Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers

Classification of fMRI independent components using IC-*fingerprints* and support vector machine classifiers

Federico De Martino,[a] Francesco Gentile,[a] Fabrizio Esposito,[a,b] Marco Balsi,[c] Francesco Di Salle,[a,d] Rainer Goebel,[a] and Elia Formisano[a,*]

[a]*Department of Cognitive Neurosciences, Faculty of Psychology, University of Maastricht, Maastricht, The Netherlands*
[b]*Department of Neurological Sciences, University of Naples "Federico II", Italy*
[c]*Department of Electronic Engineering, "La Sapienza" University, Rome, Italy*
[d]*Department of Neurosciences, University of Pisa, Italy*

We present a general method for the classification of independent components (ICs) extracted from functional MRI (fMRI) data sets. The method consists of two steps. In the first step, each fMRI-IC is associated with an IC-*fingerprint*, i.e., a representation of the component in a multidimensional space of parameters. These parameters are post hoc estimates of global properties of the ICs and are largely independent of a specific experimental design and stimulus timing. In the second step a machine learning algorithm automatically separates the IC-*fingerprints* into six general classes after preliminary training performed on a small subset of expert-labeled components.

We illustrate this approach in a multisubject fMRI study employing visual structure-from-motion stimuli encoding faces and control random shapes. We show that: (1) IC-*fingerprints* are a valuable tool for the inspection, characterization and selection of fMRI-ICs and (2) automatic classifications of fMRI-ICs in new subjects present a high correspondence with those obtained by expert visual inspection of the components.

Importantly, our classification procedure highlights several neurophysiologically interesting processes. The most intriguing of which is reflected, with high intra- and inter-subject reproducibility, in one IC exhibiting a transiently task-related activation in the 'face' region of the primary sensorimotor cortex. This suggests that in addition to or as part of the mirror system, somatotopic regions of the sensorimotor cortex are involved in disambiguating the perception of a moving body part.

Finally, we show that the same classification algorithm can be successfully applied, without re-training, to fMRI collected using acquisition parameters, stimulation modality and timing considerably different from those used for training.
© 2006 Elsevier Inc. All rights reserved.

* Corresponding author. Department of Cognitive Neurosciences, Faculty of Psychology, University of Maastricht, Postbus 616, 6200 MD, Maastricht, The Netherlands. Fax: +31 43 3884125.
E-mail address: e.formisano@psychology.unimaas.nl (E. Formisano).
Available online on ScienceDirect (www.sciencedirect.com).

Introduction

Non-inferential or exploratory multivariate methods are being increasingly used in fMRI data analysis. These methods provide a characterization of the data, which does not rely on the statistical testing of a few stringent hypotheses and generate potentially valuable information on the nature of signal and noise in the fMRI time series. The value of such information consists in being complementary to that of statistical inferential maps. In some cases, however, the amount of information generated by these exploratory methods may be overwhelming and not easily interpretable.

In spatial Independent Component Analysis (ICA), for example, fMRI time series are decomposed into a large number (up to the number of scans) of spatial modes (independent components, [ICs]), with associated time courses (McKeown et al., 1998). In most cases, some of these components reflect interesting spatiotemporal patterns of stimulus-induced or spontaneous brain activity; other components reflect signal artifacts or noise (McKeown et al., 2003). The basic assumption in spatial ICA is that fMRI time series can be modeled as linear mixtures of latent sources, which can be blindly recovered under the constraint that their spatial distributions are mutually statistically independent. Several recent methodological and applied contributions indicate that this approach outperforms Principal Component Analysis (PCA) and can be a useful complement to standard hypothesis-driven univariate analysis (McKeown et al., 1998). In contrast to PCA, however, in which extracted components are naturally ordered according to explained variance, ICA does not provide any intrinsic order of the ICs. The experimenter is thus confronted with the problem of selecting and interpreting a subset of 'interesting' and 'meaningful' components.

In previous fMRI applications of ICA, selection of interesting components has been performed using various approaches. The simplest approach relies on the visual inspection of IC spatial maps/time courses (Bartels and Zeki, 2005, Calhoun et al., 2001a,

b). Selection of ICs based on their visual inspection, however, is very time consuming and highly dependent on the experience of the researcher. In most cases, ICs have been selected according to the amount of linear correlation of their time course with a model of the expected responses (McKeown et al., 1998; Schmithorst and Brown, 2004; Moritz et al., 2005) or related measures in the temporal frequency domain (Moritz et al., 2003). These approaches, however, appear to contrast with the data-driven nature of ICA. As an explorative tool, ICA may be particularly useful for detecting patterns of activity whose temporal dynamics cannot be easily modeled, such as in the case of hallucinations (van de Ven et al., 2005), epileptic seizures or in sensory or cognitive paradigms in which expected hemodynamic responses may be very diverse (Duann et al., 2002; Formisano et al., 2004; Castelo-Branco et al., 2002). Furthermore, ICA is being increasingly used for the study of 'resting state' functional connectivity (van de Ven et al., 2004; Greicius et al., 2003, 2004) or as a de-noising step, which requires the selection of components reflecting noise and artifacts (Thomas et al., 2002). In all these cases, selection of ICs based on strong expectations on the profile of the IC time courses is insufficient.

Alternatively, selection of ICs has been performed using strong a priori assumptions on the spatial layout of the activation (Castelo-Branco et al., 2002; van de Ven et al., 2004). In this approach, distributed brain networks are detected by selecting ICA components that load heavily in predefined regions of interest (ROIs). A priori expectation on one or more ROIs, however, is not always available and, as in ROI-based univariate analysis, interesting processes occurring outside the predefined ROIs are ignored.

Other post hoc measures obtained from estimated ICs have been used for their sorting/selection. In analogy to PCA, McKeown et al. (1998) sorted the ICs according to their variance contribution to the original mixture. In fMRI data, however, neurophysiologically interesting phenomena are usually weaker than some of the sources representing structured noise. Thus, ranking of the ICs in this way may be not informative. Formisano et al. (2002) characterized the ICs using a combination of three descriptive measures (kurtosis of the spatial distribution, one-lag autocorrelation of the IC time course and clustering of the IC's spatial layout). The underlying idea was that 'meaningful' components aggregate in clustered regions in the three-dimensional space defined by these three measures. This heuristic criterion proved to be effective in isolating task-related components in a simple paradigm without using stimulus timing information.

The aim of the present article is twofold. First, we introduce the IC-*fingerprint*, a visual tool that aids the experimenter in displaying and characterizing the ICs. An IC-*fingerprint* is a representation of the component in a multidimensional space of descriptive measures, which can be visualized as a polar diagram. In line with Formisano et al. (2002), the underlying assumption is that ICs reflecting similar process types (e.g., BOLD activation, structured noise, movement) have similar *fingerprints*. To preserve the data-driven nature of ICA and the generality of the approach, the descriptive measures that define the space of the *fingerprints* are post hoc estimates of global properties of the ICs and do not rely on strong temporal or spatial hypotheses.

