ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 36 (2007) 88–99

Dynamic discrimination analysis: A spatial–temporal SVM

Janaina Mourão-Miranda,[a,*] Karl J. Friston,[b] and Michael Brammer[a]

[a]*Brain Image Analysis Unit, Biostatics Department, Centre for Neuroimaging Sciences (PO 89), Institute of Psychiatry, KCL, De Crespigny Park, London SE5 8AF, UK*
[b]*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL 12 Queen Square, London WC1N 3BG, UK*

Received 12 September 2006; revised 25 January 2007; accepted 6 February 2007
Available online 23 February 2007

Recently, pattern recognition methods (e.g., support vector machines (SVM)) have been used to analyze fMRI data. In these applications the fMRI scans are treated as spatial patterns and statistical learning methods are used to identify statistical properties of the data that discriminate between brain states (e.g., task 1 vs. task 2) or group of subjects (e.g., patients and controls). We propose an extension of these approaches using temporal embedding. This makes the dynamic aspect of fMRI time series an explicit part of the classification. The proposed pattern recognition approach uses both spatial and temporal information. Temporal embedding was implemented by defining spatiotemporal fMRI observations and applying a support vector machine to these temporally extended observations. This produces a discriminating weight vector encompassing both voxels and time. The resulting vector furnishes discriminating responses, at each voxel without imposing any constraints on their temporal form.

Keywords: Machine learning methods; Support vector machine; Classifiers; Functional magnetic resonance imaging data analysis; Dynamic analysis

Introduction

Neuroimaging techniques have changed the way neuroscientists address questions about functional anatomy. Many questions about brain function, previously investigated using electrophysiological recordings in animals, can now be addressed non-invasively in humans yielding important results in cognitive neuroscience and neuropsychology. During a standard fMRI experiment, hundreds of volumes or scans comprising brain activations at thousands of locations are acquired. Despite the multivariate nature of the data, the most widely used analysis methods employ mass-univariate approaches, which fit a general linear model (GLM) to each voxel's time series, often with the aim of revealing areas differentially activated between two or more brain states (Friston et al., 1995). However, given that most brain functions are distributed processes, involving a network of brain regions, it would seem sensible to use the spatially distributed information contained in the fMRI data to aid our understanding of those functions.

Recently, multivariate pattern recognition methods have been applied to fMRI data (Cox and Savoy, 2003; Carlson et al., 2003; Wang et al., 2003; Mitchell et al., 2004; LaConte et al., 2005; Mourão-Miranda et al., 2005, 2006; Haynes and Rees, 2005; Davatzikos et al., 2005; Kriegeskorte et al., 2006). In these approaches statistical learning methods are used to characterize differences in activation patterns between brain states (e.g., task 1 vs. task 2) or groups of subjects (e.g., patients and controls). This process involves using training data set to find the best spatial patterns that discriminate between two or more experimentally defined brain states. The discriminative performance of these patterns is then evaluated on a test data set.

One assumption made in pattern recognition approaches is that all training examples with the same label (e.g., task 1) have the same properties. For example, in a blocked design, all scans within the task block are considered to belong to the same category and, from a statistical point of view, can be exchanged with each other. However, it is clear that this exchangeability or stationarity assumption could easily be questioned. For example, neuronal adaptation might cause systematic variations in the signal as the block proceeds. This is quite common in learning and other paradigms that induce adaptation or repetition suppression. Even if neuronal activity reaches a steady-state very quickly, the delay and dispersion induced by the hemodynamic response function will cause temporal structure in the observed hemodynamic responses. Clearly, it would be useful to make the temporal patterns in fMRI data an explicit part of the classification. We propose that this can be achieved simply with temporal embedding. The pattern recognition approach would then use both spatial and temporal information. Temporal embedding can be implemented by defining spatiotemporal fMRI observations. Application of support vector machines to these temporally extended observations produces a discriminating weight vector, covering both voxels and time, yielding information about dynamic changes in discriminating regions. The temporal profile of the discriminating vector embodies the task-related responses during the block, without imposing any constraints on their form. A region with only a transient response during the task block will be detected with

* Corresponding author. Fax: +44 20 7919 2116.
E-mail address: Janaina.Mourao-Miranda@iop.kcl.ac.uk (J. Mourão-Miranda).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.02.020

the same efficiency as a sustained response. In contrast, standard GLM analysis or the standard spatial SVMs may not detect transient responses because they are not part of the explicit or implicit statistical model. Critically, the proposed approach does not make any assumptions about the time-dependent expression of neuronal activity or the temporal expression of this activity entailed by a specific form of hemodynamic response function.

Spatiotemporal observations have been used previously as input to classifiers in the context of neuroimaging (Mitchell et al., 2004). The novelty of the present work lies in using the classifier weight vector to infer where (in the brain) and when (in time) the discriminating information occurs.

In this paper, we present an application of spatiotemporal SVM to a block design experiment; we compare the results with those obtained using the standard spatial SVMs. We show that, by using the spatiotemporal approach, one can perform a dynamic discrimination analysis, showing how the regions discriminating between two cognitive states change over time. We will show that spatiotemporal SVM disclosed important transient responses in distributed (prefrontal–amygdala) brain systems that would have been overlooked by conventional GLM or SVM analyses.

Material and methods

Subjects

We used fMRI data from 16 male right-handed healthy US college students (age 20–25). Participants did not have any history of neurological or psychiatric illness. All subjects had normal vision and had given written informed consent to participate in the study after the study was explained to them. The study was performed in accordance with the local Ethics Committee of the University of North Carolina.

Data acquisition

The data for this study were collected at the Magnetic Resonance Imaging Research Center at the University of North Carolina on a 3 T Allegra Head-only MRI system (Siemens, Erlangen, Germany). The fMRI data were acquired using a T2* sequence with 43 axial slices (slice thickness, 3 mm; gap between slices, 0 mm; TR = 3 sec; TE = 30 ms; FA = 80°; FOV = 192 × 192 mm; matrix, 64 × 64; voxel dimensions, 3 × 3 × 3 mm). In each run 254 functional volumes were acquired.

Experimental design

Stimuli were presented in a blocked fashion. There were three different active conditions: viewing unpleasant (dermatological diseases), neutral (people) and pleasant images (pretty women in swimsuits), and a control condition (fixation). Each run comprised six blocks of the active condition (each consisting of 7 images volumes) alternating with fixation control blocks (of 7 images volumes). Blocks of each of the three stimuli classes were presented in random order.

Data representation

In the current approach, each duty cycle (block of each particular type of stimuli and the subsequent control block) is treated as a single spatiotemporal observation (Fig. 1). Thus, the time window runs from the beginning of one block to the beginning of the next. This means the number of observations per subject is the number of blocks. The fMRI data are represented by a spatiotemporal observation of size $M = J_v \times T_t$ where J_v is the total number of voxels and T_t is the number of time points included in the time window. A single feature in one observation is defined by y_{jti}, that is the fMRI data signal of a voxel j at a given time point t in the duty cycle i.

For comparative purposes we define spatial observations using two different approaches. In the first approach we used single fMRI volumes within the block as spatial observations (the number of observations per subject is the number of blocks × number of volumes within the block). In the second approach the spatial observations were created by averaging the fMRI volumes within each active block and subtracting by the average of the fMRI volumes during the preceding and following control blocks (the number of observations per subject corresponds to the total number of blocked presentations in the experiment).

