting of more than two classes and are optimally combined with approaches that reduce the number of voxels used for classification (e.g. region-of-interest or searchlight approaches).

© 2014 Published by Elsevier Inc.

Introduction

During the past years, the traditional voxel-wise analysis of fMRI data has been complemented by so-called Multi-Voxel Pattern Analysis (MVPA; e.g. Haynes and Rees, 2006; Norman et al., 2006). In contrast to the conventional activation-based statistical analysis, which examines each voxel separately for activation differences between experimental conditions using, for example, the General Linear Model (GLM; Friston, 1995), MVPA represents an information-based analysis (Kriegeskorte et al., 2006) that exploits information contained in spatial (i.e. multi-voxel) activation patterns.

Various decoding algorithms have been proposed in MVPA fMRI studies. These algorithms differ in terms of their complexity and include

correlation-based approaches (Haxby et al., 2001), Gaussian-naïve Bayes classifiers (Mitchell et al., 2004), linear discriminant analysis (Cox and Savoy, 2003; Kriegeskorte et al., 2006), linear and non-linear support vector machines (SVMs; Cox and Savoy, 2003; LaConte et al., 2005; Mourão-Miranda et al., 2005) and sparse logistic regression (SLR; Miyawaki et al., 2008; Ryali et al., 2010; Yamashita et al., 2008). More recently, the combination of classification algorithms and voxel selection strategies (e.g. De Martino et al., 2008; Langa et al., 2011; Yamashita et al., 2008) and use of multiple classifiers has been proposed (e.g. Kuncheva and Rodríguez, 2010).

In this paper, we present an fMRI decoding approach based upon self-organizing maps (SOMs). SOMs were developed to visualize high-dimensional data by converting the topology of data points into simple geometrical relationships on a two-dimensional grid (Kohonen, 2001). By preserving only the most important topological relationships, this algorithm abstracts from high-dimensional input and provides insight into the underlying data structure. These properties rendered it an important tool for data exploration in various domains. *Supervised* SOMs

^{*} Corresponding author at: Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University. P.O. Box 616, 6200 MD Maastricht, The Netherlands.

E-mail address: lars.hausfeld@maastrichtuniversity.nl (L. Hausfeld).

(SSOMs) inherit these characteristics and extend the SOMs such that they can classify unseen samples. The decoding of fMRI data with SSOMs differs with respect to other approaches in two relevant aspects.

First, SSOMs are not restricted to binary comparisons and thus allow for inherent multiclass decoding (i.e. when the fMRI measurements include more than two conditions). Many studies that analyzed fMRI data by means of MVPA employed experimental designs with two conditions or translated multiclass problems into a series of binary comparisons (one-versus-one or one-versus-all schema) paired with post-hoc processing (e.g. majority voting) to determine the predicted class (e.g. Beauchamp et al., 2009; Cox and Savoy, 2003; Ethofer et al., 2009; Kamitani and Tong, 2005; Mourão-Miranda et al., 2006; Reddy et al., 2010; Swisher et al., 2010; Walther et al., 2009). One reason is that one of the most applied classification algorithms in MVPA fMRI applications, the SVM, is in its basic form restricted to two class problems (but see Crammer and Singer (2002) and Weston and Watkins (1999) for multiclass SVM formulations and Martínez-Ramón et al. (2006) for decoding of fMRI data using multiclass SVM). So far, only few fMRI studies applied MVPA in a multiclass setting without employing binary comparisons and necessary post-processing. In these studies decoding was performed with naïve Bayes (Brouwer and Heeger, 2009; Mitchell et al., 2004) or neural network classifiers (Hanson et al., 2004; Polyn et al., 2005).

Second, SSOMs offer the possibility to visualize the underlying distribution of fMRI activity patterns used to train the decoding model. These visualizations can reveal the topology of the underlying activity patterns in high-dimensional feature (voxel) space. When analyzing fMRI activation patterns by MVPA, the experimenter is interested in the degree of separation between classes in feature space. In the case of two classes, the interpretation of classification results is straightforward, i.e. the accuracy of predictions describes the degree of separation of the response patterns. However, for classification of response patterns with more than two classes, focusing on overall classification accuracy or performances of many binary comparisons provides only a partial, non-intuitive view on class distributions in high-dimensional voxel space. Thus, when performing multiclass classification an additional post-processing step is usually required to infer and visualize underlying class distributions (or topology), e.g. via confusion matrices. For example, Abdi et al. (2009) visualized class distributions in a multiclass setting by first computing pairwise comparisons between classes and visualizing, based on these outcomes, the underlying topology using Principal Component Analysis (PCA) and a bootstrapped Multidimensional Scaling (MDS). It should be stressed that, in these cases, class distributions are deduced using the classification performance of the decoding algorithm. Conversely, SSOMs visualize the data directly and thus do not require additional post processing. Another approach that allows visualizing data in fMRI is representational similarity analysis (RSA; Kriegeskorte et al., 2008a, 2008b). In RSA, representational dissimilarity matrices (RDMs) are calculated using pairwise distances (e.g. correlation distance) between activity patterns evoked by distinct stimulus classes. Subsequently, topology of activity patterns can be visualized by MDS of the dissimilarity matrices. A notable distinction compared to SSOMs is that RDMs are calculated from the entire feature set with equal weighting, whereas SSOMs-based topologies rely on the optimized combination of features, which results from learning of the stimulus labels. Thus, SSOM-based visualization may lead to a better abstraction of high-dimensional activity patterns.

Below we illustrate how SSOMs can be used in the context of fMRI decoding to perform multiclass classification and visualize class topology. We present results on simulated fMRI data to illustrate the validity and usefulness of SSOM classification and visualization in the decoding of single data sets as well as in analyses involving multiple cross-validation splits and/or multiple subjects. Furthermore, the proposed method is applied to the analysis of two fMRI studies. First, we examined the performances of our SSOM-based method in a challenging

case of brain-based decoding of speaker identity using new fMRI data collected while participants listened to short non-linguistic vocalizations from three speakers (*dataset 1*, 3-class problem). Second, we re-analyzed publicly available fMRI data from the Haxby et al. (2001) study, where participants were presented with pictures of eight different object categories (Haxby et al., 2001; *dataset 2*, 8-class problem). Many other classification strategies have already been tested using this dataset, which thus provides a good benchmark for evaluating our method. Furthermore, the rich variety of visual categories employed allows showcasing the added value of visualizing the class topologies using SSOMs. For both simulated and real fMRI data, we compared the classification performances of SSOMs to linear and non-linear SVMs and a kNN classifier.

