ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 1055 – 1065

The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data

Janaina Mourão-Miranda,[a,*] Emanuelle Reynaud,[b] Francis McGlone,[c,d] Gemma Calvert,[e] and Michael Brammer[a]

[a]*Biostatics Department, Centre for Neuroimaging Sciences, Institute of Psychiatry, KCL, London, UK*
[b]*Institut de Psychologie, Université Lumière Lyon 2, Lyon, France*
[c]*Unilever R&D, Cheshire, UK*
[d]*Sir Peter Mansfield MR Centre, Nottingham University, UK*
[e]*Department of Psychology, University of Bath, Bath, UK*

Received 28 March 2006; revised 25 July 2006; accepted 9 August 2006
Available online 28 September 2006

In the present study, we compared the effects of temporal compression (averaging across multiple scans) and space selection (i.e. selection of "regions of interest" from the whole brain) on single-subject and multi-subject classification of fMRI data using the support vector machine (SVM). Our aim was to investigate various data transformations that could be applied before training the SVM to retain task discriminatory variance while suppressing irrelevant components of variance. The data were acquired during a blocked experiment design: viewing unpleasant (Class 1), neutral (Class 2) and pleasant pictures (Class 3). In the multi-subject level analysis, we used a "leave-one-subject-out" approach, i.e. in each iteration, we trained the SVM using data from all but one subject and tested its performance in predicting the class label of the this last subject's data. In the single-subject level analysis, we used a "leave-one-block-out" approach, i.e. for each subject, we selected randomly one block per condition to be the test block and trained the SVM using data from the remaining blocks. Our results showed that in a single-subject level both temporal compression and space selection improved the SVM accuracy. However, in a multi-subject level, the temporal compression improved the performance of the SVM, but the space selection had no effect on the classification accuracy.

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

Introduction

In recent years, fMRI has become one of the most widely used non-invasive methods for the investigation of human brain function in vivo. Most investigations focus on mapping areas in the brain activated in response to a specific experimental task. In standard experimental designs, the subjects perform active and control tasks during the scanning section, and statistical analysis methods are applied in order to identify the areas differentially activated between the two types of task. To answer this question, univariate statistical methods are commonly applied to the time series at each voxel (Friston et al., 1995a). These methods often ignore the spatial correlation in the data, i.e. the interrelationship between the activities at different brain areas. The standard univariate method is to fit a general linear model (GLM) to each voxel's time series. Considering that most brain functions are distributed processes involving a network of brain regions, it would seem to be desirable to use the spatially distributed information contained in the fMRI data to give a better understanding of brain functions. Some multivariate techniques have been also applied to fMRI data (e.g. Friston et al., 1995b; McIntosh et al., 1996; McKeown et al., 1998), but such techniques are used much less frequently than univariate approaches.

Recently, pattern recognition methods have begun to be applied as multivariate techniques for fMRI data analysis (Cox and Savoy, 2003; Carlson et al., 2003; Wang et al., 2003; Mitchell et al., 2004; LaConte et al., 2005; Mourao-Miranda et al., 2005; Haynes and Rees, 2005; Davatzikos et al., 2005; Kriegeskorte et al., 2006). In contrast with the standard approaches that try to map cognitive tasks to brain regions, pattern recognition approaches allow the mapping from a pattern of brain activity (i.e. observed fMRI data) to a subject's cognitive states, in other words a "brain reading" approach. The fMRI data are treated as a spatial pattern, and statistical pattern recognition methods (e.g. machine learning algorithms) are used to obtain the mappings. Such algorithms are designed to learn and later predict or classify multivariate data

* Corresponding author. Brain Image Analysis Unit, Centre for Neuroimaging Sciences (PO 89), Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK. 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 © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.08.016

based on statistical properties of the data set. A promising machine learning approach is the support vector machine (SVM). SVM is a kernel-based device designed to find functions of the data that enable classification. These techniques are based in statistical learning theory (Vapnik, 1995) and have emerged as powerful tools for statistical pattern recognition (Boser et al., 1992). The SVM analysis consists of two phases: training and testing. During the training phase, the algorithm finds statistical properties in the fMRI training data that discriminate between two or more brain states. After the training, during the test phase, the algorithm can predict or classify the brain state of a test data (i.e. a new example or subject).

There have been a number of studies using SVM to classify fMRI data. Some studies have used whole brain data as input to an SVM classifier (Mourao-Miranda et al., 2005; LaConte et al., 2005), others have applied a feature selection method to choose regions of interest (ROIs) and then used the time series from the ROIs as input to the classifier. This strategy has been used both at single-subject (Cox and Savoy, 2003 and Mitchell et al., 2004) and multi-subject (Wang et al., 2003) level. An intermediate approach was used by Kriegeskorte et al. (2006). They proposed scanning the whole brain with a "searchlight" whose contents were analyzed multivariately at each location in the brain. It is important to emphasize that there are two fundamentally different motivations to perform pattern classification on fMRI data. The first is that pattern classification can be used as to infer cognitive or mental states from whole brain activity in a single or multi-subject level. In these applications, one can use the whole brain as input to the classifier and the SVM algorithm finds, by itself, the most discriminating regions common to subjects, and this information can be presented as a map showing the most discriminating regions between two different brain states. However, other studies use pattern classification to compare the role of different cortical areas in information processing. In these cases, the use of ROIs is not performed to achieve good classification but in order to make regionally specific assignment of decoding accuracy. The problem in selecting ROIs is that the selection depends on a user-driven choice, usually some *a priori* hypothesis. In addition, considering the anatomical variability among different subjects, it seems difficult to choose common ROIs for different subjects at a multi-subject level.

In this paper, we considered various data reduction or transformations that can be applied before training the support vector machines and we have evaluated their effect on the accuracy of classifiers trained on single and multi-subjects. The transformations we considered involve averaging or selecting data in time or space. Effectively, this restricts the kernel of the SVM to subspaces of the original data. By choosing the subspace carefully, it is possible to improve the classification performance. A good subspace will be one that retains discriminatory variance, while avoiding subspaces that are irrelevant for classification. One obvious benefit of averaging is that classification-irrelevant noise is suppressed. By averaging over-scans in fMRI, one suppresses scan-to-scan noise, increasing the signal to noise rate (SNR). By selecting data in space (i.e. defining "regions of interest" in the whole brain), one can focus on a subset of brain regions evaluating the role of these regions in performing different brain functions.