Second, we formulate the problem of selecting 'meaningful' components in the more general context of their (automatic) classification. After transforming the ICs in the multidimensional space of *fingerprints*, this problem can be formulated as

subdividing the ICs in maximally disjoint classes and finding the optimal separating set of boundaries (hypersurfaces). Many different (supervised and unsupervised) algorithms may be used for this purpose (see Mitchell, 1997). Here, we describe and validate a supervised method for the classification of the ICs based on least squares Support Vector Machines (ls-SVMs). SVMs refer to a class of machine learning algorithms introduced by Vapnik at the end of the 70s (Vapnik, 1979). Ls-SVMs are a variant of SVM which have been proved to be effective in many problems of classification and pattern recognition (Suykens et al., 2002).

We illustrate our approach in the context of a multisubject fMRI study with visual structure-from-motion (SFM) stimuli (Kriegeskorte et al., 2003). We show that the set of measures that defines the IC-*fingerprints* is informative with respect to the problem of selecting and classifying fMRI-ICs and allows a reliable detection of interesting activation patterns. Furthermore, we show that an ls-SVM classifier, which is trained with a small subset of data from one subject, can automatically classify these fMRI-ICs in all other subjects with high correspondence to an expert classification. Finally, we show that the same classification algorithm can be successfully applied, without re-training, to fMRI collected using magnetic field, acquisition parameters, stimulation modality (auditory versus visual) and timing (event-related versus block design) considerably different from the SFM experiment used for training.

Methods

General description of the approach

Fig. 1 illustrates schematically the proposed approach. Individual fMRI time series are decomposed using spatial ICA in sets of ICs (see *ICA decomposition*). Obtained ICs are 'transformed' in a multidimensional space using a number of descriptive parameters (eleven in the current implementation). These parameters are computed from spatial distributions, spatial maps and time courses of ICs and characterize the global statistical and spatiotemporal nature of the sources (see Characterization of the ICs in a multidimensional space). We define IC-*fingerprint* as the representation of an IC in this multidimensional space of parameters (Fig. 1A).

In order to classify the ICs, two steps are required (Fig. 1B). The first step consists in *training* the SVM classifier(s) using a limited subset of IC-*fingerprints* (typically the subset of ICs corresponding to an individual single run time series). This step requires expert user's intervention and supervision in labeling the ICs (see Supervised training). The second step operates fully automatically and consists in *classifying* all other components using the trained classifier(s) (see Automatic classification of ICs). Fig. 1C visualizes, for the application presented in this article, the proportion of data which has been used for training (red) and testing (blue) of the classification procedure.

ICA decomposition

Let **X** be the $T \times M$ (T = number of scans, M = number of time courses) matrix of the fMRI time series, **S** the $N \times M$ matrix whose rows S_i ($i = 1, \ldots, N$) contain the spatial processes ($N \leq T$ = number of processes) and **A** the $T \times N$ mixing matrix whose columns A_j ($j = 1, \ldots, N$) contain the time courses of the N processes and is assumed to be of full rank. The problem of the
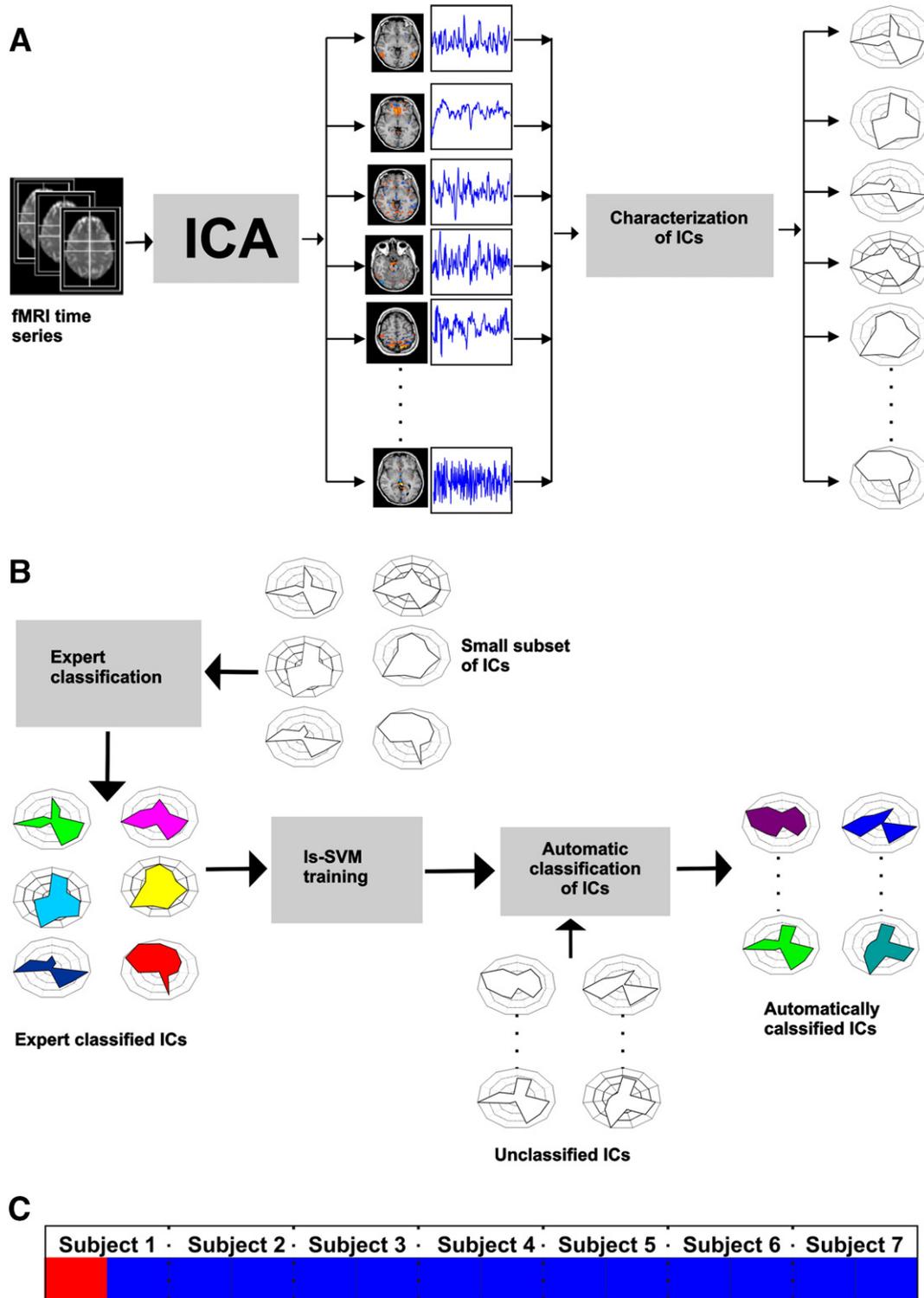
Fig. 1. General description of the approach for the characterization and classification of fMRI-ICs; (A) Independent Component Analysis of fMRI data and representation of the ICs in a multi dimensional space of *fingerprints* (see also Fig. 2). (B) Classification of IC-*fingerprints* by an ls-SVM-based algorithm. The algorithm is trained on a small subset of data labeled by an expert. (C) Proportion of data which has been used for training (red, 1/14) and testing (blue, 13/14) the classifier in the present article. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ICA decomposition of fMRI time series can be formulated as the estimation of both matrices of the right side of the following equation:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} \tag{1}$$

under the constraint that the processes S_i are (in the ideal case) spatially independent. No a priori assumption is made about the mixing matrix $\mathbf{A}$, i.e., about the time courses of the processes. In this model, all the spatial components, with the possible exception of one, are assumed to be non-Gaussian. Structured

(non-Gaussian) artifacts in the data (e.g., head movements, machine and physiological artifacts) are not explicitly modeled, but instead are treated as independent sources and are expected to be represented in one or more of the components.