Methods

Self-organizing maps (SOMs)

Self-organizing maps (also called Kohonen maps or networks) are a special type of neural network that was first proposed by (Kohonen, 1982). Several properties have rendered it an important tool for exploration, visualization and abstraction of high-dimensional data (Kohonen, 2001). The SOM typically consists of a two-dimensional rectangular grid of nodes or units (in the following nodes and units are exchangeable terms) each associated with a model of the high-dimensional input data. The weights of these models are iteratively adapted during a learning process, which changes models to optimally span the range of input data. SOM nodes with their associated weights organize such that similar patterns in the high-dimensional space are grouped in clusters using a non-linear competitive learning strategy (Kohonen, 2001). More specifically, nodes that are close to each other in the two-dimensional SOM after training represent patterns of the input space that are similar (i.e. have a small distance). Furthermore, regions in high-dimensional feature space that are more densely populated are represented by more nodes compared to sparse regions with few data points. These properties arise from the learning process that occurs in three main stages (Kaltch et al., 2008). First, the nodes are initialized to span the range of input values for each dimension. Second, the competitive, unsupervised learning scheme assigns each input pattern to the closest node, i.e. the *best matching unit* (BMU), using Euclidean distance in most implementations. Then, the weight vectors of the BMU and neighboring units defined by a neighborhood function are updated to better match the input pattern (i.e. these associated models are shifted towards the input pattern). These training steps are repeated until a specified number of iterations or a convergence criterion is met (e.g. Cheng, 1997; Kohonen, 2001). In a third step, the resulting nodes and attached models can be visualized using several strategies (e.g. Vesanto, 1999).

Formally, a SOM consists of a rectangular two-dimensional grid with U units. Each unit i is described by a N -dimensional model or weight vector $\mathbf{m}_i = [m_{i1}, \dots, m_{iN}]$ where N is the number of input features. The basic organization of SOMs is usually a rectangular or hexagonal lattice (i.e. SOM nodes have four or six neighbors, respectively). We used a hexagonal lattice because they are preferable for visualization purposes (Kohonen, 2001; Vesanto, 1999). For SOM learning, training samples $\mathbf{x}_k = [x_{k1}, \dots, x_{kN}]$ ($k = 1, \dots, K$) are iteratively presented and the best-matching unit (BMU) $\mathbf{m}_{BMU}$ is selected according to smallest distance, i.e.

$$\|\mathbf{x}_k - \mathbf{m}_{BMU}\| = \min_i (\|\mathbf{x}_k - \mathbf{m}_i\|), \quad (1)$$

where $\|\cdot\|$ denotes Euclidean distance. In the following, weights of map units are modified with the following update rule:

$$\mathbf{m}_{t+1} = \mathbf{m}_t + \alpha_t h_{BMU}(r_t) \|\mathbf{x}_k - \mathbf{m}_i\|, \quad (2)$$

where t denotes the learning iteration, α_t the learning rate and $h_{\text{BMU}}(r_t)$ the Gaussian neighborhood kernel around winning unit $\mathbf{m}_{\text{BMU}}$ with radius r_t . After weight adaption, BMUs are redefined with the new unit-specific weight vectors. Both learning rate and radius of the neighborhood are decreasing functions over time. This learning leads to an early stage that sets the general layout of the map by allowing large adjustments and a fine-tuning stage with small changes. In this paper we used the batch computation of the SOM and not the stepwise, recursive algorithm as it is faster and more robust (e.g. Kohonen, 2001). The MATLAB-based SOM-toolbox (<http://www.cis.hut.fi/somtoolbox/>) was used for SSOM training.

Supervised SOMs (SSOMs)

Kohonen (2001) outlined how unsupervised SOMs could be modified to predict unseen instances (see Wongravee et al. (2010) and Xiao et al. (2005, 2006) for applications of this type of SSOM). The basic idea is to append a vector that specifies class membership to the SOM input. In this way, the training incorporates both the measured data (multivoxel patterns) as well as the class label for model learning simultaneously, i.e. the algorithm does not differentiate between labels and input data. Melssen et al. (2006) proposed two alternatives for a supervised SOM algorithm. Both of these approaches consist of two separate SOMs, one representing the input data (multivoxel patterns in this paper) and the other reflecting the assigned class labels. During the training process a shared BMU or separate BMUs for the two maps are determined by taking distances in both SOMs into account. The weight vectors of SOM units for both maps are updated as in usual SOMs. In this paper, we apply the supervised SOM as proposed by Kohonen (2001) mainly for lower computational costs.

The extension suggested by Kohonen (2001) is based on a modified input vector $\mathbf{x}_k^* = [\mathbf{x}_k \ \mathbf{c}_k]$ for model training that results from concatenating input trials $\mathbf{x}_k$ and a C -dimensional (C denotes the number of classes) class-vector $\mathbf{c}_k = [c_{k1}, \dots, c_{kC}]$ where $c_{ki} = 1$ if trial k belongs to class i and $c_{kj} = 0$ ($j \neq i$) otherwise (1-of- K coding; Bishop, 2006). Similarly, a vector $\mathbf{v}_i = [v_{i1}, \dots, v_{iC}]$ is appended to the weight vectors of SSOM-units $\mathbf{m}_i$ to form $\mathbf{m}_i^* = [\mathbf{m}_i \ \mathbf{v}_i]$ with $N + C$ elements. After SSOM training, map units are ascribed to one class by inspecting the last C elements of the map weight vectors $\mathbf{v}_i$: the index with the largest value determines the label of map unit $\mathbf{m}_i$. In general, class-vectors $\mathbf{c}_k$ have norm τ (e.g. $\mathbf{c}_k = [0 \ \tau \ 0]$ would be the class-vector coding the second class out of three). Large τ lead to better class separation of the supervised SOM but simultaneously increase the risk that SSOMs reflect the 'artificial' concatenated inputs $\mathbf{x}_k^*$ rather than the original inputs $\mathbf{x}_k$, which might lead to poor generalization performance. The parameters τ and U for the two fMRI datasets were set according to results from the respective simulated datasets ($\tau = 0.2$ and $U = 64$ for dataset 1; $\tau = 1.0$ and $U = 100$ for dataset 2) and not optimized using measured data. Alternatively τ and U could have been selected using grid search-like strategies in nested cross-validation. Here, the parameters chosen based on the simulations provided a good trade-off between data-driven organization and supervision leading to reasonable degree of clustering necessary for generalization and visualization (see supplemental material S1 for choices of τ and U). SOM units were initialized with weights that corresponded to eigenvectors of the largest U eigenvalues of the training data.

For the prediction of testing trials, the elements of the weight vectors containing class information are detached and unseen instances $\mathbf{x}_{\text{test}}$ (i.e. without class vector $\mathbf{c}_{\text{test}}$) are presented. In contrast to the usual approach in which a trial is classified according to the label of the BMU, we employed the 10 best-matching units (10-BMUs) to accumulate evidence for classification (see Fig. S3 for results of k -BMUs). As SOMs reflect the topology underlying the input data, using k -BMUs ($k > 1$) can be seen as taking the neighborhood of the BMU into account,

which leads to more robust classification performances (e.g. Hausfeld et al., 2012; Silva and Del-Moral-Hernandez, 2011). In particular, we computed a classification index

$$CI_c = \sum_{i=1 \dots 10} (v_{ic} \cdot \exp(-\|\mathbf{x}_{\text{test}} - \mathbf{x}_{\text{BMU}i}\|^2)), \quad (3)$$

to obtain evidence that trial $\mathbf{x}_{\text{test}}$ belongs to class c (v_{ic} denotes the class specific certainty of i th best-matching unit and $\mathbf{x}_{\text{BMU}i}$ is the model of the i th best-matching unit). The supervised SOM predicts an unseen trial according to the class obtaining largest CI .