It is important to highlight that the optimal data transformation to retain the task discriminatory variance may be different in a single and multi-subject context. In a multi-subject context, inter-subject anatomical variability is often addressed using spatial smoothing and the analysis depends on coarse-grained regional effect as opposed to the fine-grained local pattern that is important for single-subject analysis. To address this issue, we compared the effects of temporal compression and space selection using classifiers trained on single and multi-subjects. We used different approaches to perform temporal compression and space selection. The first approach for temporal compression was averaging the time points within each block, minus the average of the time points within the preceding and following control blocks. In the second approach for temporal compression, we used parameter estimates from a linear model of voxel-specific time series as inputs to the SVM. The "regions of interest" were defined by using two different strategies: The first one was based on the results of the GLM (univariate approach) to select the most activated voxels, and the second was based on results of a first-pass SVM (multivariate approach) to select the most discriminating voxels.

Materials 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 gave 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 runs were acquired using a T2* sequence with 43 axial slices (slice thickness, 3 mm; gap between slices, 0 mm; TR = 3 s; 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 girls in swimsuits), and a control condition (fixation). In each run, there were 6 blocks of the active condition (each consisting of 7 images volumes) alternating with control blocks (fixation) of 7 images volumes. Six blocks of each of the 3 stimuli were presented in random order.

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, transformed into standard space (Talairach and Tournoux, 1988). The data were smoothed in space using an 8-mm Gaussian filter (FWHM) for the multi-subject classifiers. However, to preserve the fine-grained local pattern in the single-subject classifiers, we used unsmoothed data.

In addition, 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 128s). The removal of low frequency components was not necessary in the temporal compression approach. As we subtracted the mean volume of the previous and posterior control blocks from the mean volume of the blocks, the temporal compression approach itself incorporates a correction for continuous baseline variability. Finally, a mask was applied to select voxels which contain brain tissue for all subjects.

We used singular value decomposition (SVD) or principal components analysis (PCA) to reduce the raw data to its eigenvariates (i.e. reducing the size of the problem from the number of voxels to the number of spatial modes, i.e. scans). In the multi-subject analysis, the SVD was performed across data for all training subjects. In the single-subject analysis, the SVD was performed across data from all training blocks. In both cases, the training and the test data were projected onto the basis computed using the training data. The SVMs were trained and tested with the projected data. A description of SVD can be found in Appendix A.

Classifier

The support vector machine (SVM)

Machine learning methods can be used to find the unknown decision function f that maps fMRI data to brain states: $f(\mathbf{v}) \rightarrow c$, where $\mathbf{v}$ represents the fMRI data (i.e. one volume image) and c is the brain state or class. The classifier is trained by providing examples of the form $<\mathbf{v}, c>$ (e.g. training data consisting of fMRI volumes and the known class label for each volume). Once the decision function is learned from the training data, it can be used to predict the brain state of a new set of data (e.g. fMRI volume from a new subject). There are different approaches to determine the decision function depending on the learning method used. It is important to have a decision function that not only classifies the training data correctly, but also does the same for the test data (new examples), i.e. one needs to find a classifier which is able to generalize well. It is known that the SVM algorithm corresponds to an optimal classifier with the best generalization yet described (Boser et al., 1992). In Appendix B, there is a brief description of the SVM, for a detailed description of the SVM, see Burges (1998).

Multi-class SVM

The classification problem can be extended to cases where the number of classes is more than two, in this case, the unknown function f takes values from a discrete set of classes: $\{c_1, \ldots c_k\}$. Typically, the multiclass problem is reduced to multiple binary classification problems that can be solved separately. There are many ways to achieve this (Allwein et al., 2000). For all approaches, after the binary classification problems have been solved, the resulting set of binary classifiers must be combined in some way to give the class prediction. In the simplest approach, each class is compared to all others, one versus all. In a second approach, all pairs of classes are compared to each other. Dietterich and Bakiri (1995) presented a general framework in which the classes are partitioned into opposing subsets using Error Correcting Output Code (ECOC). In the ECOC approach, each class is associated with a unique binary string of length n ("codewords"). Then, n binary classifiers are trained, one for each bit position of the binary string. During the test, a new example is classified by evaluating each of the binary functions to generate an n-bit string. This string is compared to each of the k codewords, and the test example is assigned to the class whose codeword is close, according to some distance measure, to the generated string.

In the present work we used the ECOC approach, which in the case of three classes corresponds to an all pairs approach. The ECOC is defined by a code matrix of binary values (Fig. 1). The number of columns is the length of the code (number of bits), and the number of rows is equal to the number of classes in the

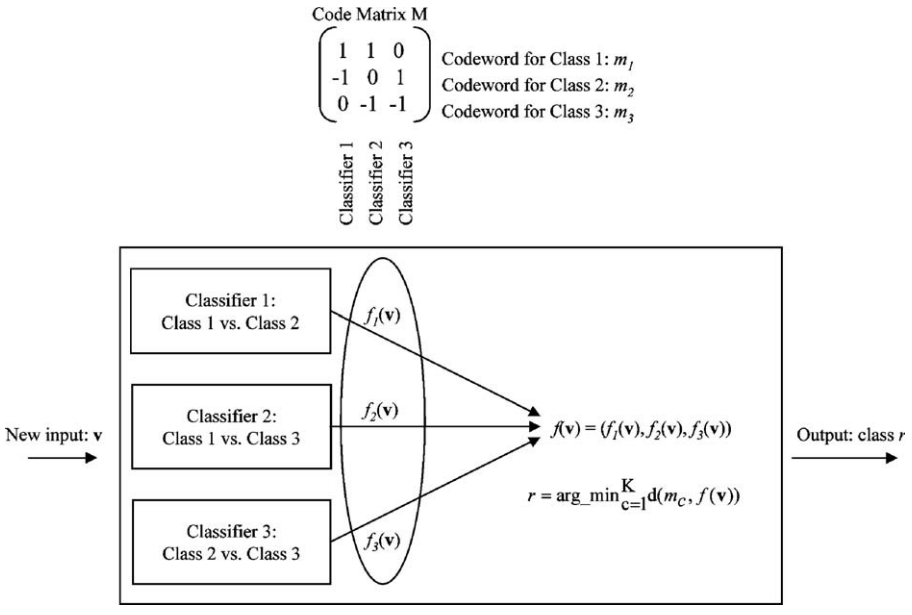

Fig. 1. Three-class SVM and the Error Correcting Output Coding (ECOC) method. The three-class SVM is reduced to three binary SVM classifiers. The prediction or classification of a new example $\mathbf{v}$ is made by evaluating each of the 3 classifiers to generate an output of 3 bits ($f(\mathbf{v})=(f_1(\mathbf{v}), f_2(\mathbf{v}), f_3(\mathbf{v}))$. The predicted class r for the example $\mathbf{v}$ will be the one whose corresponding codeword m_r is the closest to $f(\mathbf{v})$ according to the Hamming distance.