The amount of statistical dependence within a fixed number of spatial components can be quantified by means of their mutual information. Thus, the ICA decomposition of $\mathbf{X}$ can be defined as a linear transformation:

$$\mathbf{S} = \mathbf{W} \cdot \mathbf{X} \tag{2}$$

where the matrix $\mathbf{W}$ (the "unmixing" matrix) is determined such that the mutual information of the target components S_i is minimized. Matrix $\mathbf{A}$ can be computed as the pseudo-inverse of $\mathbf{W}$. Note that this definition of ICA and Eq. (2) implies that ICs are determined up to a permutation, a multiplicative constant and to the sign. In the present context, the indeterminacy with respect to the permutation is important because it implies that there is no intrinsic ordering of the ICs, which is in contrast with PCA.

We estimated $\mathbf{S}$ using cortex based ICA (cb-ICA) (Formisano et al., 2004) as implemented in BrainVoyager 2000 (Brain Innovation, Maastricht, The Netherlands). Cb-ICA uses individual anatomical constraints and a fixed point ICA algorithm (FastICA, see Hyvärinen, 1999) and allows an optimized analysis of cortical sources. After sphering the matrix $\mathbf{X}$ and reduction of the temporal dimension of the data set with PCA (see below), the hierarchical (deflation) mode of the FastICA algorithm was used and the components were estimated one-by-one. After the decomposition, voxel values of IC spatial maps were z-transformed and color coded according to the absolute value and sign (McKeown et al., 1998). It should be noted that the z-values do not pertain to any significance statistic, because no comparison is made to a null hypothesis.

Characterization of the ICs in a multidimensional space

Spatial ICA decompositions of fMRI time series result in sets of ICs, whose dimensionality is determined by the number of scans or by the (optional) reduction of the temporal dimensions with PCA. Conventionally, interpretation of results requires expert inspection of each IC's voxel values distribution (histogram) (Fig. 2A), spatial map (Fig. 2B) and time course (Fig. 2C). Temporal information can be additionally expressed in the frequency domain by computation of a power spectrum (Fig. 2D). The characterization of ICs which is described in this section has the purpose of aiding in this task and to highlight similarities among the ICs.

For each IC, values of eleven descriptive measures were derived from the IC's voxel values distribution (kurtosis, skewness, entropy), spatial layout (degree of clustering in the anatomical space) as well as their temporal (one-lag autocorrelation, entropy) and spectral (power contribution in five different frequency bands) properties. The exact definition of these measures and the rationale behind their inclusion are given in Appendix 1.

We define IC-*fingerprint* the representation of an IC as a point of the multidimensional (eleven-dimensional) space of parameters. IC-*fingerprints* are visualized using polar plots with eleven axes, each of them corresponding to one of the parameters normalized between 0 and 1 (Fig. 2E).

Classification of the ICs with least-squares support vector machines

We formulate the problem of classifying fMRI-ICs as the problem of assigning the corresponding ICs-*fingerprints* to one of C classes of sources. Based on previous experience on fMRI-ICA and a preliminary analysis of data presented in this article (see Results), we consider here six classes of sources ($C=6$): (1) the 'BOLD' class includes components that are thought to reflect consistently task-related, transiently task-related and non task-related (e.g., default state) neuronal activity; (2) the second class (MOT) includes residual motion artifacts; (3) the third class (EPI) includes the typical EPI susceptibility artifact which is maximal in the frontal part of the brain; (4) the fourth class (VESSEL) includes physiological noise with highly localized peaks (e.g., large vessels); (5) and (6) the fifth and sixth class include noise at high spatial (SDN, spatially distributed noise) or temporal (tHFN, temporal high-frequency noise) frequency. For visualization, each class is labeled with a different color (see Fig. 6A for color definition and representative exemplars of each class).

In this article we approach the classification problem using a supervised machine learning algorithm based on ls-SVMs. A complete mathematical account of this approach is beyond the scope of the article. For reader's convenience, we include a brief description of SVM and ls-SVM-based classification in a binary or multi-class case. Further mathematical details on SVM and ls-SVM can be found in Cristianini and Shawe-Taylor (2000) and Suykens et al. (2002), respectively.

Support vector machines and least-square support vector machines (binary classification)

Let us consider a training set:

$$\{\mathbf{f}_i, c_i\}, i = 1, \dots l, c_i \in \{-1 + 1\}; \quad \mathbf{f}_i \in \mathfrak{R}^d \tag{3}$$

where $\mathbf{f}_i$ is a d-dimensional IC-*fingerprint*, whose class c_i is known (e.g., defined by the user). Let us further suppose that the classes are linearly separable, which is equivalent to:

$$c_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 0 \forall i \tag{4}$$

where $\mathbf{w}$ is the normal to an hyperplane and b the distance of the same hyperplane to the origin of the multidimensional space. In linear SVMs the optimal boundary between the classes is obtained by finding the hyperplane (defined by $\mathbf{w}$ and b) that maximizes the distance to the nearest training points of the two classes. Such points are referred to as support vectors ($\mathbf{f}_{sv}$) and lie on the marginal hyperplanes defined by:

$$c_i(\mathbf{w}^T \mathbf{f}_{sv} + b) = 1. \tag{5}$$

This problem can be formulated as:

$$\min_{w,b} J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \tag{6}$$

subject to:

$$c_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 1, i = 1, \dots l. \tag{7}$$
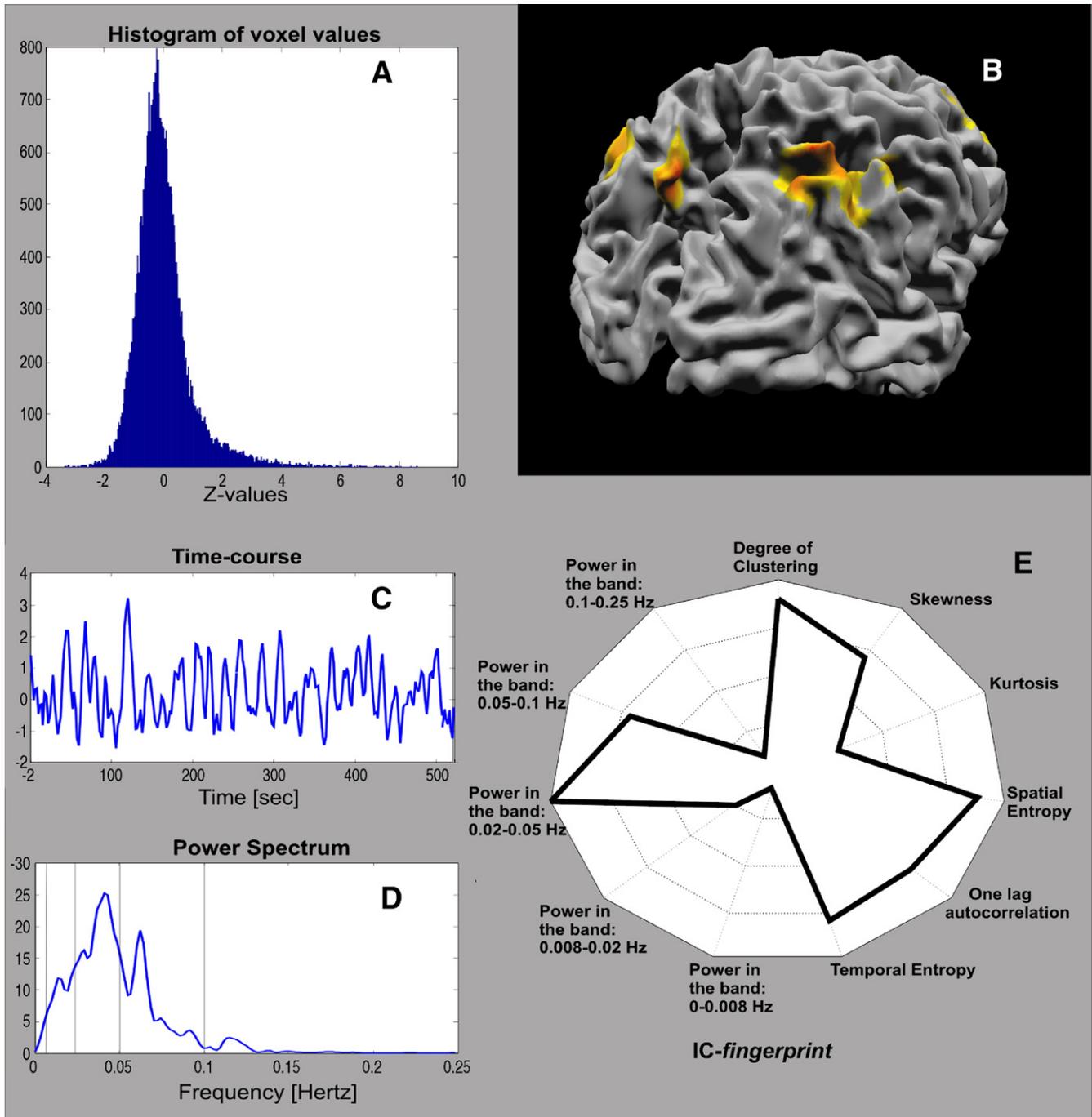
Fig. 2. Characterization of one representative component in terms of its: (A) histogram of voxel values; (B) map layout (projected on the reconstructed cortical surface of the subject); (C) time course; (D) power spectrum; (E) IC-*fingerprint*. Each axis in the polar plot corresponds to one of the normalized spatial, temporal or spectral parameters (see Appendix A for details).

Solution is obtained by Lagrangian methods (Cristianini and Shawe-Taylor, 2000). Classification of new IC-*fingerprints* $\mathbf{f}_{new}$ is obtained by evaluating:

$$sign(\mathbf{w}^T \mathbf{f}_{new} + b). \tag{8}$$

For the more general case of non-separable classes (i.e., classes with overlapping distributions) the formulation of the SVM can be modified in order to account for misclassification errors introdu-

cing additional slack variables ξ_i, $i=1,\dots,$ l. The optimization problem becomes:

$$\min_{w,b,\xi} J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + a \sum_{i=1}^{l} \xi_i \tag{9}$$

subject to:

$$c_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 1 - \xi_i, \ i=1\dots,l \tag{10}$$

and:

$$\xi_i \geq 0, \ i=1,\dots,l \tag{11}$$

where a is a positive real constant (Suykens et al., 2002).

In the present paper, we use a variant of SVM known as ls-SVMs. In the classical SVM formulation Eqs. (6–7; 9 10 11) the optimal boundary between different classes is obtained considering only the support vectors. In ls-SVMs each training point is weighted in order to obtain the distinguishing hypersurface (hyperplane). The optimization problem for the general case of non separable classes is defined as:

$$\min_{w,b,e} J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma \frac{1}{2}\sum_{i=1}^{l} e_i^2 \tag{12}$$

subject to:

$$c_i(\mathbf{w}^T\mathbf{f}_i + b) = 1 - e_i, \ i = 1,\dots,l \tag{13}$$

where γ is a positive real constant and the set of inequalities defined in Eq. (10) is replaced with a set of equalities.

SVMs (and ls-SVMs) have been modified in order to find non-linear division boundaries. In this case the data are first projected to some other Euclidean space H, using a non-linear transformation, which we call φ:

$$\varphi : \Re^d \rightarrow H.$$

A linear solution is found in the space H. as in Eqs. (6–7; 12–13). This corresponds to finding an optimal hypersurface in the original space such that:

$$c_i(\mathbf{w}^T\varphi(\mathbf{f}_i) + b) \geq \forall i \tag{14}$$

Classification of new samples is obtained evaluating:

$$sign(\mathbf{w}^T\varphi(\mathbf{f}_{\text{new}}) + b) \tag{15}$$

The use of Kernel methods (see Cristianini and Shawe-Taylor, 2000 for details) allows replacing the function φ (used for non-linear extensions) with the kernel matrix $\mathbf{K}$. In the present paper we use the Radial Basis Function (RBF) kernel whose expression is given by:

$$\mathbf{K}(\mathbf{f}_i,\mathbf{j}) = e^{-\|\mathbf{f}_i - \mathbf{f}_j\|/2\sigma^2} \tag{16}$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ refer to *fingerprint* values for two components i and j.

Extension to multi-class classification

As IC-*fingerprints* are to be assigned to one of six classes (i.e., $c_i \in \{1, 2, 3, 4, 5, 6\}$) we considered an extension of ls-SVMs to solve multi-class problems. The method we used is a generalization of the one-versus-all approach (for a complete review, see Allwein et al., 2000) and is based on Error Correcting Output Coding (ECOC) (Diettrich and Bakiri, 1995). The method requires the generation of a binary ($C \times q$) matrix (code book) that contains one row (code word) for each of the c classes and q bits. To solve the multi-class problem a binary ls-SVM is trained for each of the q columns of the code book. When a new IC-*fingerprint* is tested a binary word is generated as an output of the q binary classifiers. The point is assigned to the class that corresponds to the nearest code word. ECOC is based on the concept that redundancy helps correcting errors that may be introduced along the information channel on one or more bits (for details, see Diettrich and Bakiri, 1995). Results reported in this paper are obtained using $C=6$

(number of IC-*fingerprint* classes) and $q=5$ (length of the code word).