Multi-split and multi-subject SSOMs

One main advantage of using SSOMs is that these offer opportunities to visualize decoding models (see Fig. 1B). In this study, we visualized the decoding model, i.e. the SSOM, and results by projecting the weight vectors $\mathbf{m}_i$ onto two dimensions using principal component analysis (PCA; see e.g. Vesanto (1999) and Xiao et al. (2005) for other types of visualization). We chose the PCA-based approach because it is a simple and comprehensive linear approach that, in most cases, provided similar visualizations compared to non-linear mappings like multi-dimensional scaling (MDS).

Fig. 1. Flowchart of data analysis and creation of general SSOMs. Panel A presents the main processing components employed in this study. After data acquisition and preprocessing, the dataset was divided according to a k -fold cross-validation scheme into an independent training and testing set. A univariate selection of features (i.e. voxels) via a GLM and subsequent multivariate selection using an SVM-based approach was performed on the training data (please note that the voxel selection steps were only used for analyzing dataset 1). The set of voxels surviving feature selection were used to construct the testing set. Next, different decoding algorithms were applied to create classification models that were validated using the validation set. Panel B shows the different processing steps to derive a general SSOM. Models of SSOMs (obtained in the *Classification Models* step in panel A) were aligned to be represented in a common high-dimensional space with procrustes transformation. Then, simulated annealing was employed to match nodes of single SSOMs that, in the end, were used to construct the general SSOM by averaging node membership and testing trial occupation.

SSOMs are trained for each cross-validation split and each of these single-split SSOM possesses its own topology. In order to have an understanding of the common topology across cross-validations it is necessary to create a *multi-split SSOM*, i.e. a representation that generalizes from SSOMs of single splits. Similarly, to identify the common topology of SSOMs across single subjects requires establishing a *multi-subject SSOM*. Here, we suggest abstracting from single to multiple SSOM properties (i.e. weight vectors, label certainties, and training/testing trial occupation) with a two-step procedure: First, single-split or subject-specific response patterns are transformed into a common space and, second, an optimal correspondence between the nodes of single SSOMs is found.

SSOM alignment

The first obstacle to create general SSOMs is the fact that underlying activation patterns might not be based on the same voxels (e.g. in this study voxel selection depends on properties derived from the training set and not a region-of-interest [ROI] or searchlight approach for which the same voxels are used across splits). To resolve this alignment problem we follow an approach recently proposed by Haxby et al. (2011) that was developed to align response patterns of different participants by creating a common space. Here, a procrustean transformation is applied to align the centroids of all nodes with the same label of single SSOMs to find a transformation that maps SSOM nodes into a common space. The parameters for the linear transformations (i.e. rotation and translation; we did not include a scaling parameter as data were already normalized) are estimated in a three-step process (cf. Haxby et al., 2011). First, the single SSOM weight vectors are aligned to an iteratively updated reference to form an *initial reference SSOM*. A randomly chosen single SSOM served as the first reference. Consecutively, this reference was updated by the average of the old reference and the newly aligned SSOM. The reference obtained in the end of this stage formed the *initial reference*. For the second processing stage, all single SSOMs were aligned to the *initial reference SSOM* and the average of the aligned single SSOMs formed the *final reference*. As a last step, single SSOMs were aligned to the *final reference*.

The estimated parameters of the final procrustean transformation for each SSOM were used to convert the weight vectors of single SSOMs to a common space. Note that for cases in which the same voxels are selected across splits the alignment via procrustean transformation is not needed and can be skipped.

Node matching

In this step, the aim is to find a matching of nodes across different splits (or subjects) to form the *general SSOM*. Once the matching is found, node properties (i.e. coordinates, node labeling, training- and testing trial occupation) are averaged accordingly.

In order to find a good matching of nodes we defined an error function and used the simulated annealing algorithm (SA; Kirkpatrick et al., 1983) to approximate its global minimum. The error function E we aimed to minimize was

$$E = \omega_1 E_1 + \omega_2 E_2 + \omega_3 E_3, \quad (5)$$

where E_1 , E_2 and E_3 denoted distance, connection and correspondence error, respectively (see supplemental material S2). For SA all errors were rescaled to lie between 0 and 1 and the parameters of the error function were $\omega_1 = 0.7$, $\omega_2 = 0.2$, and $\omega_3 = 0.1$. We chose this error weighting to base the matching mostly on the obtained SSOM node weights and less so on node labels. In order to ensure a visually comprehensive connection layout, we introduced the connection error, however, with small weighting. In each step of the SA, we perturbed two randomly selected nodes of one randomly chosen SSOM. A node matching with a lower energy state was always accepted whereas for a higher energy state the new matching was accepted with probability

$p = \exp[(E_{old} - E_{new})/T_{iter}]$, where T_{iter} decreases with iterations ($T_0 = 10$, $T_{iter} = \gamma^{iter-1} \cdot T_0$ and $\gamma = 0.9$ for dataset 1; $T_0 = 1$, $T_{iter} = \gamma^{iter-1} \cdot T_0$ and $\gamma = 0.75$ for dataset 2). The result of this procedure was a node matching between SSOMs that minimized the energy function (Eq. (5)) and a general SSOM could be created. To determine node labels and training/testing trial node occupation we averaged the single SSOM counterparts. This general SSOM can be visualized similarly to single SSOMs (see above).

SSOM visualization

For visualizations of SSOMs, we projected the trained SSOMs onto the first two principal components using customized functions provided by the MATLAB-based SOM toolbox (see 2.1). The resulting grid summarizes the data topology based on categorical structure in the high-dimensional voxel space. In addition, for dataset 2, class-specific maps were derived by computing a weighted average over the (linear) models of all nodes of the considered class. For weighting, we ranked the nodes with respect to their distance from the center of mass of the respective class, i.e. the weighting was computed according to $(\gamma - rank_{disk} + 1)/Y$ where Y denotes the number of nodes of the respective category. Note that the maps in Fig. 12 do not show the average activation for each category but – owing to the SSOM algorithm – they display the set of features (voxels) that discriminate one category optimally from the other categories.