Pre-processing

The data were pre-processed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK). All the scans were realigned to remove residual motion effects and transformed into standard space (Talairach and Tournoux, 1998). The data were smoothed in space using an 8-mm Gaussian filter (FWHM). When block averaging was not used, the baseline and the low frequency components were removed by applying a regression model for each voxel. The low frequency components were modeled by a set of discrete cosine functions (cut-off period 128 s). The removal of low frequency components was not carried out when averaging examples within the block; in this case as we subtracted the mean volume of the previous and subsequent control blocks from the mean volume of the blocks, the temporal compression approach itself incorporates a correction for continuous baseline variations. Finally, a mask was applied to select voxels which contain brain tissue for all subjects. We used singular value decomposition (SVD) to reduce the raw data to its eigen-variates. The SVD was performed across data for all training subjects. The training and the test data were projected onto the resulting singular vectors or basis (a description of SVD can be found in Appendix A). It is important to note that this application of SVD is not a dimension reduction device because we used all of the singular vectors (i.e., principal components); it simply allows the SVM to work with smaller vectors, with no loss of information.

Support vector machine (SVM)

Support vector machines are kernel-based devices that find functions of the data that facilitate classification. They are based in statistical learning theory (Vapnik, 1995) and have emerged as powerful tools for statistical pattern recognition (Boser et al., 1992). In the linear formulation, a SVM finds, during the training phase, the hyperplane that separates the examples in the input space according to a class label. The SVM classifier is trained by providing examples of the form $<\mathbf{x},c>$, where $\mathbf{x}$ represents a spatial pattern and c is the class label. Once the decision function is learned from the training data it can be used to predict the class of a new test example. In the present study, $\mathbf{x}$ represents a spatiotemporal observation (described in Fig. 1 as V_i) and c is the task performed ($c = 1$ for task 1 and $c = -1$ for task 2). A brief summary of the essential concepts of the SVM

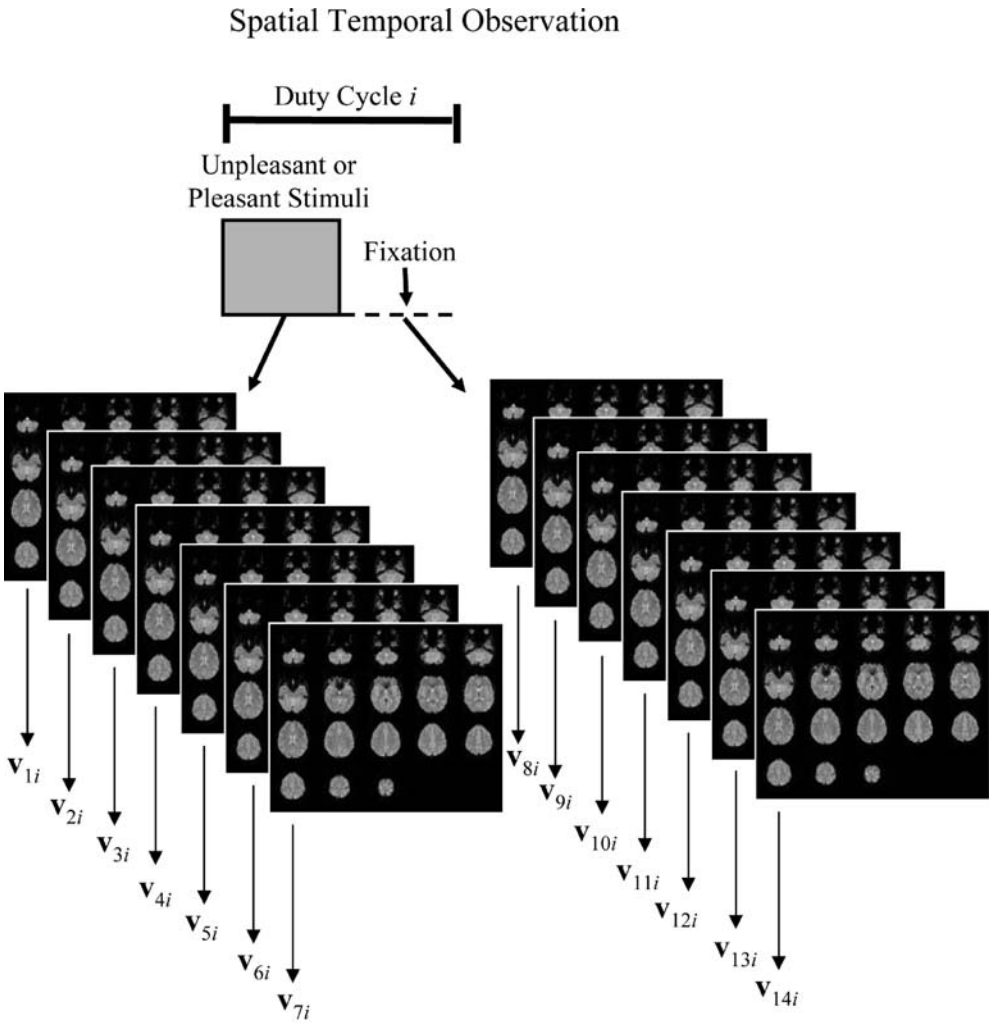

Fig. 1. Representation of a spatiotemporal observation: each fMRI volume is represented by an J_v dimensional vector ($\boldsymbol{v}_{ti}$), where J_v is the number of brain voxels and t refers to the time point in the duty cycle i. By concatenation all volumes within the duty cycle i we define a spatiotemporal observation ($\mathbf{V}_i$).

can be found in Appendix B. For a detailed description see Burges (1998) and Schölkopf and Smola (2002). In our example, we had 3 class labels (unpleasant, neutral and pleasant), which we treated as three two-way discriminations. In the following description, we focus on the unpleasant vs. pleasant comparison.

We used a linear kernel SVM that allows direct extraction of the weight vector as an image (i.e., the discriminating spatiotemporal pattern). A parameter C, that controls the trade-off between having zero training errors and allowing misclassifications, was fixed at $C=1$ for all cases (default value). The SVM toolbox for Matlab was used to perform the classifications (http://ida.first.fraunhofer.de/~anton/index.html).

Dynamic discrimination maps

The separating hyperplane is orthogonal to the direction along which the training examples of both classes differ most, i.e., the weight vector. If the input space is the voxel space (one voxel per

dimension) the weight vector ($\boldsymbol{w}$) will be the direction along which the volumes of either tasks differ most. Hence, it represents a map of the most discriminating regions (i.e., a discrimination map). Given two classes, task 1 and task 2, with the labels +1 and −1, a positive value in the discrimination means that this voxel had higher activity during task 1 than during task 2 in the training examples that contribute most to the overall classification (i.e., the support vectors). In our case the input space covers voxels and time points, consequently $\boldsymbol{w}$ encodes a spatiotemporal map showing, for each voxel, how the discrimination between the tasks changes over time.

Space–time separability

In addition, to evaluate the independence of time and space in the weight vector profile, we performed a principal component analysis of the spatiotemporal weight vector ($\boldsymbol{w}$) by treating it as a space by time matrix. Quantitatively, it might be asked whether the spatiotemporal mode identified by the SVM is dominated by a single mode, suggesting space–time separability, or whether there is spatiotemporal structure that would otherwise have gone undetected. This question is resolved by considering whether the second and subsequent eigenvalues are negligible, in relation to the first. A useful device here is the Kaiser Criterion (Kaiser, 1960) and is probably the one most widely used. This criterion says that we should consider only [normalized] eigenvalues greater than one (one is the average size of the eigenvalues in a full decomposition). In essence, this is like saying that, unless a mode accounts for at least as much variance as an original variable (voxel), we drop it.