Supervised training

For training the SVM-based classifier and optimizing the various parameters involved (e.g., σ in Eq. (16)) we employed two sets of data. A first set was generated by inspecting and separating the ICs obtained from the cb-ICA decomposition of one run (subject 1, see below) into the six classes of sources. A second set was formed with simulated samples of the same six classes. These samples were drawn from a multivariate Gaussian distribution, whose mean and standard deviation were estimated from the first set. Optimization of parameters and learning of the distinction between the six ICs classes in the multidimensional space (i.e., definition of the separating hypersurfaces) was achieved using a cross-validation technique (Mitchell, 1997), which involved iterative training on the simulated data set and evaluation of performances on the real data set. We trained $R=50$ multi-class classifiers in order to take into account the variability introduced by the coding scheme.

Note that the use of simulated data, while introducing an assumption on the distribution of the IC-*fingerprints*, allowed keeping at a minimum the amount of real data employed during training (and thus of ICs to be inspected and labeled). An alternative approach would be to substitute the simulated data with additional real data (e.g., more data from one subject or multiple subjects).

Automatic classification of ICs

Unclassified IC-*fingerprints* were processed by each of the R classifiers. The final classification of IC-*fingerprint*s was obtained from the outcomes of the R classifications following a simple majority rule.

In order to evaluate the performances of our approach the unclassified ICs were also classified by an expert, on the basis of the visual inspection of the IC maps, time courses and power spectra (see Fig. 2A–D). The ls-SVM-based classifications were then compared with the expert classifications and true and false positive rates were estimated accordingly.

Furthermore, for comparison purposes, we report the classification performances obtained with a linear discriminant analysis, using the same training data and the same expert-labeled ICs as benchmark. In this analysis, covariance of the classes was assumed to be equal and was estimated by pooling the data across classes (Duda et al., 2000).

fMRI data

Visual structure-from-motion stimulation (block design)

We tested the proposed approach on data from a block-designed visual experiment using structure from motion (SFM) stimuli (Kriegeskorte et al., 2003). Stimuli were moving dots evoking the percept of faces or complex random shapes. The experimental conditions comprised two different types of SFM stimuli with the dots either fixed to the surface of the object (classical SFM) or moving on it (on-surface SFM). Motion and static control stimuli were also included in the stimulation protocol. In total, there were nine random-dot stimulus conditions, including moving faces (classical SFM and on-surface SFM); moving random shapes (classical SFM and on-surface SFM); static faces (on-surface SFM);

static random shapes (on-surface SFM); moving-dot control matched to classical SFM; moving-dot control matched to on-surface SFM; static-dot control (for a complete description of the stimuli see Kriegeskorte et al., 2003). Each condition appeared twice in each run, except for the two moving dot control conditions, each of which appeared only once in each run. There were, thus, 16 stimulation periods separated by 17 fixation periods. Because each period had a duration of 16 s, an experimental run lasted 8 min and 48 s. The condition sequence was pseudorandom but symmetrical.

Seven subjects between 21 and 34 years of age participated in the study (average age, 25.3 years). Each of the seven subjects underwent four runs of the SFM experiment. Results presented in this article refer to the analysis of the first two runs of each subject. Subjects were instructed to continually fixate a central cross visible throughout the experiment and to classify each stimulus presented as either face or non-face as soon as they could by pressing one of two buttons (two-alternative forced choice).

Functional scans consisted of 20 transversal slices collected at 1.5 T (Magnetom Sonata; Siemens, Erlangen, Germany) using a single-shot gradient-echo echo-planar imaging sequence (in-plane resolution 3.125×3.125 mm^2; slice thickness 5 mm; gap 0 mm; slice acquisition order interleaved; field-of-view (FOV) 200×200 mm^2; acquisition matrix 64×64; time to repeat (TR), 2000 ms; time to echo (TE), 60 ms; flip angle (FA), 90°). Each subject underwent a high-resolution T1-weighted anatomical scan at 1.5 T (Magnetom Sonata, see above), using either a three-dimensional magnetization-prepared-rapid acquisition-gradient-echo sequence lasting 8 min and 34 s (192 slices; slice thickness 1 mm; TR 2000 ms; TE 3.93 ms; FA 15°; FOV 250×250 mm^2; matrix 256×256) or a three-dimensional T1-fast-low-angle shot sequence lasting 16 min and 5 s (200 slices; slice thickness 1 mm; TR 30 ms; TE 5 ms; FA 40°; FOV 256×256 mm^2; matrix 256×256).

The fMRI data sets were subjected to a series of pre-processing operations. (1) Slice-scan-time correction was performed by resampling the time courses with linear interpolation such that all voxels in a given volume represent the signal at the same point in time. (2) Head movements were detected and automatically corrected minimizing the sum of squares of the voxel-wise intensity differences between each volume and the first volume of the run. Each volume was then resampled in three-dimensional space according to the optimal parameters using trilinear interpolation. (3) Temporal high-pass filtering was performed to remove temporal drifts of a frequency below three cycles per run (3/528 s). (4) After co-registration to the anatomical images collected in the same session the functional volumes were projected into Talairach space.

After these pre-processing steps, each of the 14 functional time series (7 subjects, 2 functional runs per subject) was decomposed using cb-ICA, which included PCA-based reduction of dimensionality to 60 dimensions (retaining more than 99% of the variance/covariance of the data) and 60×14 independent components were extracted. Each IC was projected into the space of the eleven parameters and corresponding IC-*fingerprints* were obtained as described above. ICs from the first functional run of Subject 1 were used for expert labeling (using displays as in Fig. 2) and for training the classification algorithms. The remaining components were used for testing and validating the automatic classification.

Auditory presentation of sentences (block and event-related design)

To investigate the validity of our approach in an experiment other than the SFM data set used for training, we extracted and classified

without re-training of the ls-SVM classifier ICs from two fMRI time series collected using a different stimulation modality (auditory presentation of sentences) and different stimulation timing. The two time series are part of the publicly available Functional Imaging Analysis Contest (FIAC) 2005 data set (see Poline et al., 2006 and http://www.madic.org) and refer to the single-subject data (Subject 3). In particular, we considered the first run of the block design and the first run of the event-related design. Below an essential description of these data is provided; details on the rationale of the experiment and on acquisition and stimulation procedures can be found in Dehaene-Lambertz et al. (2006, see Experiment II).

The experiment examined the functional specialization of cortical language areas using an adaptation paradigm with spoken sentences and was performed in a 3-T whole body system (Brucker, Germany). Functional images comprised 30 axial slices obtained with a T2-weighted gradient echo, EPI sequence (TR 2.5 s; TE, 35 ms; FA 80°; FOV 192×192 mm; pixels 64×64 pixels). Anatomical images were obtained using a high resolution ($1 \times 0.9 \times 1.4$ mm), T1-weighted sequence. In the block design, 20-s blocks of six sentences were presented in which either the speaker voice or the linguistic content of the sentences, or both, were repeated. Stimulation blocks were followed by 9-s 'silence' blocks. In the event-related design one sentence was presented every 3333 ms. The same conditions as in the block design were presented, but they were defined by the transition between two sentences (Dehaene-Lambertz et al., 2006).