fMRI datasets

Simulated fMRI data

A 3-class fMRI dataset including 3 runs was simulated following procedures described in (De Martino et al., 2008). For each class we simulated 30 trials using SNRs (i.e. signal amplitude/standard deviation of noise) of 0.35 and 0.5 (referred to as *low* and *high* SNR, respectively) and CNRs (amplitude difference between conditions/noise standard deviation) of 0.15, 0.25, and 0.35 (referred to as *low*, *medium* and *high* CNR, respectively). SNR and CNR values were randomly assigned independently to each voxel with a standard deviation of 0.1 (SNR) and 0.01 (CNR). The underlying anatomy and region definitions were based on a real dataset employing auditory stimuli (see 2.6.2). For generating spatio-temporal activation patterns, boxcar time-courses were convolved with hemodynamic responses that varied across voxels with respect to its time to positive peak (drawn from normal distribution with $\mu = 4$ s and $\sigma = 0.5$ s). Subsequently, we added temporally auto-correlated noise to obtain the simulated signal with an autocorrelation of $\rho \sim N(0.5, 0.1)$ for each voxel (cf. De Martino et al., 2008). Finally, voxel time-courses were sampled at a repetition time (TR) of 2.6 s that was also used for acquiring the auditory fMRI dataset (see 2.6.2) resulting in ~180 TRs per run (trials lasted between 5 and 7 TRs). We simulated two regions in the auditory cortex with different response properties. Voxels of the first region (432 voxels) responded to all classes but with larger amplitude to one of the C classes. In the second region (1453 voxels), voxels responded to sounds but did not differentiate between classes. These responsive regions were embedded within a dataset of temporally correlated noise including a total of 16,505 voxels (cortex mask).

In a second set of simulations that we used for estimation of parameters for the dataset 2, specifications of the above-mentioned simulations remained the same but the number of classes was set to 8 (corresponding to the fMRI data) and 12 trials per class were simulated. We defined one region (523 voxels) that differentiated classes with an SNR of 0.5 and CNR of 0.25.

Speaker identification study

To examine the classification approach with a challenging dataset (i.e. difficult to decode), we acquired fMRI data while 5 participants performed a speaker identification task. Vocalized sounds (<1 s) were

played during a 1.2 s period of silence between image acquisitions in a slow event-related design (5–7 TRs between single trials) and subjects were asked to indicate the speaker identity after sound presentation. Functional runs were collected at 3 Tesla (Allegra, Siemens) and consisted of 18 slices positioned parallel to the Sylvian fissure obtained with a T2-weighted gradient echo, EPI sequence (TR 2.6 s, TE 30 ms, TA 2.4 s; voxel size $2 \times 2 \times 2 \text{ mm}^3$). Anatomical images were obtained using a high resolution ($1 \times 1 \times 1 \text{ mm}^3$), T1-weighted MPRAGE sequence. Presented sounds were 30 different short non-linguistic vocalizations (e.g. “aww”, “uuh”) of three speakers (1 female, 2 males referred to as *f*, *m1*, *m2*). Before scanning, participants were familiarized with the voice identities with a short practice session consisting of 20 vocalizations not presented during fMRI data acquisition.

Visual object categories study

We reanalyzed a publicly available data set from the Haxby et al. (2001) study. In this study, participants viewed pictures of faces, cats, houses, chairs, scissors, shoes, bottles, and scrambled version thereof. Provided data were converted to BrainVoyager QX format (v2.4, Brain Innovation). We restricted our analysis to the Region of Interest (ROI) defined by the mask included in the dataset covering the ventral temporal object-selective cortex. Thus, neither univariate nor multivariate voxel selection was necessary. For more specifications on the experiment and data the reader is referred to the original study (Haxby et al., 2001).

Data processing and analysis

Preprocessing

The preprocessing of fMRI data consisted of slice-scan-time correction, motion correction, transformation into Talairach space, temporal high-pass filtering (0.005 Hz) including removal of linear trends and spatial smoothing with a Gaussian kernel (2 mm FWHM) using BrainVoyager QX.

Cross-validation and voxel selection

For cross-validation of the first dataset and corresponding simulations, we divided each run into two half-runs (the first half contained trials 1 to 15 and the second trials 16 to 30 of the respective run; conditions were balanced over half-runs) and performed 6-fold cross-validations to assess classification performance. For model training we defined datasets consisting of 5 of the 6 half-runs. The remaining half-run was used to determine how well the trained model was able to generalize. The second dataset (visual object categories) was analyzed

in a leave-run-out scheme. This resulted in a 12-fold cross-validation for five subjects and 11-fold cross-validation for one subject.

For the first dataset we performed a series of feature selection steps (see Fig. 1A). We limited the analysis on an anatomical mask covering auditory responsive regions of temporal cortex and reduced the number of voxels with a GLM, which was computed using the training set by selecting the strongest responding 2000 voxels across conditions. An ensemble feature selection method (e.g. Abeel et al., 2010) was used to define feature sets with different numbers of voxels. This feature selection approach relies on bootstrapped aggregation of SVM-based feature ranking. In this work, for each of 25 bootstrap samples, features surviving univariate selection were ranked according to model weights of a linear SVM ($C_{svm} = 1$; one-versus-one scheme) and ranks across bootstrap samples were averaged to obtain the final ranking. We created 12 differently sized feature sets by removing iteratively the lowest ranked 20% voxels starting with 1000. This resulted in feature sets of 1000, 800, 640, 512, 410, 328, 262, 210, 168, 134, 107, and 86 voxels.

For both sets of data the single-trial response (in percent signal change) was fitted to a hemodynamic response model for each voxel. The obtained β values indicating response amplitude were used as features representing single trials. We normalized features of both training and test trials with an inter-quartile-range (IQR) normalization (median and 1st and 3rd quartile were estimated using training data): $x_i^{IQR} = 1.35 * (x_i - Q_{50}) / (Q_{75} - Q_{25})$, where x_i is the estimated β value of the i th trial and Q_{50} , Q_{25} and Q_{75} are the median, first and third quartile, respectively. Compared to z-score normalization, using the median and IQR for normalization is more robust to outliers. The scaling factor assures that z-score and IQR normalization is comparable for normally distributed data. For simulations we decoded the three simulated classes and, similarly for the real dataset, we classified the identity of the three speakers (*f*, *m1*, *m2*), whose vocalizations were presented to the subjects. We compared outcomes of SSOMs with a kNN classifier ($k = \{1, 2, 3, 5, 10, 15, 20\}$), a linear SVM ($C_{svm} = 1$), and a non-linear SVM (RBF-kernel; grid search of soft-margin and kernel parameter $C_{svm} [2^{-6}, 2^{-4}, \dots, 2^4]$ and $\sigma_{svm} [2^{-7}, 2^{-5}, \dots, 2^3]$). Parameters for kNN and non-linear SVMs were optimized in 5-fold cross-validation. For SVM classification, the multiclass classification problem was transformed into binary classification using the one-versus-one scheme. We used the Spider toolbox (www.kyb.tuebingen.mpg.de/bs/people/spider) for SVM (linear and non-linear) and the Bioinformatics toolbox for kNN classification (www.mathworks.nl/products/bioinfo/).