Classifier performance

We evaluated the performance of the classifier using a leave-one-subject-out cross validation test. In each trial we used observations from all but one subject ($S-1$ of the S subjects) to train the classifier. Subsequently, the class assignment of the test subject was calculated during the test phase. This procedure was repeated S times, each time leaving observations from different subject out. The classifier accuracy was measured by the proportion of observations correctly classified.

T-maps

For comparative purposes we computed a standard mass-univariate t-test on a time point by time point basis at each voxel, testing for the same effect as the classifier (spatiotemporal SVM).

Results

Classifier accuracy

In Fig. 2 we present the mean accuracy for the SVM trained with spatiotemporal observations (A), with single fMRI volumes (B) and with a mean volume per block (C). Error bars indicate the standard error across 16 leave-one-subject-out cross validation tests (for each test the classifier was trained using data of 15 subjects and tested with a new subject). The mean accuracy obtained by using spatiotemporal observations was 90%. It was higher than the mean accuracy obtained by using single fMRI volumes as training examples (74%) and similar to the accuracy obtained by using the average of volumes as training examples (90%). It is interesting to notice that in cases (A) and (B) the signal to noise rate (SNR) is comparable since there is no averaging of examples in both cases whereas case (C), by virtue of the averaging process involved, produces a higher SNR. The temporal embedding compensates, in terms of classification performance, for the increase of dimensionality and low SNR. Using spatio-

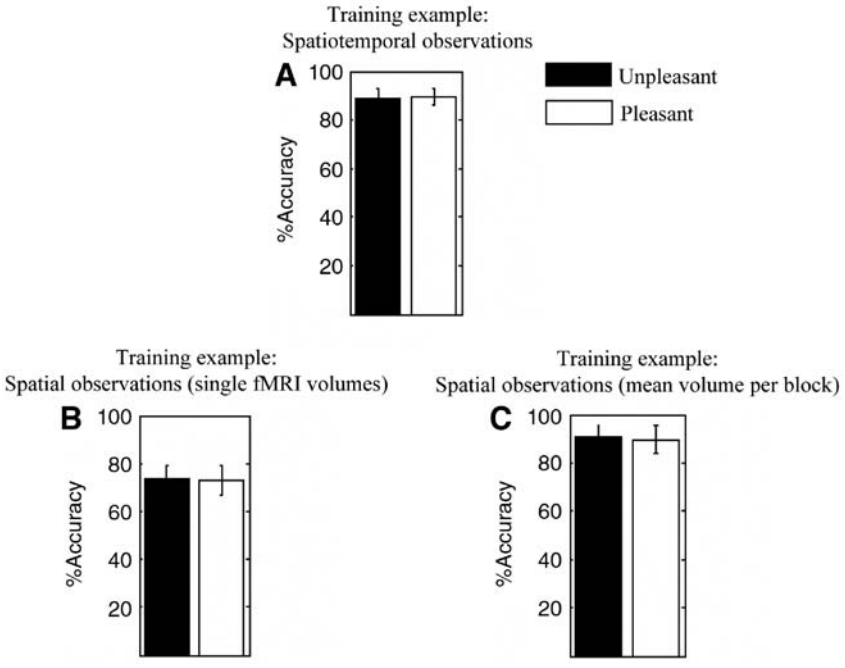

Fig. 2. Classification accuracy of multi-subject SVMs using spatiotemporal observations as input to the classifier (A), using single volumes as input to the classifier (B) and using a mean volume per block as input to the classifier (C). Error bars indicate the standard error across 16 leave-one-subject-out cross validation tests (for each test the classifier was trained using data of 15 subjects and tested with a new subject).

temporal SVM we obtain the same accuracy as that achieved by the spatial SVM, but in contrast to the latter, which involves block averaging, we gain temporal information.

Weight vector—spatial SVM

In Fig. 3 we present the spatial profile of the weight vector for the SVM trained with single fMRI volumes (A) and with a mean volume per block (B). It shows the most discriminating regions between unpleasant and pleasant stimuli. The light/dark blue colors represent areas with high negative values, i.e., relatively more activation for pleasant stimuli, the red/orange colors represent areas with high positive values, i.e., relatively more activation for unpleasant stimuli, and the regions with lower discriminating weight are colored in green, cyan and yellow. The weight vector obtained by training the SVM with single fMRI volumes is noisier than the weight vector obtained by training the SVM with volumes averaged across block although both maps have many peaks in common (presumably representing areas with close to stationary responses). The most discriminating regions, identified by the second approach, are striate visual cortex, which was more activated for unpleasant stimuli and extra-striate visual cortex, which was more activated for pleasant stimuli.

Weight vector—spatiotemporal SVM

In Figs. 4A and B we present the spatiotemporal profile of the weight vector for the SVM trained with the whole block as a spatiotemporal observation. This can be interpreted as a dynamic discrimination map, i.e., for each time point or TR within the duty cycle; it shows the discriminating weight of each voxel. The first seven rows correspond to time points during picture presentation (Fig. 4A: T1–T7) and the following seven rows correspond to time points during fixation (Fig. 4B: T8–T14). The color scale identifies the most discriminating regions for each time point (light/dark blue for negative values, i.e., relatively more activation for pleasant stimuli, and red/orange for positive values, i.e., relatively more activation for unpleasant stimuli) in relation to the regions with lower discriminating weight (green, cyan and yellow). The color scale was adjusted to be comparable with the color scale of the *t*-maps testing for the same effect as the classifier (details described bellow).

The discrimination volumes during the first and second TRs (Fig. 4A, $T=1$ and $T=2$) do not contain many voxels with a highly discriminating weight. During the third TR we observe the first discriminating areas, between unpleasant and pleasant, appearing in red and blue. It is possible to observe how the visual cortex discriminates between unpleasant and pleasant stimuli through time. Its discrimination starts on the third TR, increases continuously until the fifth or sixth TR and decreases after the end of the image presentation. On the other hand, discrimination in the amygdala is very transient, appearing only in the fifth TR. In other regions, like orbito-frontal cortex the discrimination is also more transient than in the visual cortex. It is these dynamic or transient task-related responses that the spatiotemporal SVM was designed to highlight.

To evaluate the temporal pattern of the weight vector and to show how it correlates with the BOLD time series we selected specific voxels based on maxima in the discrimination map and plotted their BOLD time series (averaged over all duty cycles of all subjects) for unpleasant (red lines in Figs. 5A, C, 6A, C and E) and pleasant stimuli (blue lines in Figs. 5A, C, 6A, C and E). In addition, we plotted the temporal profile of the weight vector (plotted in green in Figs. 5B, D, 6B, D and F) and the difference in BOLD response between unpleasant and pleasant stimuli (plotted in blue in Figs. 5B, D, 6B, D and F). The temporal profile of the weight vector for discriminating between unpleasant and pleasant duty cycles is highly correlated with the difference in BOLD responses to unpleasant and pleasant stimulus blocks.