multiclass learning problem. As shown in Fig. 1, each column of the code matrix represents one classifier (column 1: classifier between Class 1 and Class 2, column 2: classifier between Class 1 and Class 3 and column 3: classifier between Class 2 and Class 3). Each class is represented by a codeword (first line corresponds to Class 1 and so on). The prediction or classification of a new example $\mathbf{v}$ is made by evaluating each of the 3 classifiers to generate an output of 3 bits, $f(\mathbf{v})=(f_1(\mathbf{v}), f_2(\mathbf{v}), f_3(\mathbf{v}))$, where $f_1(\mathbf{v})$ is the output of the classifier 1, $f_2(\mathbf{v})$ is the output of the classifier 2 and $f_3(\mathbf{v})$ is the output of the classifier 3. The predicted class for the example $\mathbf{v}$ will be the one whose corresponding codeword is the closest to $f(\mathbf{v})$ according to the Hamming distance (i.e. the number of bits which differ between two binary strings) (Hamming, 1950). The classification can be seen as a decoding operation and the class of an input $\mathbf{v}$ is computed as:

$$\arg\_\min_{c=1}^{K} d(m_c, f(\mathbf{v})) \quad (1)$$

where m_c is the codeword for the class c and $\arg\_\min_{x\in S} d(x)$ returns one of such x that minimizes the function d.

In case of fMRI data, each linear SVM classifier is described by a weight vector which corresponds to a volume with the most discriminating regions between two classes. In case of three classes, we will obtain one discriminating volume for each classifier: Class 1 vs. Class 2, Class 1 vs. Class 3 and Class 2 vs. Class 3.

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

Analysis

In the present study, we investigated the effect of temporal compression and space selection on SVM classification in a multi-subject and single-subject level. We used the "leave-one-subject-out" cross-validation approach for the multi-subject classifiers, in each iteration, we trained the SVM using data from all but one subject (i.e. the SVM received all data from all training subjects with no explicit distinction between within and between subject effects) and tested its performance in predicting the class label of the test subject's data. For the single-subject classifiers, we used the "leave-one-block-out" approach, i.e. for each subject, we selected randomly one block per condition to be the test block and trained the SVM using data from the remaining blocks. We constrained the block selection in a way that all blocks but the first and last one were chosen the same number of times as test blocks.

We trained and test SVM classifiers using combination of the following approaches:

Whole data

In this approach the whole data were used as input to the classifier, after the pre-processing steps described in Section 2.4.

Temporal compression I

The temporal compressed data were created by averaging the time points within each active block and subtracting the average of the time points of the preceding and following control blocks. This corresponds to fitting a top-hat function to the voxel time series and using the associated parameter estimate as the input to the SVM.

Temporal compression II

In this approach, we used parameter estimates from a linear model of voxel-specific time series as inputs to the SVM. To account for the delay and dispersion induced by the hemodynamic response, we used a block-specific estimator based upon convolving a top-hat function with a hemodynamic response function. This procedure was applied to each block within a particular class providing one parameter estimate per block.

Space selection I (ROIs based on the GLM)

In this approach, a subset of voxels were selected within a predefined mask. For each "leave-one-subject-out" test, a mask was created based on a GLM analysis of the training subjects (mixed effects multi-subject analysis). For the "leave-one-block-out" approach, a mask was created based on a GLM analysis of the training blocks (fixed effect single-subject analysis). In both cases, the GLM mask included all of the voxels activated (p-value<0.001 uncorrected) in at least one of the SPMt contrasts: condition 1 vs. condition 2, condition 1 vs. condition 3 and condition 2 vs. condition 3.

Space restriction II (ROIs based on SVM)

In this approach, we used the training data to build a first-pass SVM classifier (a three-class SVM consists of a classifier between conditions 1 and 2, a classifier between conditions 1 and 3 and a classifier between conditions 2 and 3). Each binary classifier corresponds to a weight vector. For each classifier, we ranked the voxels according to their values of the weight vector and labeled the highest h voxels as discriminating voxels, where h was the number of voxels activated in the GLM analysis for the same contrast and using the same training subjects. The SVM mask included all of the voxels labeled as discriminating in at least one of the classifiers.

In the multi-subject classifiers, we tested the effect of temporal compressions I and II and combinations of temporal compression I with space selection I and II. In the single-subject analysis, we only tested the effects of the temporal compression I and space selection I on the SVM accuracy.

Classifier performance

The classifier accuracy was measured by the ratio of the number of examples correctly classified to the total of tested examples. Differences between classification accuracy for different levels of temporal compression and space selection were determined using a generalized linear model with categorical explanatory variables (analogous to ANOVA) and suitable error structure (binomial) to reflect the nature of the data (percentage scores). These analyses were performed in R (http://cran.r-project.org/).

GLM analysis

We used SPM2 (Wellcome Department of Imaging Neuroscience, London, UK) to perform the GLM analysis. The effects of the experimental conditions on the response variable were modeled as box car functions smoothed with a hemodynamic impulse response and used as regressors in the GLM (Friston et al.,

1995a,b). The statistical model included global and low frequency confounds. A random or mixed effect analysis was performed in SPM2 as a second level analysis.

Results

The results comparing the effects of temporal compression and space selection on multi-subject and single-subject classifiers are presented in Figs. 2–4 and Tables 1 and 2. In Figs. 2–4, a summary of the analysis steps and the accuracy results for the different cases are presented. All approaches led to classification accuracy significantly above chance. If the three-class SVM were operating in a random fashion, we would expect a third of the stimuli to be correctly classified, i.e. the expected accuracy would be 33.33%.

Multi-subject classification

In Fig. 2 and Table 1, we can see the effect of temporal compression on the accuracy of the multi-subject SVM classifier. We compared the performance of three different approaches: (A) whole brain with no compression (i.e. using each fMRI single volume as one example), (B) average of the time points within each block, minus the average of the time points within the preceding and following control blocks (temporal compression I) and (C) temporal compression based on GLM parameters (temporal compression II). It is possible to see that both temporal compression approaches lead to an improvement on the classification accuracy. We applied the generalized linear model using a temporal compression as a categorical explanatory variable (no compression, temporal compression I and temporal compression II). There was a significant effect of temporal compression on the SVM accuracy, however, there was no significant difference between the accuracy resulting from temporal compression I and temporal compression II.

In Fig. 3 and Table 1, we present the results of the SVM using combinations of temporal compression I and two different approaches for space selection: voxel selection based on the GLM results (space selection I) and voxel selection based on the results of a first-pass SVM (space selection II). Statistical analysis based on the generalized linear model including spatial selection and temporal compression as categorical explanatory variables showed that there was a main effect of temporal compression, no main effect of space selection and no interaction. That means space selection did not change the SVM accuracy in a multi-subject level, however, the temporal compression significantly improved the SVM performance.