Following standard preprocessing (see Goebel et al., 2006 for details), we decomposed the data sets using cortex-based ICA, which included PCA-based dimensionality reduction to 60 dimensions, and characterized extracted components using IC-*fingerprints*. We then proceeded in classifying the IC-*fingerprints* with the ls-SVM-based classifier, which had been trained on the visual SFM experiment as described above.

Results

ICA analysis and characterization of the ICs

Figs. 3A and 4A illustrate the results of the ICA decomposition in Subject 1 (run 1) in the visual SFM experiment. Together with IC maps, visualized on the flattened representation of the subject's cortex, and time courses these figures include the eleven-dimensional *fingerprint* of each IC.

As expected, a subset of components included ICs that were consistently related to the stimulation protocol (Fig. 3A). The spatial maps of these components encompassed a widespread set of visual ventral and parietal areas, whose activation has been detected with univariate hypothesis-driven analysis (Kriegeskorte et al., 2003). In particular, ICA decomposition highlighted three distinct spatiotemporal patterns reflecting: (1) activation in early visual areas (violet component in Fig. 3A); (2) co-activation of ventral visual regions including the lateral occipital cortex (LOC) and the fusiform-face area (FFA) and regions along the intra-parietal sulcus (IPS) (red component in Fig. 3A) and (3) co-activation in the motion complex hMT+ and temporo-parietal regions (light blue in Fig. 3A). Condition-based averages of these component time courses reflect the different functional role of these networks in the processing of the SFM stimuli (see insert in Fig. 3A) and highlight a strong selectivity for SFM faces compared to SFM control surfaces in the case of the ventro-parietal component (see Discussion). A fourth task-related
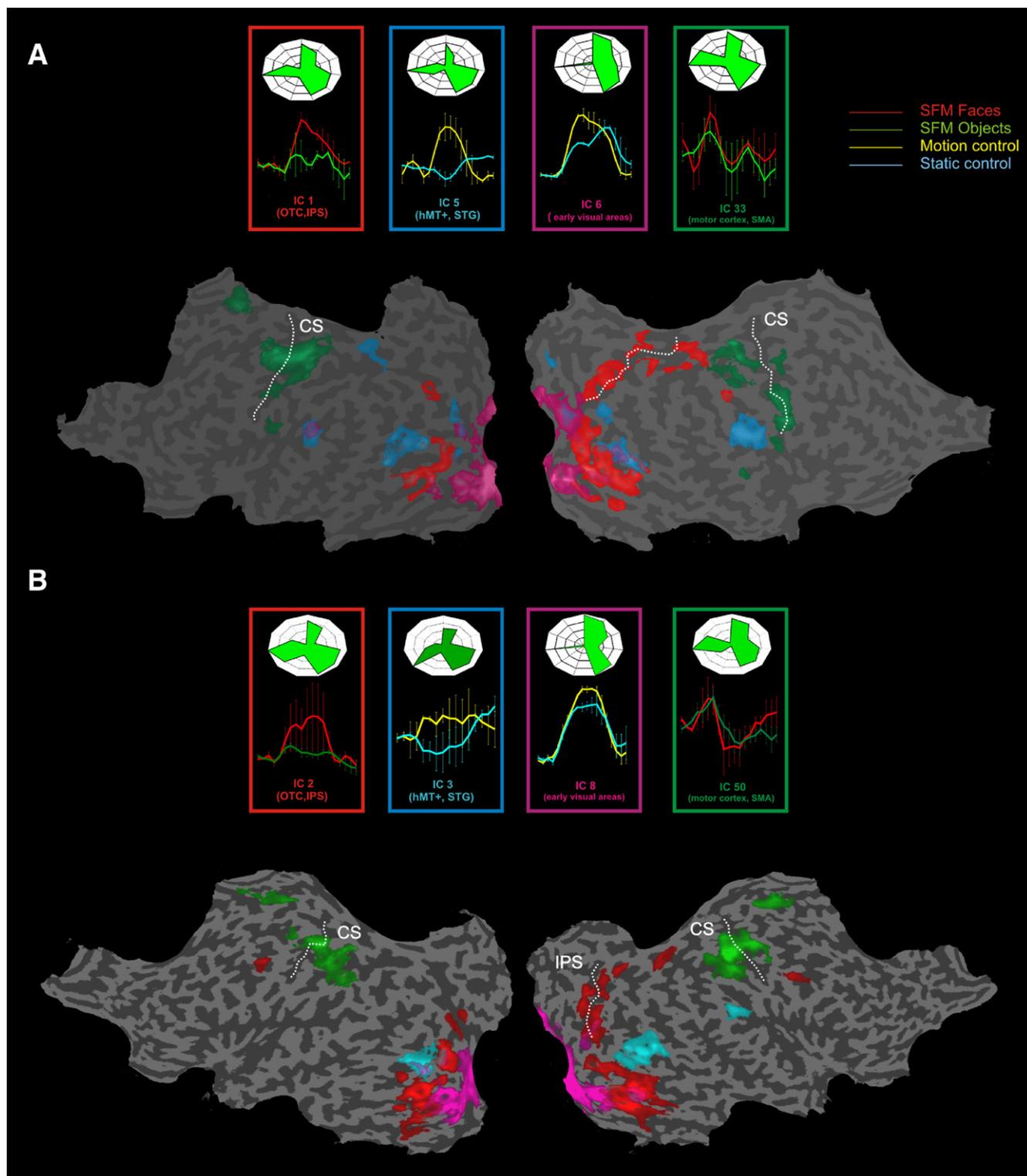
Fig. 3. Independent components consistently related to the stimulation protocol. Component maps are projected on the flattened representation of the subject's cortex: (A) Subject 1 (BS), run 1 (training set); (B) Subject 2 (AH), run 1 (automatic classification). Colored-matched inserts show condition-based averages of the IC time courses for the most relevant comparisons. Numbering of ICs is based on the order of extraction in the ICA decomposition. CS=central sulcus; IPS=intraparietal sulcus; STG=superior temporal gyrus; SMA=supplementary motor area; OTC=occipito-temporal cortex.