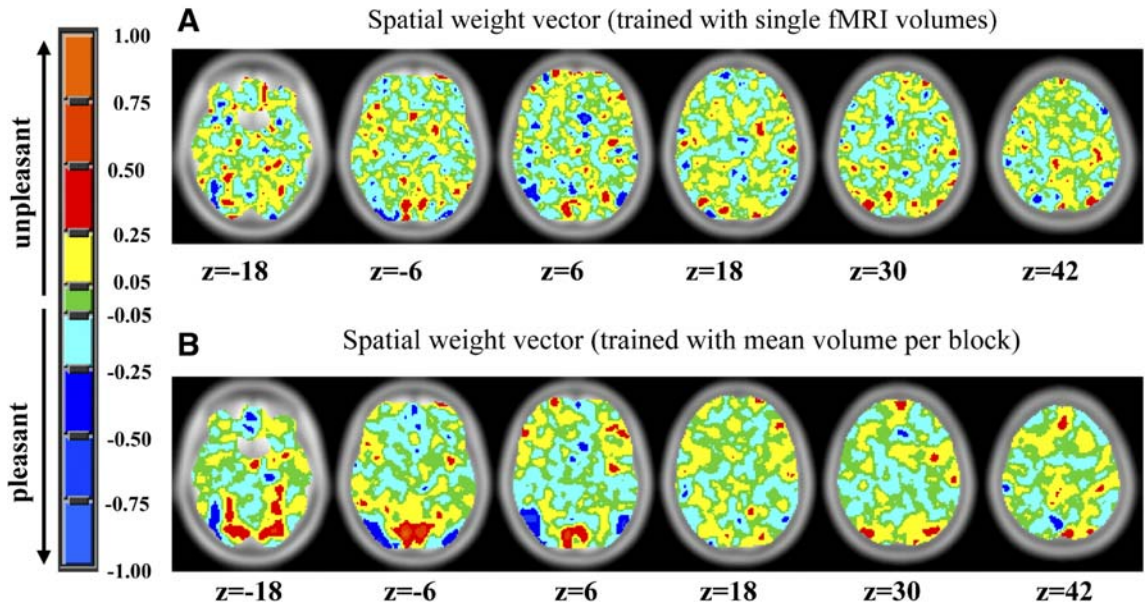

Fig. 3. Spatial profile of the weight vector (trained leaving the subject number 1 out as test data) for discriminating between unpleasant and pleasant stimuli, obtained using single volumes as input to the SVM (A) and using a mean volume per block as input to the SVM (B). The maps were rescaled to have the same mean and variance (the values and colors are equivalent in both maps).

By observing the time courses of the BOLD responses (Figs. 6C and E) we can see that in the visual cortex (striate and extra-striate) they appoximate the shape predicted by convolving the block design with a standard hemodynamic response function (HRF). However, in areas like amygdala (Figs. 5A and C) and orbito-frontal cortex (Fig. 6A) the BOLD response conforms to a different temporal pattern.

T-maps

To compare the weight vector and statistical *t*-maps we rescaled the *t*-maps to have the same mean and variance as the weight vector. By plotting a 2-dimension histogram we observed that for most of the voxels there is a linear relationship between the *t*-values and the

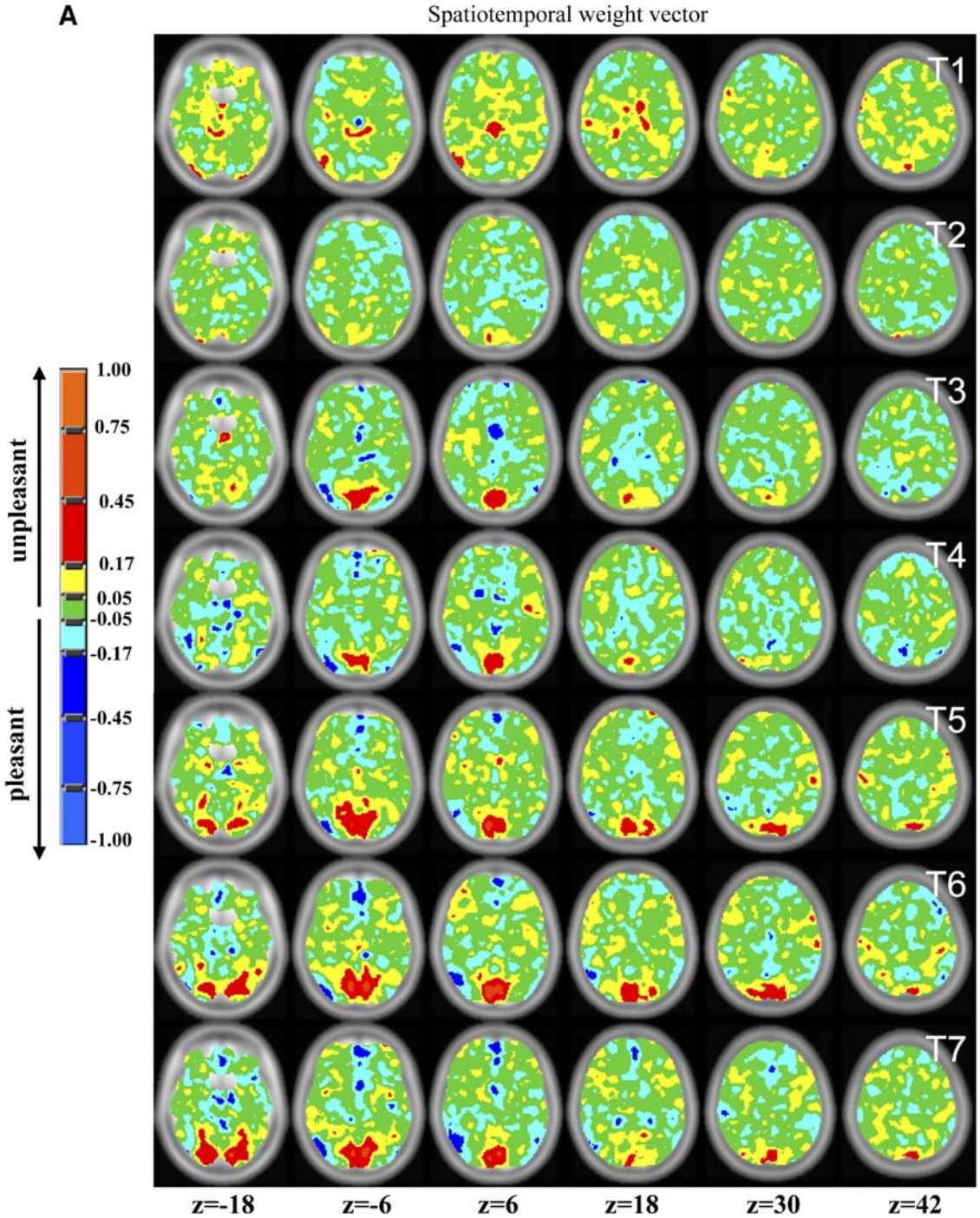

Fig. 4. The spatiotemporal profile of the weight vector (trained leaving the subject number 1 out as test data), for discrimination between unpleasant and pleasant stimuli. The weight vector shows for each time point within the duty cycle the discriminating weight of each voxel. The first seven rows correspond to time points during picture presentation (4A: T1–T7) and the following seven rows correspond to time points during the fixation period (4B: T8–T14). Regions with high values in weight vector are colored in light/dark blue (negative values) and in red/orange (positive values) and regions with low value in the weight vector are colored in green, cyan and yellow.