Single-subject classification

In Fig. 4 and Table 2, we present the results of the single-subject SVM classifiers using combinations of temporal compression I and space selection I. The mean accuracy over all classes was 69.34%

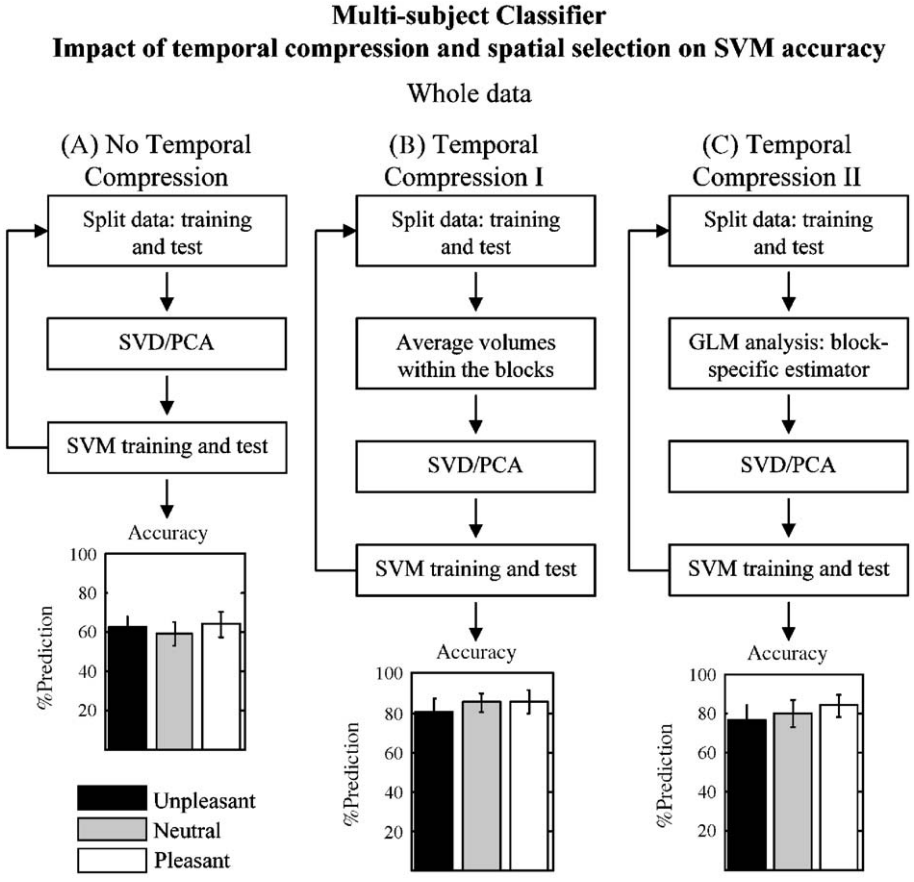

Fig. 2. Classification accuracy of multi-subject SVMs using different levels of temporal compression: (A) no temporal compression, (B) temporal compression I and (C) temporal compression II. In each case, a summary of the analysis step is presented. Error bars indicate the standard error across 16 leave-one-subject-out cross-validation tests (for each test, the classifier was trained using data of 15 subjects and tested with a new subject).

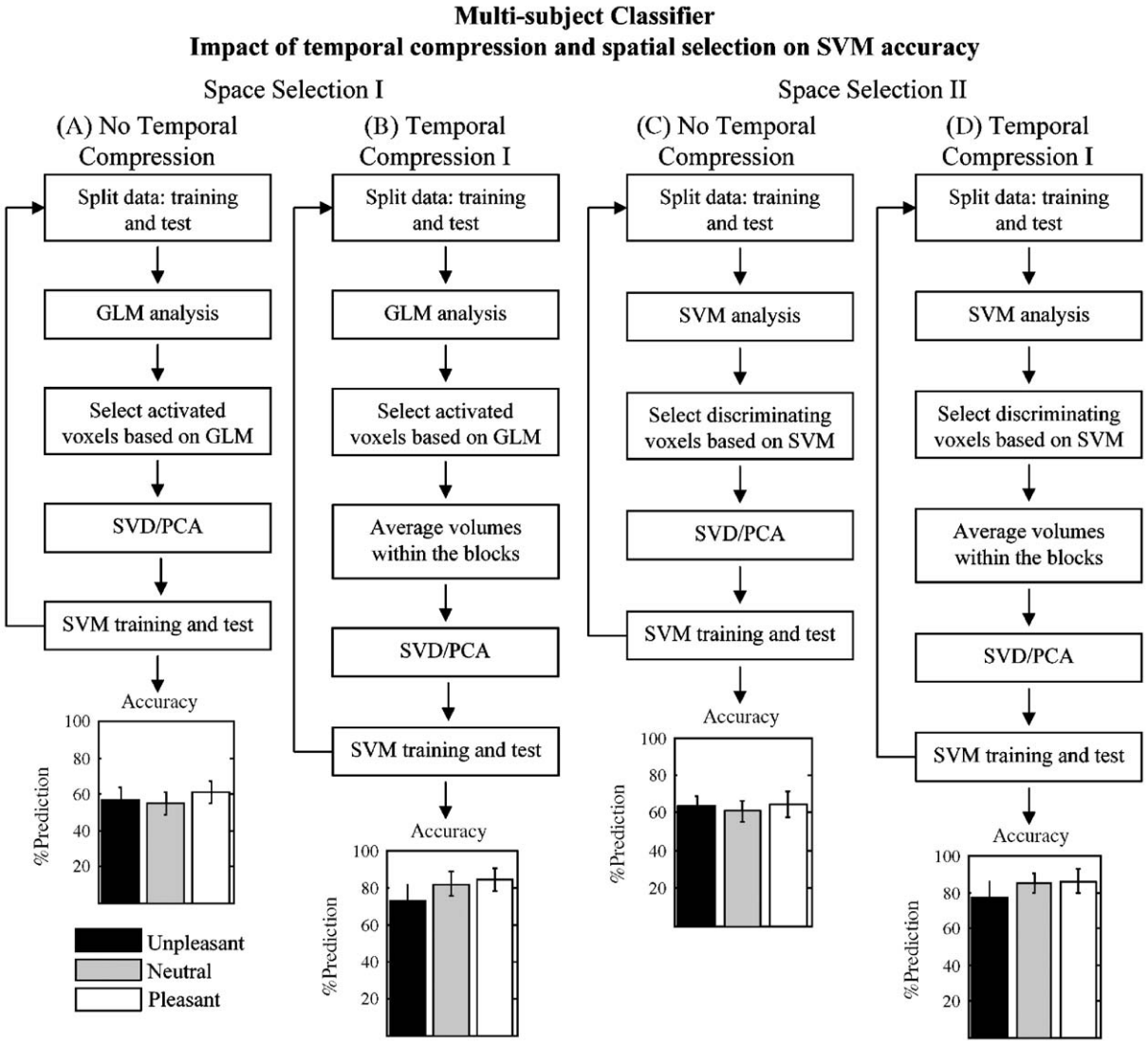

Fig. 3. Classification accuracy of multi-subject SVMs using different levels of temporal compression and space selection: (A) space selection I and no temporal compression, (B) space selection I and temporal compression I, (C) space selection II and no temporal compression and (D) space selection II and temporal compression I. In each case, a summary of the analysis step is presented. 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).