component (green in Fig. 3A) reflected the activation in the hand region of the central sulcus (CS) during the motor response at the beginning of each block. Importantly, IC-*fingerprints* associated

to these components showed a high degree of similarity, with a high value of degree-of-clustering and temporal autocorrelation. Note also the high contribution of low and mid-frequency range
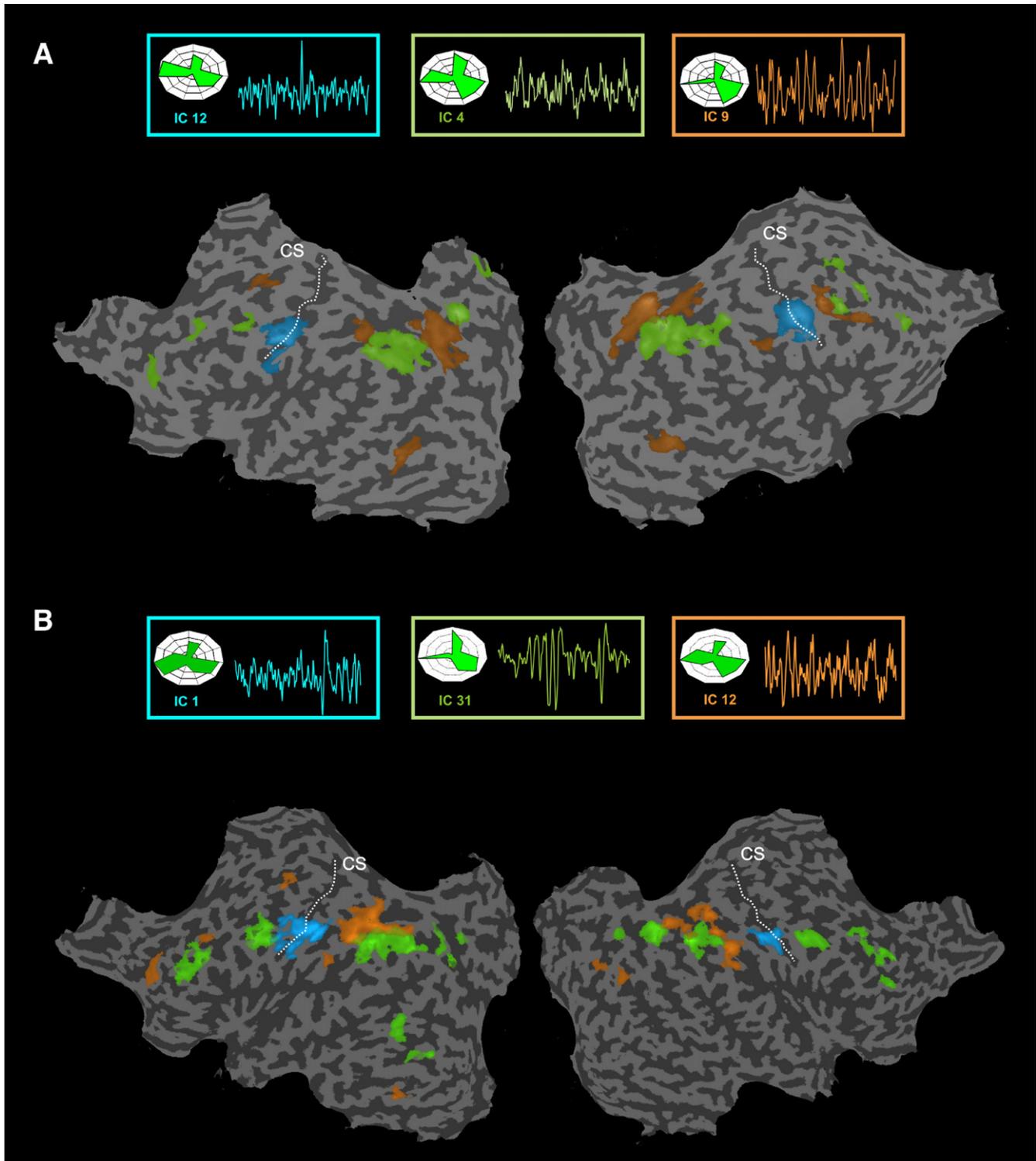
Fig. 4. Other independent components reflecting neurophysiologic sources. Component maps are projected on the flattened representation of the subject's cortex; (A) Subject 1 (BS), run 1 (training set); (B) Subject 2 (AH), run 1 (automatic classification). Colored-matched inserts show the IC time courses. Numbering of ICs is based on the order of extraction in the ICA decomposition.

in all these components, with a sharper peak for the primary visual component in the range that includes the stimulation frequency.

Inspection of the ICs with *fingerprints* most similar to those of stimulus-related ICs revealed interesting activation patterns in other specific regions or in functionally-connected networks. Although the time courses of these ICs are not (or not consistently) related to the stimulation protocol, their spatial layout together with the statistical properties of their histogram and their spectral properties are very similar to those of consistently task-related ICs,

suggesting a common neuronal/BOLD nature of the underlying sources. Fig. 4A shows three of such ICs, including two segregated fronto-parietal networks (green and brown) and a component located in the lower part of the motor cortex bilaterally (light blue).

Other ICs with comparable *fingerprints* reflected the typical *default mode* networks (maps not shown), consistently found in several recent works (Raichle et al., 2001; Greicius et al., 2003).

The remaining components reflected the contribution of artifactual sources, as it is commonly found in fMRI-ICA decompositions (McKeown et al., 2003; Thomas et al., 2002). These sources included large vessels, subject's motion, signal changes due to EPI susceptibility artifacts and noise at high spatial and high temporal frequencies. The ICs reflecting the same type of sources were associated with similar and distinctive *fingerprints*. Typical samples of these ICs and corresponding *fingerprints* are shown in Fig. 6A.

Automatic classification of the ICs

Visual structure-from-motion stimulation

The analysis of Subject 1 (run 1) described above served as the basis for generating a training data set. After training and optimization of the ls-SVM classifier, ICs resulting from the ICA decomposition of remaining data sets (13 runs, run 2 of subject 1, runs 1 and 2 of subjects 2–7) were submitted to the classifier and automatically labeled.

In each subject, the automatic classification identified as 'BOLD' a number of ICs ranging from a minimum of eight to a maximum of fifteen. In all cases, the class of BOLD ICs included the subset of consistently task-related components, accounting for the activation of early visual areas, hMT+, ventral and parietal areas, and 'hand' sensorimotor cortex (see e.g. subject 2 in Fig. 3B). The *default mode* ICs were also consistently included in the BOLD class. Furthermore, in all subjects this same class included the fronto-parietal ICs and the IC with localized activity along the CS (see e.g. subject 2 in Fig. 4B). To further investigate this latter IC we analyzed in detail its spatial layout and the time course of activity around the component peaks (Fig. 5). Consistently across subjects, the prominent peak of this IC was located in correspondence of the 'face' region of the primary sensorimotor cortex, located along the CS inferiorly and medially to the 'hand' sensorimotor region (Lotze et al., 2000). Condition-based averaging, across all subjects, of the time course of activation in the individually ICA-defined 'face' region revealed a transient activation at the stimulus onset in all conditions involving the perception of faces. Activation time courses in the FFA and in the 'hand' M1/S1 are reported for comparison.

The ls-SVM-based algorithm classified accurately most of the artifacts-related ICs in the data set. The EPI susceptibility artifact was successfully labeled in eight out of thirteen runs. The spatiotemporal features for this class were very similar to those observed for the class 'BOLD' (high degree of clustering and high autocorrelation at the first lag) but a much higher power contribution at the lowest frequency band was present. The SDN class was consistently characterized by a low degree of clustering while resembling the class BOLD in the other parameters. The highest degree of similarity was observed between the class VESSEL and the class BOLD. The class tHFN was consistently represented across runs by a very characteristic

fingerprint (high spectral contribution in the highest frequencies, and low kurtosis values). The class that presented the most variable *fingerprints* was the class reflecting residual motion artifacts (MOT).