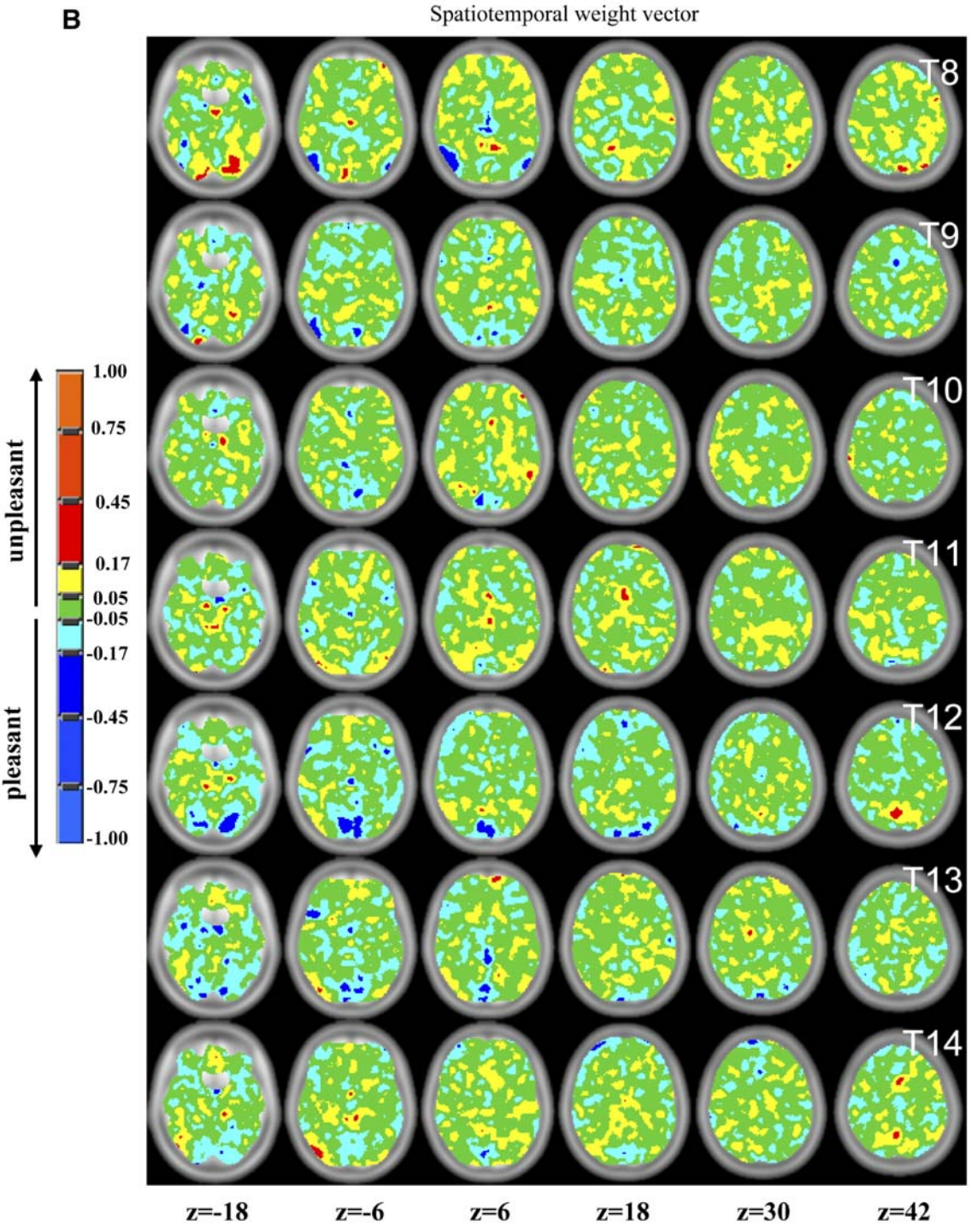

Fig. 4 (*continued*).

values of the weight vector. This means that the values and colors are equivalent in both maps. We adjusted the color scale so that the values corresponding to a p-value < 0.001 in the t-maps were highlighted (i.e., red and dark blue clusters correspond to a p-value < 0.001). There is a high degree of similarity between the t-maps and the weight vector maps, especially in visual areas (see Supplementary material). This result is expected considering the large responses in the visual areas. However, there are some differences between the maps. In general, the t-maps are flatter and there are more localized peaks in the weight vector maps. To illustrate these differences we show two slices at the fifth time point where the amygdale and the orbito-frontal cortex appear as discriminating areas in the weight vector but are less prominent in the t-map (Fig. 7).

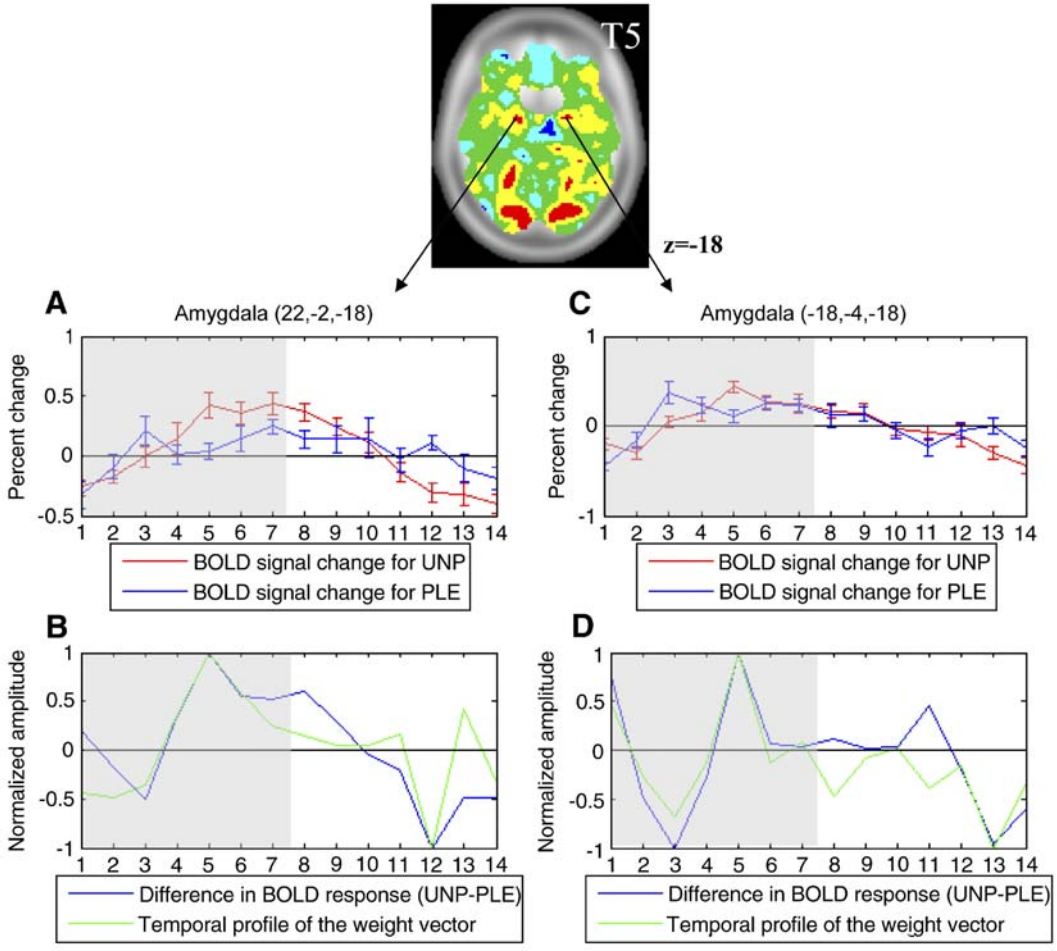

Fig. 5. (A) and (C) BOLD response (% signal change) averaged over all duty cycles of all subjects for unpleasant (red line) and pleasant stimuli (blue line) for a voxel selected in the right and in the left amygdale, respectively. (B and D) The temporal profile of the weight vector is shown in green and the difference in BOLD response between unpleasant and pleasant condition in blue. The area in gray represents the period of stimulus presentation.