Generalization performance was assessed with accuracy (i.e. the number of correct predictions divided by the total number of testing

Fig. 2. Visualization of single-split SSOM. The shading of nodes indicates which of the three classes the node represents. The radius of the inner gray circles denotes the number of training (A) and testing trials (B) for which the node was among the 10-BMUs. The pie charts in (C) show the class membership of testing trials of the respective units (the radii of pie charts denotes as in B the number of testing trials). See text for details.

Fig. 3. Examples of information provided by SSOM topology. SSOMs show the same information as in Fig. 1A but for two special cases. In (A) the simulated 3-class dataset contains two classes that have similar activation patterns. Panel (B) shows a case in which binary classification was carried but the simulated data actually contained three classes.

trials) and true positive rates (TPR) for each class to assess class-specific performance.

Statistical testing

For simulations, we tested whether performances with discriminative voxels were higher than simulations without discriminative voxels using a Monte Carlo approximate permutation test (e.g. Good, 2000). We created 1999 randomly shuffled resamplings of the 20 simulations with and 10 simulations without discriminative voxels. The difference between the two types of simulations of the true groups was compared to the empirical distribution created by the resampling procedure. The significance was calculated by dividing the count of

instances in the permutation distribution that was larger than the actual difference (one count was added to both numerator and denominator). The 95% confidence interval of the estimated p-value is given by $\hat{p} \pm 1.96 * \sqrt{\hat{p}(1-\hat{p})/n}$, where $\hat{p}$ is the estimated p-value and n the number of permutations. Comparisons between classification algorithms were done similarly. However, in contrast to the above-mentioned procedure, we randomly changed the sign of the elements of the difference vector between two classification approaches 1999 times and compared the actual average difference to the differences obtained with the Monte Carlo procedure.

For real data we tested – for each voxel selection level – whether classification performance was better than expected by chance by

Fig. 4. Visualizations of single-split and multi-split SSOMs. Shadings of nodes denote the three associated classes and inner circles of single-split SSOMs reflect the class specificity of testing trials for which nodes were BMUs (diameter scales with class specificity and color denotes the prevalent label of testing trials). Pie charts in the multi-split SSOM show class membership of testing trials and radii denote the number of training trials. Scales of principal components are equal to Fig. 1. The general SSOM is scaled up by a factor of 2.0 compared to single-split SSOMs.

Fig. 5. Classification accuracy of simulated data for different SNR and CNR. Solid lines denote decoding performance for simulations with informative voxels and dotted lines show performance without informative but responsive voxels. Note that results depict an average decoding performance across simulations with independent noise patterns and randomly assigned voxel specificity in the discriminative region. Markers with black edge indicate classification performances above chance ($p < .05$; exact permutation test).

permuting class labels at a single subject level. Due to high computational costs (see Table S.1) the number of permutations was limited to 99. However, we also examined in detail a selected voxel selection level, for which we performed 999 permutations at single subject level. Corresponding single-subject accuracies and their confidence interval were derived as described for the simulations. For group statistics we employed an exact permutation test. Specifically, we determined the difference between the accuracy with true labels and the average permutation accuracies (observed 95% CI of average permutation accuracy for individual subjects $0.33 \pm .098$). The amount of occurrences of larger differences in the permutation distribution divided by the number of permutations denoted the significance. Comparisons between classification algorithms were performed with the same exact permutation test by computing accuracy differences between two classification approaches.

Results and discussion

Simulations

SSOM properties and visualization

Fig. 2 shows a SSOM obtained for one data split (low SNR, medium CNR; 262 voxels). The maps visualize the model as follows. First, weight vectors $\mathbf{m}_i$ of SSOM-units are projected onto two dimensions using PCA. Each node of the SSOM has a colored shading according to the class it represents (determined by the winning index of the class-defining vector $\mathbf{v}_i$). It can be seen that the SSOM grid reflects the properties of the 3-class dataset by forming a triangular-like shape with a clustering of nodes at the tips that represent one class. The shape indicates that, after training of the SSOM algorithm, nodes representing voxel activation patterns of the training set cluster at three regions in the high-dimensional input space. The occurrence of three clusters is expected because input vectors contain class information leading to the class-distinguishing SSOM. To investigate the model's properties, we computed the 10-BMUs for training and testing trials. The radii of the gray circles in Fig. 2A (B) indicate for each node the amount of trials that were best described by the node weight vectors $\mathbf{m}_i$ (i.e. the number of occurrences that this node belonged to the 10 BMUs). It can be seen

that the model succeeds in describing training trials: nodes at the tips of the triangle are more often among the BMUs (Fig. 2A). In contrast, testing trials occupy more often nodes in the center of the triangle revealing that the model cannot differentiate as well between classes for the testing trials compared to trials in the training set (Fig. 2B). To provide a visual account of the generalization performance of the SSOM, pie charts describe to which class the testing trials belong (Fig. 2C). Here, colors indicate the proportion of classes of testing trials whereas the radius reflects the amount of trials falling onto single nodes (same as in Fig. 2B). It can be seen that nodes in the center of the triangle show similar proportions of testing trials for all classes. Nodes at the tips, however, do not attract many testing trials but are more specific with respect to classes. For this model it can be seen that testing trials falling on nodes at tips are most often consistent with the node's class label, thereby suggesting above chance

Fig. 6. Voxel selection of simulated data for different SNR and CNR combinations. Solid lines depict the ratio of informative versus non-informative voxels among selected voxels. Dashed lines indicate the number of selected voxels within the discriminative region divided by total number of informative voxels. Gray lines show the respective results for simulations without informative but responsive voxels.

classification (classification accuracy for this split: 0.40). Note that examining the performance of SSOMs on test trials is essential for a correct interpretation of topologies. In fact, due to supervised nature of the training, a clustering of nodes from the same class occurs to some extent also for random data. However, this is meaningful only if it generalizes to new trials. Hence, visualizations should only be consulted when classification performance is above chance level (similar to MVPA studies presenting discriminative maps). In such cases, the SSOMs visualization can provide interesting insights into the decoding process. For example, when performing 3-class classification and the underlying representation of two classes is similar, SSOMs would show smaller distances for these classes in comparison to the other class (see Fig. 3A for an example). Similarly, strong evidence for separating one class into two classes would lead to subclusters of nodes within a class (Fig. 3B). In addition, other measures besides classification accuracy can be conveyed ‘at a glance’ (e.g. misclassification for specific classes).

To create a general SSOM from single-split SSOMs (Fig. 4 shows projections of SSOMs for different splits) we applied the alignment and node-matching procedure (see 2.3) that assures that class centroids overlap in common space. This linear transformation is then applied to produce a general SSOM. Note that for visualization purposes we did not present the classes of testing trials as pie charts for single-split SSOMs but their class-specificity, which we defined as the ratio of the classification index (cf. Eq. (3)) for the class with most and medium number of trials.