using whole brain, 75% using temporal compression, 81.84% using space selection and 85.41% using a combination of temporal compression and space selection. Even though both approaches, temporal compression and space selection, lead to an improvement of the mean SVM accuracy, these effects were not significant in the generalized linear model including a temporal compression and space selection as categorical explanatory variables. The possible reason for this result is the very low number of training and test examples used in the single subjects classifiers. As we applied the "leave-one-block-out" cross-validation test in a data set with only six blocks, in the temporal compression approach, there were only five training examples and one test example for each class. Considering this limitation, we performed a second analysis including only space selection as a categorical explanatory variable in the model (whole brain and space selected data). By using this model, we observe a significant effect of space selection on the accuracy of single-subject's SVM classifier.

SPMt and discriminating maps

As described in the Materials and methods section, the three-class SVM corresponds to three binary classifiers. Each classifier corresponds to a discriminating volume, where the value of each voxel indicates the importance of such voxel in differentiating between two brain states or stimuli. For each contrast, we present the results of the SPM*t* (multi-subject analysis) together with the discriminating volume (multi-subject classifiers) when using the whole data and the temporal compressed data (temporal compression I). This allows a comparison between the classical GLM results and the SVM results. Figs. 5–7 represent the pair wise comparisons (Fig. 5 unpleasant vs. neutral, Fig. 6 unpleasant vs. pleasant and Fig. 7 neutral vs. pleasant). In each case, panel A shows the result of the GLM analysis and panels B and C show the SVM results using the whole data and temporal compressed data, respectively. Voxels with p-value<0.001 (uncorrected) in the SPM*t* are shown in color scale (blue scale for

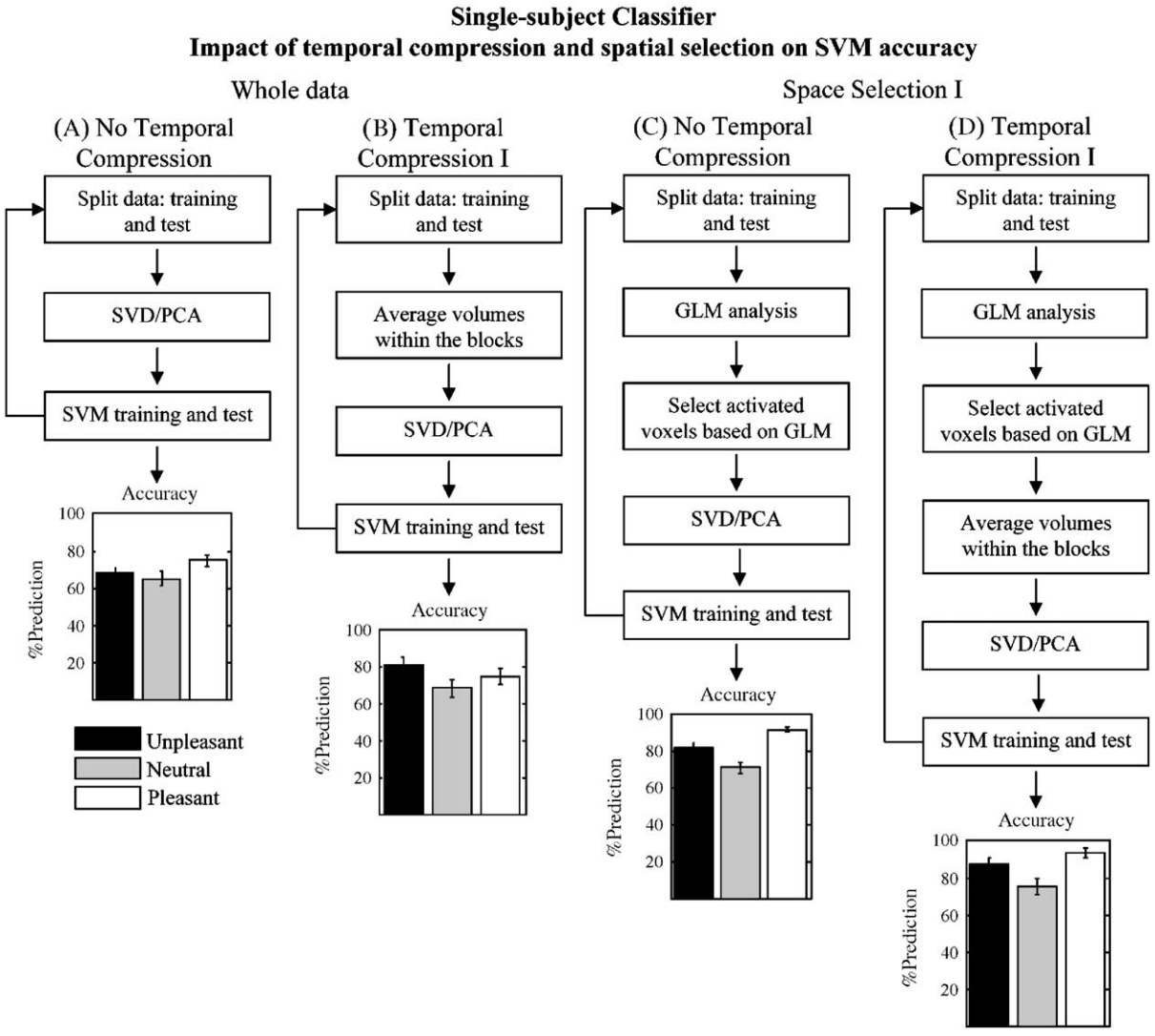

Fig. 4. Classification accuracy of single-subject SVMs using different levels of temporal compression and space selection: (A) whole brain and no temporal compression, (B) whole brain and temporal compression I, (C) space selection I and no temporal compression and (D) space selection I and temporal compression I. In each case, a summary of the analysis steps is presented. Error bars indicate the standard error across 16 single-subject SVMs.

negative values and red scale for positive values). The discriminating volumes were thresholded in a value that revealed the same number of voxels as in the thresholded SPM*t* of equivalent effect.

These results show that the discriminating volumes obtained by using the temporal compressed version of the data (Figs. 5C, 6C and 7C) tend to be closely related to SPM*t* (Figs. 5A, 6A and 7A). The discriminating volume obtained by using the whole (Figs. 5B, 6B and 7B) seems to be influenced more by noise in the data.