Space–time separability

The spectrum of eigenvalues obtained by PCA decomposition of the spatiotemporal weight vector is presented in Fig. 8. The first mode is higher than the others suggesting some degree of space–time separability. However, it is clear that both the second and third modes cannot be discounted and therefore the spatiotemporal structure cannot be factorized.

Discussion

We propose a new approach for performing dynamic discrimination analysis of fMRI data. By defining each duty cycle in a block design as one spatiotemporal observation or training example, we obtain a weight vector that covers both voxels and time. This spatiotemporal discriminator shows, for each time point within the duty cycle, the discriminating weight of each voxel. We applied this approach to a blocked design to discriminate between two brain states resulting from looking at unpleasant (class 1) and pleasant images (class 2). In addition, we compared these results with standard SVMs analysis using just spatial observations as training examples.

Recently, SVMs have been applied to the analysis of fMRI data (Cox and Savoy, 2003; Wang et al., 2003; Mitchell et al., 2004; LaConte et al., 2005; Mourão-Miranda et al., 2005; Davatzikos et al., 2005). There have been a number of different approaches to define the data input to train and test the classifiers. Depending on the information given as input to the classifiers one can use SVM as a pattern recognition approach that looks at different levels or scales of the neuroimaging data. Cox and Savoy (2003) used a number of different approaches to select a subset of voxels and averaged data from individual blocks over a window starting two acquisitions after the beginning of the block and ending at the end of the block. Each averaged block was treated as a single observation for training and testing the classifiers. Wang et al. (2003) trained multi-subject classifiers using two different approaches to define the input vector. In the first approach they used the mean fMRI activity in each of several ROIs defined anatomically in individual subjects as input to multi-subject SVM. In the second approach, they transformed the data to the Talairach space and selected the n most active voxels from across the brain. Additionally, Mitchell et al. (2004) explored a variety of methods for encoding the fMRI data as input to the classifier. They considered feature selection methods that select voxels based on both their ability to distinguish the target classes (discriminability), and on their ability to distinguish the target classes from the fixation condition (activity). In addition, they combined the voxel-selection method with the space-compression

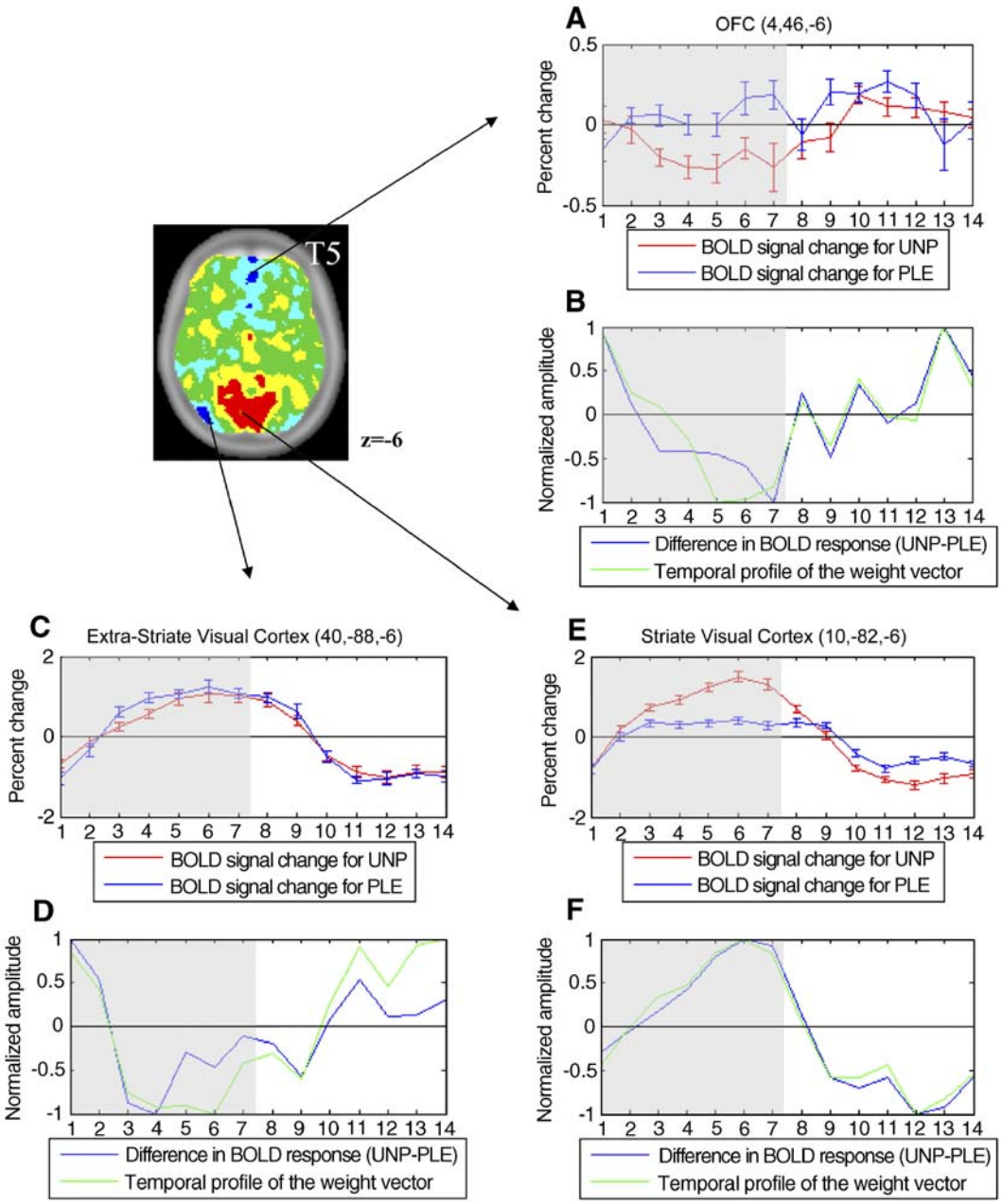

Fig. 6. (A, C and E) BOLD response (% signal change) averaged over all duty cycles of all subjects for unpleasant (red line) and pleasant stimuli (blue line) for voxels selected in the orbito-frontal cortex (OFC), extra-striate visual cortex and striate visual cortex respectively. (B, D and F) The temporal profile of the weight vector is shown in green and the difference in BOLD response between unpleasant and pleasant condition in blue. The area in gray represents the period of stimuli presentation. The agreement between these two profiles is evident.

method by computing the mean of active voxels per ROI. Mourão-Miranda et al. (2005) and LaConte et al. (2005) used individual time points of whole brain as input the classifiers. Davatzikos et al. (2005) used parameter estimate images from the GLM analysis (i.e., regression coefficients) as input to the classifiers. They included each single event during the fMRI experiments in the GLM as an individual regressor.

There are two fundamentally different motivations for performing pattern classification on fMRI data. The first is that classifiers can be used as a "recognition device" to predict cognitive states from brain activity. In these applications the aim is to optimize classification accuracy, without explicit reference to the origin of the discriminating information. However, in the present work the spatiotemporal information was not used to improve the classifier performance but to perform a dynamic discrimination analysis to disclose how regions discriminating between two cognitive states change their behavior over time.