The center of Fig. 4 shows the corresponding general SSOM derived from single-split SSOMs. The nodes that share class labels cluster together forming triangular shapes in the PCA mapping. Similar to properties observed in single SSOMs, nodes between class clusters possess a higher attraction for testing trials, however, with less class-specificity. Nodes at the tips attract fewer testing trials but are more class-specific. It can be seen that for some of these nodes, the majority of testing trials did not belong to the class attributed to the node according to the SSOM algorithm. In those cases this indicates misclassification and, more specifically, which classes are being confused. In the results for the particular simulation presented in Fig. 4, a faithful classification of the yellow class can be observed whereas the purple and orange one are often confused. This is also reflected in TPRs (yellow: 0.60, purple: 0.30, orange: 0.33; classification accuracy: 0.41).

Performance of SSOMs compared to other decoding approaches

For statistical evaluation of performances, we created 20 simulations with different noise patterns and randomized voxel's class preference in the discriminative region (as described in 2.5). In addition, we created 10 simulations in which we replaced voxels in discriminative region with voxels that were still responsive but no

longer discriminative (i.e. these simulations did not contain any information about conditions).

As can be seen in Fig. 5 all decoding algorithms were able to classify class labels successfully independent of SNR for medium and high CNR (summed accuracies across selection levels, $p = .005 \pm .0031$ [significance estimate and 95% confidence interval thereof] for all decoding algorithms) and for high SNR and low CNR ($p < .005 \pm .0031$ for all decoding algorithms). We found that classification accuracy increased with a larger number of features.

Focusing on SSOMs, we found that they perform better compared to k NN classification for simulations with medium CNR and high SNR (summed accuracies across selection levels, $p = .0215 \pm .0064$) and high CNR and low and high SNR ($p = .011 \pm .0046$). Level-specific comparisons showed that SSOMs could better discriminate classes compared to k NN classification for voxel selections consisting of medium-sized feature sets (i.e. 328–640 voxels; $p = .0555 \pm .0100$) for three SNR/CNR combinations (SNR/CNR: low/high, high/medium, and high/high). This suggests that although SSOMs and k NN classification share properties like computing distances to training samples (k NN) and nodes (SSOM) to predict test instances, SSOMs outperform k NN classification. One possible explanation is that in contrast to k NN classification, SSOM abstracts from individual training patterns to form SSOM nodes. This might lead to more robust generalization due to less sensitivity to noise or outlier samples.

Compared to both types of SVMs, the decoding accuracy of SSOMs was not significantly different (summed accuracies across selection levels, $p = .135 \pm .150$ for all SNR and CNR settings). Focusing on performances for single selection levels SVMs outperformed SSOMs only when the number of voxels in the feature set was large (i.e. 800 and 1000 voxels; $p = .045 \pm .0091$) for simulations with three SNR/CNR combinations (low SNR/high CNR, high SNR/medium CNR, and high SNR/high SNR).

To better understand the decoding results, we examined the sets of voxels selected by the ensemble feature selection. As can be seen in Fig. 6, the number of voxels that belonged to the region with discriminative (i.e. informative) voxels increased until the largest set of features (dashed lines). This indicates why decoding algorithms show a general increase in performance. However, the number of informative voxels decreased compared to the number of noise voxels (i.e. voxels that did not contain information about class identity) for increasing sets of features (solid lines). This might explain the performance plateau for SSOMs for high CNR cases as the larger number of noisy voxels entering the set of features counteracts the increase of informative voxels. SVMs are more robust to this increase of noisy voxels by their intrinsic weighting of features, which is in line with the observed performance increase observed also for large sets of voxels. Thus, for cases with a large number of features (e.g. whole-brain classification), SVMs should be chosen over SSOMs.

Fig. 7. Group maps resulting from the ensemble feature selection approach. For each subject the most important 262 voxels were selected and projected onto the group-aligned cortex. The colors depict for how many subjects a region was included in the combined univariate-multivariate selection procedure.

Fig. 8. Average classification accuracy for real fMRI data. Solid lines show average accuracy and dashed lines denote empirical chance level computed by permuting class labels. Markers with black edge indicate significant classification performance on a group level ($p < .05$; exact permutation test).

Speaker identification study

Voxel selection

Fig. 7 shows the group map of the voxel ranking for the left and right hemispheres established by the ensemble feature selection procedure. We used cortex-based alignment (CBA) to transform single-subject selection maps into a common space (Goebel et al., 2006). For the group results, we computed maps that focus on the consistency of important regions across subjects. In particular, we selected the highest ranked 262 voxels for each subject and created a group map thereof that depicts for how many subjects a region was included in the combined univariate-multivariate selection procedure. Voxels selected with high consistency among subjects were found in left planum temporale (PT), bilateral middle superior temporal gyrus (STG)/Heschl's sulcus

(HS), lateral Heschl's gyrus (HG), right posterior superior temporal sulcus (STS), and left anterior and posterior portions of STG. The right STS and HG/HS are in agreement with regions found to be involved in processing of voices (Belin et al., 2000; Moerel et al., 2012) and in the discrimination of individual speakers (Formisano et al., 2008). However, in contrast to Formisano et al. (2008) – where subjects were passively listening to the stimuli – we also observed left anterior STG as being important for speaker classification. This might be due to the active speaker discrimination task employed in our new experiment (see also Andics et al. (2010) that found the same region to be involved in an active speaker identification task).

Performance of SSOMs compared to other decoding approaches

We compared the SSOM decoding performance with those of kNN and SVM (linear and RBF-kernel) classification. Fig. 8 shows the average decoding performance for the different voxel selections and classification algorithms. We found that all algorithms perform above chance level for medium numbers of voxels ($p < .05$).

Compared to kNN classification, SSOMs perform similarly (summed across selection levels, $p = .2903$) but with a better decoding accuracy for medium numbers of voxels. In particular, SSOMs were found to classify better compared to kNN classification for 210, 262 and 512 voxels ($p = .0323$). This finding is in line with the results of the simulations (see 3.1.2). In addition, SSOMs show similar performance to linear and non-linear SVMs for this set of data (summed across selection levels, $p > .40$). The increase of classification performance with larger voxel sets can be observed earlier in SSOMs and reaches its maximum at 262 voxels. In contrast, linear and non-linear SVMs show a more gradual increase and classify most accurately using 640 voxels. For single voxel selection levels, we found that SSOMs performed better compared to linear SVMs for 262 voxels ($p = .0323$) and worse than linear and non-linear SVMs for large number of voxels, i.e. 640 and 800 voxels ($p = .0323$).

In addition to decoding accuracy, we also examined class-specific outcomes. In Fig. 9 (upper left panel) we plotted for SSOMs the average TPR for each of the three voices (see Table 1 for single subject results). We found that the female speaker is decoded more truthfully compared

Fig. 9. Class-specific TPR for decoding algorithms. Colored solid lines denote single speaker TPRs (purple – f, yellow – m1, orange – m2), the black lines classification accuracy and dashed lines indicate TPR for permutations. Markers on the classification accuracy show decoding performance above chance.

Table 1

Individual and group classification performance of SSOM with 262 Voxels. For accuracy bold print denotes significant performance compared to empirical chance level. TPR = true positive rate, ACC = classification accuracy.