Table 1
Multi-subject classifiers—mean accuracy for the three-class SVM obtained using combinations of temporal compression and space selection approaches

Method	Mean accuracy
Whole brain (no temporal compression)	62.00%
Temporal compression I	83.68%
Temporal compression II	80.55%
Space selection I	57.59%
Space selection II	62.85%
Space selection I and temporal compression I	79.86%
Space selection II and temporal compression I	82.99%

Discussion

In the present study, we investigated the impact of temporal compression (averaging across multiple scans) and space selection

Table 2
Single-subject classifiers—mean accuracy for the three-class SVM obtained using combinations of temporal compression and space selection approaches

Method	Mean accuracy
Whole brain	69.34%
Temporal compression I	75.00%
Space selection I	81.84%
Space selection I and temporal compression I	85.41%

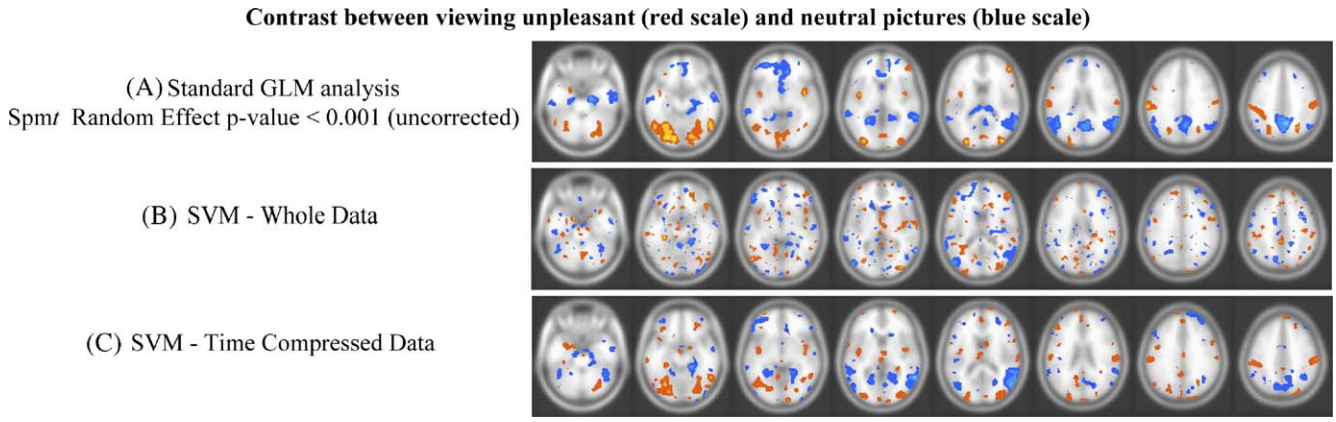

Fig. 5. (A) SPM*t* showing the contrast between viewing unpleasant and neutral pictures and the SVM discriminating volume for the same contrast (B) using the whole data and (C) using the temporally compressed data. The discriminating volumes were thresholded to give the same number of voxels as in the thresholded SPM*t* with *p*-value < 0.001 (uncorrected). In both cases, we used a blue scale for negative values and a red scale for positive values.

(i.e. selection of "regions of interest" from the whole brain) on classification performance using SVM trained on single- and multi-subject fMRI data. Our results showed that in a multi-subject level the temporal compression improves the performance of the SVM, but the space selection has no effect on the classification accuracy. However, in a single-subject level, both temporal compression and space selection may improve the SVM accuracy.

Previous studies have used spatially selected data to train the SVM (Cox and Savoy, 2003; Wang et al., 2003; Mitchell et al., 2004). Cox and Savoy (2003) trained single subject's classifiers to distinguish between 10 classes using data from predefined regions of interest during a subset of trials (i.e. voxel time series). The ROIs were defined by using two different methods: voxels that vary significantly across at least one of the categories of stimuli according to ANOVA and voxels inside "object processing areas" according to a univariate correlation analysis. Wang et al. (2003) trained multi-subject classifiers using two different feature selection approaches to reduce the dimensionality of the input feature 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 feature or voxel selection 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 method by computing the mean of active voxels per ROI. The space selection approach may be efficient for training single-subject classifiers, but, by virtue of cross-subject anatomical variability, it may not be a good feature selection method for training multi-subject classifiers.

In contrast, other studies used whole brain data as input to the classifier without prior selection of spatial features (Mourao-Miranda et al., 2005; LaConte et al., 2005). Mourao-Miranda et al. (2005) applied SVM to perform multi-subject classification of brain states from single fMRI volumes. They demonstrated that SVM outperforms Fisher Linear Discriminant (FLD) in classification performance as well as in robustness of the spatial maps

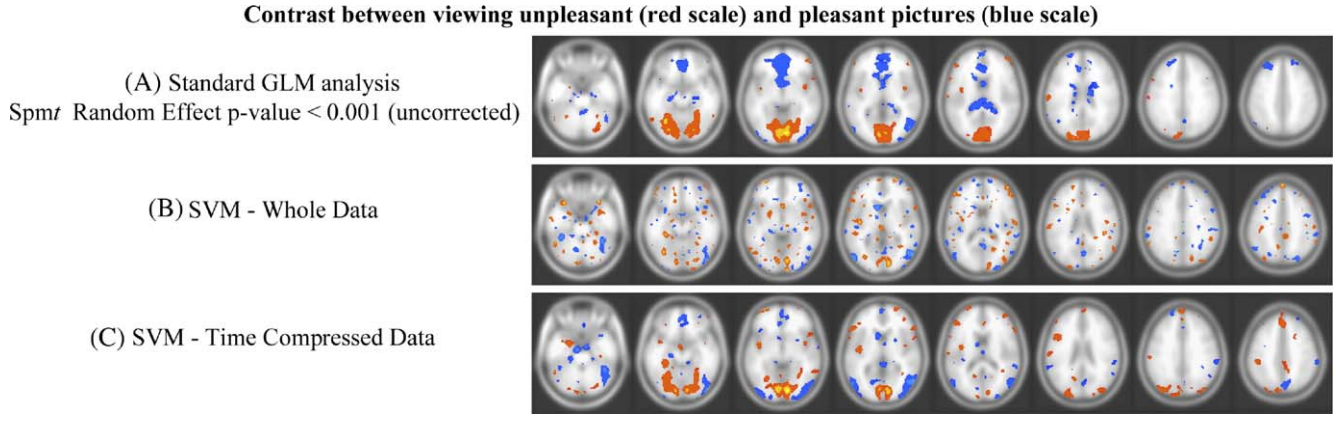

Fig. 6. (A) SPM*t* showing the contrast between viewing unpleasant and pleasant pictures and the SVM discriminating volume for the same contrast (B) using the whole data and (C) using the temporally compressed data. The discriminating volumes were thresholded at a value that gave the same number of superthreshold voxels as in the thresholded SPM*t* with *p*-value < 0.001 (uncorrected). In both cases, we used a blue scale for negative values and a red scale for positive values.