SVM and designed experiments

It is worth considering the motivation for SVM in the context of designed experiments. In this context, SVMs furnish useful

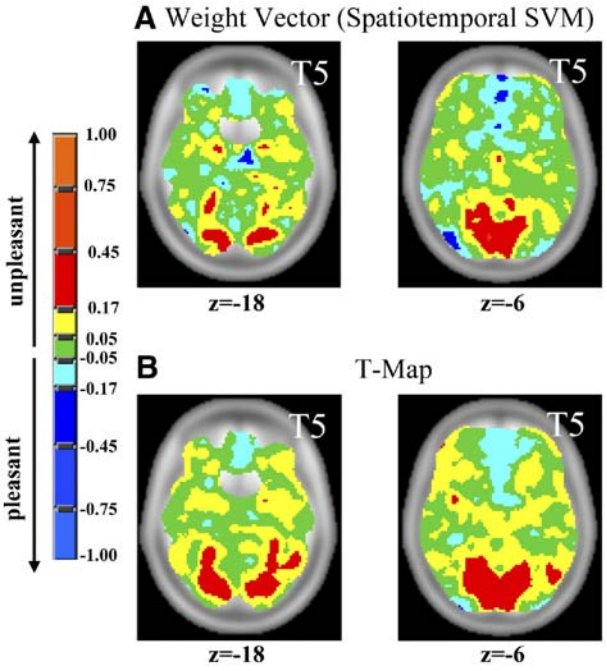

Fig. 7. Figure showing some of the differences between the spatiotemporal weight vector and the paired *t*-test at the fifth time point. Both maps were rescaled to have same mean and variance (the values and colors are equivalent in both maps). We adjusted the color scale to give emphasis to the values corresponding to a p-value<0.001 in the *t*-maps (i.e., red and dark blue clusters correspond to a p-value<0.001).

constraints on the inversion of over-determined models, mapping brain responses to designed experimental factors. These constraints lead to sparsity on the model parameters, where the parameters are forced to be a mixture of a small number of data features (i.e., images). This is useful because the images embody natural scaling and spatial coherence characteristics that one might expect to see in the distributed responses, encoded by the parameters or weights. In our case, we have extended this constraint to spatiotemporal characteristics. Inference about the ensuing parameters is based on classification performance. Crucially, classification performance is used as a surrogate for inference following multivariate model inversion. Classification performance per se is not interesting because one already knows the classes from which the data were sampled (because the experiment was designed). This means that optimizing classification performance is not the key issue when applying SVM to designed fMRI experiments. The key issue is the latitude afforded by SVM to infer plausible and interesting models of structure–function mappings. This is the use of SVM on which we have focused our attention.

Mitchell et al. (2004) trained classifiers to decide whether subject were examining a sentence or a picture during a particular time interval. The input to the classifier was an 8-second interval of fMRI data (4 s of stimulus presentation, followed by 4 s of blank screen). They demonstrated that it is possible to use the information contained in spatiotemporal observations to predict a cognitive state. In contrast we are interested in describing where (in the brain) and when (in time) the discriminating information occurs. This localization in space and time is essential when applying classifiers to designed neuroimaging studies when we are interested in modeling the underlying networks during changes in cognitive state.

Spatial vs. spatiotemporal SVM

In the present work, instead of giving only spatial patterns of brain activity (based on single observations or parameter estimates like time-averages) as input to the SVM, we used spatiotemporal patterns as input to the classifier. The SVM receives not only the spatial information that characterizes the brain state but also the temporal information for each voxel. The output of this approach is the spatiotemporal weight vector which shows for each voxel the discriminating responses during the duty cycle without imposing any constraints to their form. For comparison, we also trained the SVM using spatial patterns of brain activations based on single fMRI volumes and on averages of volumes across time points within blocks. The approach based on single volumes is comparable with the spatiotemporal approach in terms of SNR since there is no averaging of examples in both cases. By averaging the spatial patterns of fMRI activation within the blocks we assume stationarity of the neuronal and hemodynamic responses. This approach has the advantage of increasing the SNR by suppressing the scan to scan variability. However, the drawback is that areas that do not conform to the stationarity assumption will not be detected optimally. Our results showed that the accuracy of the spatiotemporal SVM (90%) was better than the accuracy of the spatial SVM trained with single fMRI volumes (74%) and similar to the accuracy of the spatial SVM trained with averages of fMRI volumes (90%). However, the important difference between the spatial and the spatiotemporal approaches is the weight vector profile. By using spatial patterns as training examples we obtained a weight vector that covers all voxels in the brain, i.e., a map showing the most discriminating regions between unpleasant and pleasant stimuli (Figs. 3A and B). By using spatiotemporal patterns as training examples we obtained a weight vector covering both voxels and time, i.e., a dynamic discrimination map, showing for each time point the most discriminating regions between unpleasant and pleasant stimuli. In Fig. 4 we can see that the most discriminating regions change with time. The visual cortex appears as a discriminating region

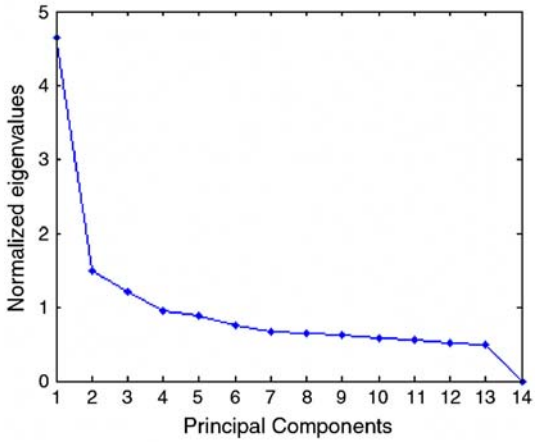

Fig. 8. Normalized eigenvalues of the spatiotemporal weight vector. The principal component analysis was performed by treating the weight vector as a space (lines) by time (columns) matrix.

during most of the time during which stimuli are presented (T1–T7). However, other discriminating regions have a high value of the weight vector for a much more limited period, e.g., amygdala and orbito-frontal cortex. When we examine the time series of discriminating regions we can see that in the visual areas (Figs. 6C and E) they resemble the prediction of the block design. However, in other regions such as the amygdala (Figs. 5A and C) and orbito-frontal cortex (Fig. 6A) the temporal profile of the BOLD response is very different from that predicted by a simple box-car model. The results shown in Fig. 5 have a high degree of biological validity because we would expect to amygdala to respond to emotionally salient stimuli and, furthermore, be involved in time-dependent learning of the current experiential context. For example, Büchel et al. (1999) found that responses in the amygdala were best characterized with a time × stimulus interaction, indicating rapid adaptation of responses to salient stimuli in a trace conditioning paradigm. The important issue here is that these time-dependent prefrontal–amygdala responses were not anticipated in the current paradigm and would not have been detected by conventional GLM or SVM analyses. The spatiotemporal SVM identified these responses by relaxing inappropriate constraints on their form. One could now re-visit this paradigm using conventional analyses, armed with new knowledge about the functional anatomy of processing salient stimuli that has been furnished by our novel pattern recognition scheme.

The spatiotemporal SVM approach can be used to investigate the sequence of cognitive states that the subjects pass through when performing a complex task. One does not need to define a model including each component of the task as an individual regressor. The spatiotemporal profile of the weight vector reveals the regions involved in performance of the task. Therefore, it can be considered a data-driven method. However, there is a fundamental difference between this approach and other model-free and unsupervised approaches like ICA or clustering. SVM is a supervised method and information about the model is defined by the labels of the training examples. The requirement for using this approach is to have enough replications of a particular task or cognitive state (without overlapping events) to give statistical reliability.