Subjects	TPR			ACC
	f	m1	m2	
S1	0.533	0.367	0.367	0.422 (.027)
S2	0.533	0.333	0.400	0.422 (.038)
S3	0.467	0.333	0.467	0.422 (.035)
S4	0.633	0.400	0.333	0.456 (.013)
S5	0.400	0.567	0.233	0.400 (.072)
Average	0.513	0.400	0.360	0.424 (.032)

to male speakers for most voxel selection levels. This is in line with the acoustic properties and corresponding behavioral results showing that the distinction between two male speakers is more challenging compared to detecting the female speaker (results not shown). With respect to TPRs the class-specific performances were found to be similar between SSOMs and SVMs. In contrast, class-specific performances obtained with *k*NN classification show a different pattern. Here, the first male speaker showed the highest TPR for most selection levels.

The decoding results for the real dataset suggest that SSOMs perform similar or slightly better compared to linear and non-linear SVMs for a small and medium number of voxels. For voxel sets with many voxels SVMs could discriminate better between classes probably due to their better robustness to noisy voxels. Compared to *k*NN classification we found that SSOMs had superior classification performance and that the TPRs of SSOMs for single classes matched the behavioral results in contrast to *k*NN classification. Thus, these results suggest that a classification with SSOMs is applicable for smaller regions when inherent multiclass classification is important.

SSOM visualization

Fig. 10A shows the visualizations of general SSOMs (i.e. abstracting from single-split SSOMs) with 262 voxels for all subjects. For most subjects the nodes indicating the female voice attracted the testing trials of the very same class. In contrast, male voices were often confused among

each other. For subject S4 and S5 (Fig. 10A, lower right SSOM) one can see a different pattern revealing faithful classification also for one male voice.

Fig. 10B visualizes the group SSOM resulting from single-subject SSOMs. In the upper panel the average occupation of training trials is shown. It can be seen that in the center of class-specific clusters (i.e. at the tips of the triangle) the training trials belonged to the assigned node label. The further away from the class center a node is situated, the more training trials from the neighboring class cluster fall onto the node. This, as a proof of concept, indicates that the generalization procedure for SSOMs as proposed in Section 2.3 seems to work well when applied twice (first to form the single-subject SSOMs and second to create the group SSOM). Focusing on testing trials (Fig. 10B lower panel) the group SSOM indicated that the dataset was challenging (small class-specificity) but in line with classification accuracies and TPRs showed that all speakers are relatively well separated but especially in the case of one male speakers (m2) test trials were often mislabeled. With respect to SSOM topology, we did not observe differences in class relationships in the projections.

Visual object categories study

Performance of SSOMs compared to other decoding approaches

For the data of the visual object categories study by Haxby et al. (2001), all decoding algorithms provided high classification performances clearly above theoretical chance level of 0.125 (see Fig. 11) using voxels in object-selective ventral temporal cortex. Results showed that SSOMs resulted in higher classification accuracy compared to *k*NN classification. In addition, linear SVMs – showing the highest performance among the tested algorithms – were found to have significantly higher classification accuracy compared to the other classification techniques.

Investigating TPRs for single classes shows similar patterns of decoding performance for all classification algorithms. These show – similar to results of the original paper (see supplemental material of Haxby et al. (2001)) – that faces, houses and scrambled pictures could be identified best, followed by chairs and cats and the lowest decoding accuracies for scissors, bottles and shoes (see Fig. 11).

Fig. 10. Single subject and group SSOMs. Speaker labels for nodes are color-coded (f – purple, m1 – orange, m2 – yellow). Panel (A) shows single-subject SSOMs with class-specificity of testing trials. In (B) the group SSOM with the class-specificity for training (upper panel) and testing trials (lower panel) is presented. All SSOMs depict the decoding model with 262 voxels. Note that for group SSOMs in (B) class-specificity expressed by circle diameter is scaled equally for training and testing trials whereas the scaling of class-specificity changes for single subject SSOMs in (A). Group SSOMs are scaled up by a factor of 1.5 compared to single-subject SSOMs.

Fig. 11. Classification performance for visual object category data. Bars show the TPR for different classes and average classification accuracy. Results for the four classification algorithms employed in this study are presented in different bar colors (blue: SSOM, red: linear SVM, green: non-linear SVM, gray: kNN). Dashed lines denote significant differences between algorithms ($p < .05$; exact permutation test).

SSOM visualization

Fig. 12A shows the group SSOM of the visual object categories study. Apart from the observation that node classes and classes of testing trials were consistent (reflecting the high classification performance for this dataset), the SSOM visualization revealed several features of the dataset. First, scrambled objects formed a tight cluster separated from all other categories. The face and house categories appeared at opposite sides of the SSOM reflecting the highly differential activation patterns for these categories in object-selective ventral temporal cortex. Interestingly, the nodes of the cat category were close to the face category forming a larger cluster probably reflecting an animate/inanimate or a natural/man-made ordering principle for objects (see Kriegeskorte et al., 2008a, 2008b). The categories bottles, shoes and scissors showed overlapping or merging clusters that suggests more similar cortical activation patterns compared to other categories.

Voxel patterns used by the SSOM algorithm to classify new instances are presented in Fig. 12B. These maps show the set of voxel that separate

one category optimally from all others. In particular, the retrieved maps show opposite patterns for faces and houses (most likely reflecting regions responding to faces but not houses and vice versa) and negative weights for scrambled pictures in object-sensitive cortex. In addition, the similarity between maps for cats and faces and for shoes, scissors and bottles is in line with the proximity of these categories of the SSOM. Please note that – similar to other multiclass algorithms – weight maps relative to each binary classification cannot be retrieved with the proposed approach.

Conclusions

In this study, we applied supervised self-organizing maps to decode simulated and real fMRI datasets. The feasibility of SSOMs was shown by means of a simulated set of data. We found that SSOMs performed similarly to SVMs for small and medium numbers of voxels (i.e. 100–600 voxels) and were superior to kNN classification especially

Fig. 12. SSOM and category maps for visual object category data. Panel A shows the Group SSOM for the 8-class dataset projected using the first two principal components. Colors denote the different categories. Category maps of one exemplary subject are presented on axial slices in panel B. These maps are constructed from a weighted average of nodes belonging to the same class and represent voxel patterns that separate one category from the others.