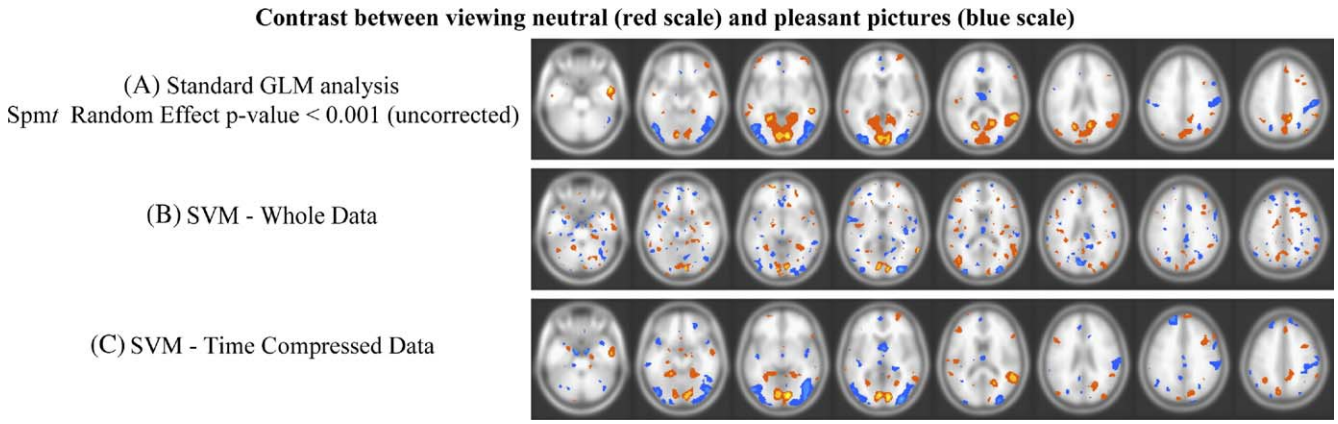

Fig. 7. (A) SPM*t* showing the contrast between viewing neutral and pleasant pictures and the SVM discriminating volume for the same contrast (B) using the whole data and (C) using the temporally compressed data. The discriminating volumes were thresholded to give the same number of voxels as in the thresholded SPM*t* with *p*-value < 0.001 (uncorrected). In both cases, we used a blue scale for negative values and a red scale for positive values.

obtained and showed that the SVM discrimination maps had greater overlap with the GLM analysis compared to the FLD. LaConte et al. (2005) evaluated the impact of pre-processing choices (spatial smoothing, temporal detrending and motion correction) on the SVM performance of single-subject classification and compared the SVM results with the Canonical Variate Analysis (CVA) for each pre-processing choice.

Another interesting approach was proposed by Kriegeskorte et al. (2006). They scanned the imaged volume with a spherical "searchlight" whose contents were analyzed multivariately at each location in the brain. The "searchlight" was centered on each voxel in turn. The resulting map showed for each voxel in the volume how well the multivariate signal in the local spherical neighborhood differentiates the experimental conditions.

The contribution of the present work was to investigate various data transformation that can be applied before training the SVM to retain task discriminatory variance, while suppressing irrelevant components of variance. We also compared the effect of these transformations on classifiers trained on single and multi-subject data. In the single-subject level, the analysis can be focused on fine local patterns using unsmoothed data. On the other side, in a multi-subject level, the analysis depends more on large regions effects on spatially smoothed data.

The first possible approach is to use the whole data as input to the classifier, i.e. to train the classifier using individual volumes as input (including all the voxels that contain brain tissue). This approach is an exploratory analysis of the data. The classifier will find the important features to distinguish the classes in common across all subjects (i.e. the discriminative regions). The disadvantage of this approach is the high computational cost, and in addition the classifier may use noise information present in the data to discriminate between the classes. To increase the signal to noise ratio (SNR), instead of using individual volumes to train the classifier, one can consider the fMRI signal stationary within the blocks and use a representative volume per block (e.g. the average of the volumes in the active block minus the average of the previous and posterior control blocks). In this case, the scan-to-scan variance within the control and "active" blocks is suppressed and the information given to the classifier is more likely to be task related. The classifier is still able to explore the whole brain space and find the most discriminating regions as there was no space selection. An extension of the temporal compression approach is to replace the block averages with a block-specific estimator based upon convolving a top-hat function with a hemodynamic response function. This properly accounts for the delay and dispersion in the response. By using the parameter estimates from a convolution model (GLM), we are effectively deconvolving the hemodynamic response from the signal. This represents a good way forward, however if the model underlying the GLM is an inaccurate reflection of the underlying response, the subsequent SVM analysis would be inaccurately biased.

A third possibility is to use spatially selected data as input to the classifier, i.e. choose some regions of interest based on predefined criteria. The advantage of this method is that it excludes from the analysis any regions that are clearly not relevant for the performance of the task or experimental condition. On the other hand, it is not easy to define *a priori* which regions are relevant. Considering the inter-subject variability, this is especially difficult when training the classifier on multi-subject data. One way to select the regions is to apply a univariate analysis on the training data to find regions with activation task related. In this case, the sensitivity of the classifier will be limited by the sensitivity of the univariate method. Another option is to use a first-pass SVM as a multivariate method to select the regions of interest. An additional possibility would be to average the voxel's time series inside regions of interest. However, by averaging the time series through the space, one reduces the spatial resolution of the data and it is not possible to obtain a detailed map with the most discriminant regions. As we were interested in keeping the spatial resolution of the data, especially in the single-subject classifiers, we did not use the spatial averaging approach.

In Fig. 2 and Table 1, we can see that for multi-subject classifiers both strategies used for temporal compression, i.e. scan averaging and parameter estimates from the GLM, improved the classification accuracy. However, there was no significant difference between the accuracy resulting from these two temporal compression approaches. This result shows that in the present data set fitting a box car function convolved with the hemodynamic response function to the voxel time series does not lead to better results than fitting a top-hat function and using the parameter estimates as input to the SVM and vice-versa. One possible explanation for these results is that the hemodynamic response function varies through the brain and by using a canonical hemodynamic function in the convolution model, we obtain a good

parameter estimate only for a subset of voxels. Although our results showed that considering the response as being stationary within the blocks and averaging the blocks' time points increases the SVM accuracy, this may not be a good approach for other experimental designs, e.g. event-related designs. Additional studies on testing these approaches in other data sets including event-related designs are necessary for more general conclusions.

The results presented in Fig. 3 and Table 1 show that for multi-subject classifiers space selection has no impact on classification accuracy. A possible reason for this result is that finding common regions of interest using a group of subjects and using the information to classify a new subject is not feasible by the tested approaches as inter-subject variability is too high for this approach to be beneficial.