Space–time separability

A fundamental aspect of spatiotemporal SVMs is that they do not assume that spatial and temporal responses factorize. Spatiotemporal factorization is an assumption implicit in conventional unsupervised approaches like independent component analysis (ICA). Unlike these approaches, we can model responses that are truly distributed in space and time; e.g., we can test for regional activations that appear in a specific temporal order, or at different times (e.g., the amygdala in our example). To demonstrate this point we performed a principal component analysis of the spatiotemporal discriminator (SVM weight vector) above by treating it as a space by time matrix. If this analysis had yielded a single significant component, the conclusion would be that the space and time components did indeed factorize. However, three significant components were detected, though the magnitude of the first was considerably greater than the second and third. From this result, we conclude that joint spatiotemporal analysis yields significant extra information over the separate spatial and temporal analysis. Indeed, a number of studies have assessed formally the issue of separability (Eckert et al., 1992; Depireux et al., 2001). The current study used a blocked experimental design and it might also be the case that the advantage of joint spatiotemporal analysis might increase with event-related designs, where dynamic changes may be more evident.

SVM vs. t-test

The comparison between the spatiotemporal weight vector maps and the *t*-maps (Supplementary material) showed that, although there are similarities between these maps, there are more localized peaks in the weight vector maps, e.g., in the fifth time point the amygdalae and the orbito-frontal cortex appear as discriminating areas in the weight vector but are less prominent in the *t*-map (Fig. 7). It is important to appreciate that performing *t*-tests on a time point by time point basis would not have revealed the amygdala involvement afforded by SVM. After correction for multiple comparisons (over space and time), the mass-univariate approach would yield non-significant results. Conversely, the multivariate approach entailed by SVM and its classification performance allow us to infer that the amygdala is a key component of a significant spatiotemporal mode, elicited by the experimental factor of interest.

In summary, we have shown that by using temporal embedding it is possible to make the temporal aspect of the fMRI data an explicit part of the classification problem. The temporal embedding was implemented by using spatiotemporal observations as training examples for the SVM. By using this approach it is possible to see dynamic changes in the brain during the performance of a task or a cognitive state that could not be detected easily by the standard spatial SVM analysis or univariate GLM approaches.

Acknowledgments

KJF was funded by the Wellcome Trust.

JMM would like to acknowledge the valuable assistance given by Joao Sato from the Department of Statistics, Institute of Mathematical Statistics, University of Sao-Paulo, Brasil for the many useful discussion during the preparation of this paper. JMM and MB thank Unilever plc (UK) for financial support for part of this project and Dr. Francis McGlone for his valuable input.

Appendix A. Singular Value Decomposition (SVD)

We define a data matrix including all training data, $\mathbf{D}_{M\times N}$ with an observation (i.e., volume image) per column and one feature per row $\mathbf{D}=[\mathbf{V}_1 \ldots \mathbf{V}_i \ldots \mathbf{V}_N]$ and $\mathbf{D}_c$ being $\mathbf{D}$ with the average over all volumes subtracted from each column. We computed the SVD of $\mathbf{D}_c$:

$$\mathbf{D}_c = \mathbf{USW} \tag{A.1}$$

The projection of the volumes onto their principal components is:

$$\mathbf{D}^p = \mathbf{U}^T\mathbf{D} \tag{A.2}$$

where $\mathbf{U}$ is an $M\times N$ matrix containing one eigenvector or PC per column, the superscript p indicates projected data or matrix and the superscript T denotes transpose. We used $\mathbf{D}^p$ as training data, and the test data were given by

$$\mathbf{D}^p_{\text{test}} = \mathbf{U}^T\mathbf{D}_{\text{test}} \tag{A.3}$$

where $\mathbf{D}_{\text{test}}$ is the data matrix with an observation per column (with the average over all volumes subtracted from each column).

In short, initially the data are in voxel space (one voxel per dimension) and after the projection the data are in the principal component (PC) space. The classification problem is solved in the PC space and the result (i.e., the weight vector) is mapped back to voxel space.

Appendix B. Support vector machine (SVM)

In the linear SVM formulation for the binary classification ($f(\boldsymbol{v})=\pm 1$) the learning function corresponds to a hyperplane ($\mathbf{H}$) that separates the examples in the input space according to the class label (i.e., an interface between the two classes). The hyperplane is described by:

$$\mathbf{H} : (\boldsymbol{w}^p)^T \boldsymbol{v}^p + b = 0 \tag{B.1}$$

where $\boldsymbol{w}^p$ is a learning weight vector T denotes transpose, b is an offset and $\boldsymbol{v}^p$ is a vector in the input space. The superscript p means projected vector or matrix onto some basis $\boldsymbol{w}=\mathbf{U}\boldsymbol{w}^p$ for computational expediency (see Appendix A).

The solution $\boldsymbol{w}^p$ is constructed by solving a constrained quadratic optimization problem and it has an expansion in terms of a subset of training examples that lie on the margin (support vectors), given by

$$\boldsymbol{w}^p = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{v}_i^p \tag{B.2}$$

where N is the number of training examples. The training examples $\boldsymbol{v}_i^p$ with non-zero coefficients α_i are called support vectors and carry all information relevant about the classification problem. The learning weight vector is the direction along which the classes differ most and is orthogonal to the hyperplane. When considering spatiotemporal observations the weight vector has a spatiotemporal information.

After the training one needs to map back the weight vector to the voxel space to relate it to spatial locations in the original fMRI data. The weight vector in the voxel space is given by:

$$\boldsymbol{w} = \boldsymbol{U}\boldsymbol{w}^p \tag{B.3}$$

$$\boldsymbol{w} = [\boldsymbol{w}_1 \ldots \boldsymbol{w}_k] \tag{B.4}$$

Where $\boldsymbol{w}_1$ is the discriminating volume during the time point 1 and k is the number of time points within the duty cycle. The maps corresponding to $\boldsymbol{w}_1$ to $\boldsymbol{w}_k$ are presented in Figs. 4A and B.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.02.020.

References

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. D. Proc. Fifth Ann. Workshop on Computational Learning Theory. ACM, pp. 144–152.

Büchel, C., Dolan, R.J., Armony, J.L., Friston, K.J., 1999. Amygdala–hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. J. Neurosci. 19 (24), 10869–10876.

Burges, C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2 (2), 121–167.

Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15 (5), 704–717.

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.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage 28 (3), 663–668.

Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 85 (3), 1220–1234.

Eckert, M.P., Buchsbaum, G., Watson, A.B., 1992. Separability of spatiotemporal spectra of image sequences. IEEE Trans. Pattern Anal. Mach. Intell. 14 (12), 1210–1213.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 20, 141–151.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. 103 (10), 3863–3868.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. NeuroImage 26 (2), 317–329.

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.

Mourão-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, S., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. NeuroImage 28 (4), 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 (4), 1055–1065.

Schölkopf, B., Smola, A., 2002. Learning with Kernels. MIT Press.

Talairach, P., Tournoux, J., 1988. A Stereotactic Coplanar Atlas of the Human Brain. Thieme, Stuttgart.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Wang, X., Hutchinson, R., Mitchell, T.M., 2003. Training fMRI classifiers to detect cognitive states across multiple human subjects. Proceedings of the 2003 Conference on Neural Information Processing Systems, Vancouver.