for voxel sets of medium size. For large feature sets (>800 voxels), SVMs outperformed SSOMs. For real data acquired from a challenging voice identification experiment, we showed that SSOMs performed similar to SVMs and kNN classification. We found that, as expected, the female speaker was easier to decode compared to the two male voices using activity patterns from auditory responsive regions of temporal cortex. Results for a dataset with known high classification performance showed that SSOMs were able to classify a dataset with 8 classes. Parameters for the SSOM algorithm were chosen according to results on simulated data; it would be interesting to extend the proposed SSOM method with automated selection procedures, e.g. based on nested cross-validation. Similarly, it would be important to extend the comparison to other classification algorithms used in fMRI data analysis, especially inherently multivariate ones like Gaussian Processes (Rasmussen and Williams, 2006), sparse logistic regression (Ryali et al., 2010; Yamashita et al., 2008), Naïve Bayes (Brouwer and Heeger, 2009; Mitchell et al., 2004), Decision Tree classifiers (e.g. Kuncheva and Rodríguez, 2010; Richiardi et al., 2011), or neural networks (Hanson et al., 2004; Polyn et al., 2005). Making use of the visualization possibilities from SOMs, we suggested one approach to plot the classification model and its generalization performance that provides information about underlying representations. To generalize from single SSOMs to a general SSOM we put forward an approach that first establishes a mapping of SSOMs into a common space and subsequently matches the SSOM-units. Group SSOMs were created and visualized by applying the summarizing approach twice (first, to create single-subject SSOMs from single splits and second, to generalize from single-subject SSOMs to a group SSOMs). SSOM visualizations for the two real datasets highlight the potential to depict the topology of the activation patterns underlying the successful decoding results.

Acknowledgments

This work was supported by Maastricht University (LH, GV, EF) and by the Netherlands Organization for Scientific Research (NWO): VICI Grant 453-12-002 (EF, LH). We thank Federico De Martino for comments on the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2014.02.006>.

References

- Abdi, H., Dunlop, J.P., Williams, L.J., 2009. How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage* 45, 89–95.
- Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., Saeys, Y., 2010. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* 26, 392–398.
- Andics, A., McQueen, J.M., Petersson, K.M., Gál, V., Rudas, G., Vidnyánszky, Z., 2010. Neural mechanisms for voice recognition. *NeuroImage* 52, 1528–1540.
- Beauchamp, M.S., LaConte, S., Yasar, N., 2009. Distributed representation of single touches in somatosensory and visual cortex. *Hum. Brain Mapp.* 30, 3163–3171.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer Verlag.
- Brouwer, G.J., Heeger, D.J., 2009. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29, 13992–14003.
- Cheng, Y., 1997. Convergence and ordering of Kohonen's batch Map. *Neural Comput.* 9, 1667–1676.
- 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, 261–270.
- Crammer, K., Singer, Y., 2002. On the learnability and design of output codes for multiclass problems. *Mach. Learn.* 47, 201–233.
- De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E., 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage* 43, 44–58.
- Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. *Curr. Biol.* 19, 1028–1033.
- Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. “Who” is saying “What?” Brain-based decoding of human voice and speech. *Science* 322, 970–973.
- Friston, K.J., 1995. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.
- Good, P.I., 2000. *Permutation tests*. Springer Verlag.
- Hanson, S.J., Matsuka, T., Haxby, J.V., 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *NeuroImage* 23, 156–166.
- Haufeld, L., Santoro, R., Valente, G., Formisano, E., 2012. Classification and visualization of multiclass fMRI data using supervised self-organizing maps. *Pattern Recognition in Neuroimaging (PRNI), 2012 International Workshop on, IEEE*, pp. 65–68.
- 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.
- Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416.
- Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534.
- Kalteh, A.M., Hjorth, P., Berndtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environ. Model. Softw.* 23, 835–845.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220, 671–680.
- Kohonen, T., 1982. Analysis of a simple self-organizing process. *Biol. Cybern.* 44, 135–140.
- Kohonen, T., 2001. *Self-organizing maps*. Springer Verlag.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3863–3868.
- Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141.
- Kuncheva, L.I., Rodríguez, J.J., 2010. Classifier ensembles for fMRI data analysis: an experiment. *Magn. Reson. Imaging* 28, 583–593.
- LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. *NeuroImage* 26, 317–329.
- Langs, G., Menze, B.H., Lashkari, D., Golland, P., 2011. Detecting stable distributed patterns of brain activation using Gini contrast. *NeuroImage* 56, 497–507.
- Martínez-Ramón, M., Koltchinskii, V., Heileman, G.L., Posse, S., 2006. fMRI pattern classification using neuroanatomically constrained boosting. *NeuroImage* 31, 1129–1141.
- Melssen, W., Wehrens, R., Buydens, L., 2006. Supervised Kohonen networks for classification problems. *Chemom. Intell. Lab. Syst.* 83, 99–113.
- Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S., 2004. Learning to decode cognitive states from brain images. *Mach. Learn.* 57, 145–175.
- Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-A., Morito, Y., Tanabe, H.C., Sadato, N., Kamitani, Y., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60, 915–929.
- Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *J. Neurosci.* 32, 14205–14216.
- Mourão-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage* 28, 980–995.
- Mourão-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., Brammer, M., 2006. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *NeuroImage* 33, 1055–1065.
- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430.
- Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310, 1963–1966.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian processes for machine learning*. MIT Press.
- Reddy, L., Tsuchiya, N., Serre, T., 2010. Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage* 50, 818–825.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Van De Ville, D., 2011. Decoding brain states from fMRI connectivity graphs. *NeuroImage* 56, 616–626.
- Ryali, S., Supekar, K., Abrams, D.A., Menon, V., 2010. Sparse logistic regression for whole-brain classification of fMRI data. *NeuroImage* 51, 752–764.
- Silva, L.A., Del-Moral-Hernandez, E., 2011. A SOM combined with KNN for classification task. Presented at the Neural Networks (IJCNN), The 2011 International Joint Conference on, IEEE, pp. 2368–2373.
- Swisher, J.D., Gatenby, J.C., Gore, J.C., Wolfe, B.A., Moon, C.-H., Kim, S.-G., Tong, F., 2010. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J. Neurosci.* 30, 325–330.
- Vesanto, J., 1999. SOM-based data visualization methods. *Intell. Data Anal.* 3, 111–126.

- Walther, D.B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 29, 10573–10581.
- Weston, J., Watkins, C., 1999. Support vector machines for multi-class pattern recognition. *Proceedings of the Seventh European Symposium On Artificial Neural Networks*.
- Wongravee, K., Lloyd, G.R., Silwood, C.J., Grootveld, M., Brereton, R.G., 2010. Supervised self organizing maps for classification and determination of potentially discriminatory variables: Illustrated by application to nuclear magnetic resonance metabolomic profiling. *Anal. Chem.* 82, 628–638.
- Xiao, Y.-D., Clauset, A., Harris, R., Bayram, E., Santago, P., Schmitt, J.D., 2005. Supervised self-organizing maps in drug discovery. 1. Robust behavior with overdetermined data sets. *J. Chem. Inf. Model.* 45, 1749–1758.
- Xiao, Y.-D., Harris, R., Bayram, E., Santago, P., Schmitt, J.D., 2006. Supervised self-organizing maps in drug discovery. 2. Improvements in descriptor selection and model validation. *J. Chem. Inf. Model.* 46, 137–144.
- Yamashita, O., Sato, M.-A., Yoshioka, T., Tong, F., Kamitani, Y., 2008. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *NeuroImage* 42, 1414–1429.