In contrast, the results presented in Fig. 4 and Table 2 show that for single-subject classifiers both temporal compression and space selection lead to an improvement on the SVM accuracy, however, these effects were not significant in the generalized linear model including temporal compression and space selection as categorical explanatory variables. The low samples size (number test examples for the single-subject classifiers) probably resulted in the low statistical power of the test. A second analysis including only space selection as a categorical explanatory variable in the model (whole brain and space selected data) showed a significant effect of space selection on the accuracy of single-subject's SVM classifier.

The fact that spatial selection improves the accuracy for classifiers trained on single-subject (unsmoothed data) but not for classifiers trained on multi-subjects (smoothed data) is consistent with the results described by Kriegeskorte et al. (2006). By using a "searchlight" whose contents were analyzed multivariately at each location in the brain, they demonstrated that there is experimental condition relevant information contained at the fine-scale spatial patterns of unsmoothed fMRI data. They argued that, at the fine spatial scale, activity patterns, like fingerprint, may be unique to each individual. However, at the coarse spatial scale, the single-subject information can be combined in standard space to obtain a group summary.

A visual inspection of the discriminating maps presented in Figs. 5–7 shows that the maps obtained by using temporal compression approach (Figs. 5C, 6C and 7C) are most robust and more similar to the SPM*t* results (Figs. 5A, 6A and 7A) than the maps obtained by using the whole data (Figs. 5B, 6B and 7B). This represents evidence for the consistency of both methods. One can expect that the most activated regions for a condition A in relation to a condition B are likely to present a high overlap with the most discriminating regions between condition A and B.

In summary, we have shown that temporal compression increases the accuracy of classifiers trained on single and multi-subject data. However, though, space selection is a good strategy for data reduction on single-subject classifiers, this is apparently not the case for classifiers trained on multi-subject data. We believe that SVM is a powerful tool for the analysis of single-subject as well as multi-subject fMRI data and anticipate rapid development in the application of SVM to clinical as well as normal subject fMRI data. It is also becoming clear that the precise way in which SVM is used for optimal results depends on the nature of the data being analyzed (e.g. single subject/multi-subject).

Appendix A. Singular value decomposition

We define a data matrix included all training data, $\mathbf{D}_{M\times N}$ with one volume per column and one voxel per row $\mathbf{D}=[\mathbf{v}_1\ldots\mathbf{v}_i\ldots\mathbf{v}_N]$ and $\mathbf{D}_c$ being $\mathbf{D}$ with the mean volume of the data set subtracted from each column. We computed the SVD of $\mathbf{D}_c$:

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

The projection of the volumes onto the principal components was carried out as:

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

where $\mathbf{U}$ is a $M\times N$ matrix containing one eigenvector or PC per column and the superscript p means projected data or matrix. 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 including the test data (one volume per column, with the mean volume of the data set subtracted from each column).

In summary, initially the data were in the 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 the voxel space.

Appendix B. Support vector machine

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

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

where $\mathbf{w}^p$ is a learning weight vector, T denotes transpose, b is an offset and $\mathbf{v}^p$ is a vector in the input space.

The learning weight vector is orthogonal to the hyperplane. If the input space is the voxel space (one voxel per dimension), the weight vector will be the direction along which the volumes of either tasks or brain states differ most. Hence, it represents a volume with the most discriminating regions, i.e. the discriminating volume (Mourao-Miranda et al., 2005). A detailed description of the SVM can be found in Schölkopf and Smola (2002) and Burges (1998). The optimal hyperplane is the one with the largest margin. Margin is the distance to a separating hyperplane from the point closest to it and for the canonical hyperplane it is defined as $1/||\mathbf{w}^p||$. To maximize the margin, one has to minimize $\frac{1}{2}||\mathbf{w}^p||^2$ subject to:

$$y_i((\mathbf{w}^p)^T\mathbf{v}_i^p + b)\geq 1 \tag{B.2}$$

where y_i is the class label and $\mathbf{v}_i^p$, i=1,…,N (N= number of training examples) are the projected fMRI volumes onto the principal components. The solution $\mathbf{w}^p$ 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:

$$\mathbf{w}^p = \sum_{i=1}^{N} \alpha_i y_i \mathbf{v}_i^p \tag{B.3}$$

The training examples $\mathbf{v}_i^p$ with non-zero coefficients α_i are called support vectors, they carry all the information relevant to the classification problem.

The class label of a test example $\mathbf{v}^p$ is computed by the hyperplane decision function given by:

$$f(\mathbf{v}^p) = \mathrm{sgn}\left(\sum_{i=1}^{N} y_i \alpha_i ((\mathbf{v}^p)^T \mathbf{v}_i^p) + b\right) \quad \text{(B.4)}$$

and the offset b is computed by exploiting:

$$\alpha_i \left[y_i ((\mathbf{v}^p)^T \mathbf{w}^p + b) - 1 \right] = 0 \quad \text{(B.5)}$$

In many real problems, a separating hyperplane may not exist (e.g. if the classes to be separated strongly overlap). To allow for the possibility of misclassification (i.e. examples violating Eq. (B.2)), Cortes and Vapnik (1995) introduced the soft margin SVM formulation by using slack variables $\xi_i \geq 0$, where $i = 1,\ldots,N$. In the soft-margin SVM, a penalization term is introduced in the objective function, i.e. one has to minimize $\frac{1}{2}\|\mathbf{w}^p\|^2 + \frac{C}{N}\sum_{i=1}^{N} \xi_i$ subject to a relaxed separation constraint,

$$y_i ((\mathbf{w}^p)^T \mathbf{v}_i^p + b) \geq 1 - \xi_i \quad \text{(B.6)}$$

The parameter C controls the compromise between having zero training error and allowing misclassification.

B.1. Weight vector or discriminating volume

As we used the projected representation of the data ($\mathbf{D}^p$) as input to the classifiers, we need to map the weight vector back to the voxel space to recover the volume with the most discriminating regions in the original space. The weight vector or discriminating volume in the voxel space is given by:

$$\mathbf{w} = \mathbf{U}\mathbf{w}^p \quad \text{(B.7)}$$

References

Allwein, E., Schapire, R., Singer, Y., 2000. Reducing multiclass to binary: a unifying approach for margin classifiers. J. Mach. Learn. Res. 1, 113–141.

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, pp. 144–152.

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.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

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.

Dietterich, T.G., Bakiri, G., 1995. Solving multiclass learning problems via error correcting output codes. J. Artif. Intell. Res. 2, 263–286.

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

Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995b. Characterizing dynamic brain responses with fMRI: a multivariate approach. NeuroImage 2 (2), 166–172.

Hamming, R.W., 1950. Error detecting and error correcting codes. Bell Syst. Tech. J. 29, 147–160.

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

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. PANAS 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.

McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 3, 143–157.

McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 6, 160–188.

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.

Mourao-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.

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.