Fig. 6B and Table 1 detail the analysis of the performances of the ls-SVM-based classification. We evaluated the correspondence of the classification obtained using the SVM-based classifier with the expert classification of all ICs.

Fig. 6B shows, for each class, a representative *fingerprint* obtained by connecting the median values along each dimension (bold line). In order to highlight the within-class range of variability, two additional *fingerprints* are superimposed to each class-representative *fingerprint*. These *fingerprints* are obtained by connecting the 5% and 95% quantiles along each dimension (dashed lines). The bottom row shows the results obtained from the ls-SVM-based classifier. For comparison, the top row shows the representative *fingerprints* of the classes obtained by expert labeling of ICs, uniquely based on the inspection of IC time courses and maps. The high correspondence between the top and bottom row *fingerprints* for the classes BOLD, SDN, tHFN indicates that the ls-SVM classifier operates as a human expert for these classes. More discrepancy and variability are noticeable for the classes VESSEL, EPI and MOT.

A characterization of the performances in terms of true positive and false positive rates is presented in Table 1. The automatic classifier reached a 94% true positive rate for the class BOLD, signifying that in 94% of the cases the classification algorithm and the human expert identically labeled these components. Restricting the comparison to the task-related ICs (i.e., to the ICs with a clear interpretation) the overlap between expert and automatic classification increased to 100%. Similarly, a 100% true positive rate was achieved for the classification of ICs in classes SDN and tHFN. For the classes EPI and VESSEL, the true positive rate was 61%. Performances dropped for the class MOT (35% true positive rate). The false positive rate was computed as the ratio between the ICs incorrectly assigned to a class and the total number of components not belonging to that class. For each class this rate was below 5%.

Table 1 also reports the classification performance obtained using a linear discriminant analysis. When compared to the expert classification, also this linear classification algorithm showed a good correspondence. However, it can be noticed that the linear discriminant classifier performed slightly worse than the ls-SVM classifier, except for MOTION and EPI classes.

We further assessed the informative nature of all measures that we included in the *fingerprints* by re-evaluating the performances of the automatic classification after removal of each measure. In all cases, removal of individual measures decreased the classification performances. Fig. 7 reports for each of the six classes the percent drop of performances of the reduced classifiers with respect to the ls-SVM-based classification obtained using all measures.

Auditory presentation of sentences

We verified the capability of our approach in identifying 'meaningful' components in a new experiment by extracting and classifying – without re-training of the ls-SVM classifier – ICs from the two fMRI time series of the FIAC data set.

Fig. 8 illustrates the spatial layout and IC-*fingerprints* of two of the components that the classifier identified as 'BOLD' for the
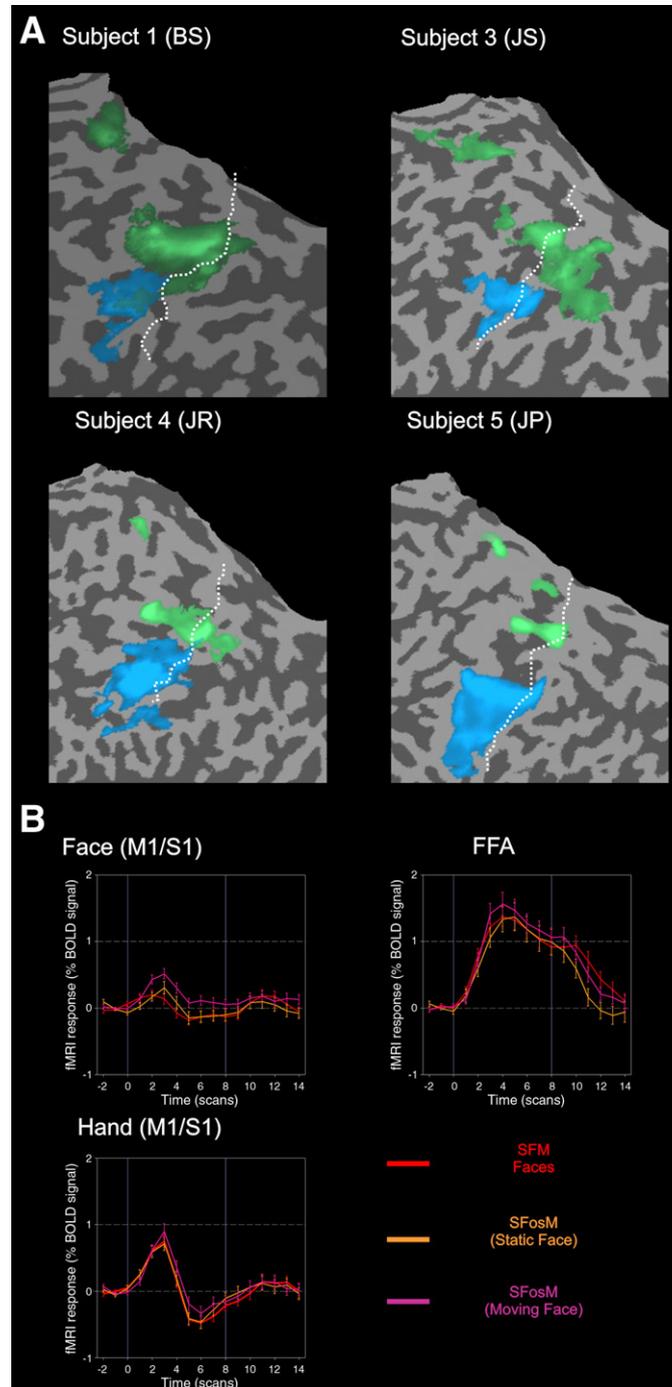
Fig. 5. (A) Details of the independent components reflecting the response in the 'face' (blue) and 'hand' (green) sensorimotor cortex along the central sulcus. Components are projected on the flattened representation of the left hemisphere of four different subjects; (B) average response to stimulation conditions involving the perception of faces in three distinctive regions.

block (Fig. 8A) and for the event-related (Fig. 8B) time series. The first component (red color) reflects the stimulus-related activation evoked by the auditory presentation in bilateral primary and secondary auditory cortical regions. The second component (blue color), which includes a more distributed temporo-frontal network with clusters located along the superior temporal sulcus and gyrus (STS/STG) and the inferior frontal gyrus, presumably reflects the linguistic processing of the sentences (see also Dehaene-Lambertz

et al., 2006 and Goebel et al., 2006). Note the visual resemblance of the IC-*fingerprints* obtained in these latter time series with those obtained in the SFM data set, despite the considerable differences between the two experiments. As in the SFM experiment, other ICs classified as BOLD were the default mode ICs (not shown, see Goebel et al., 2006). In these classifications the true positive rate (i.e., the correspondence with the expert classification) was on average 95% for the class BOLD (block: 100%, event-related:
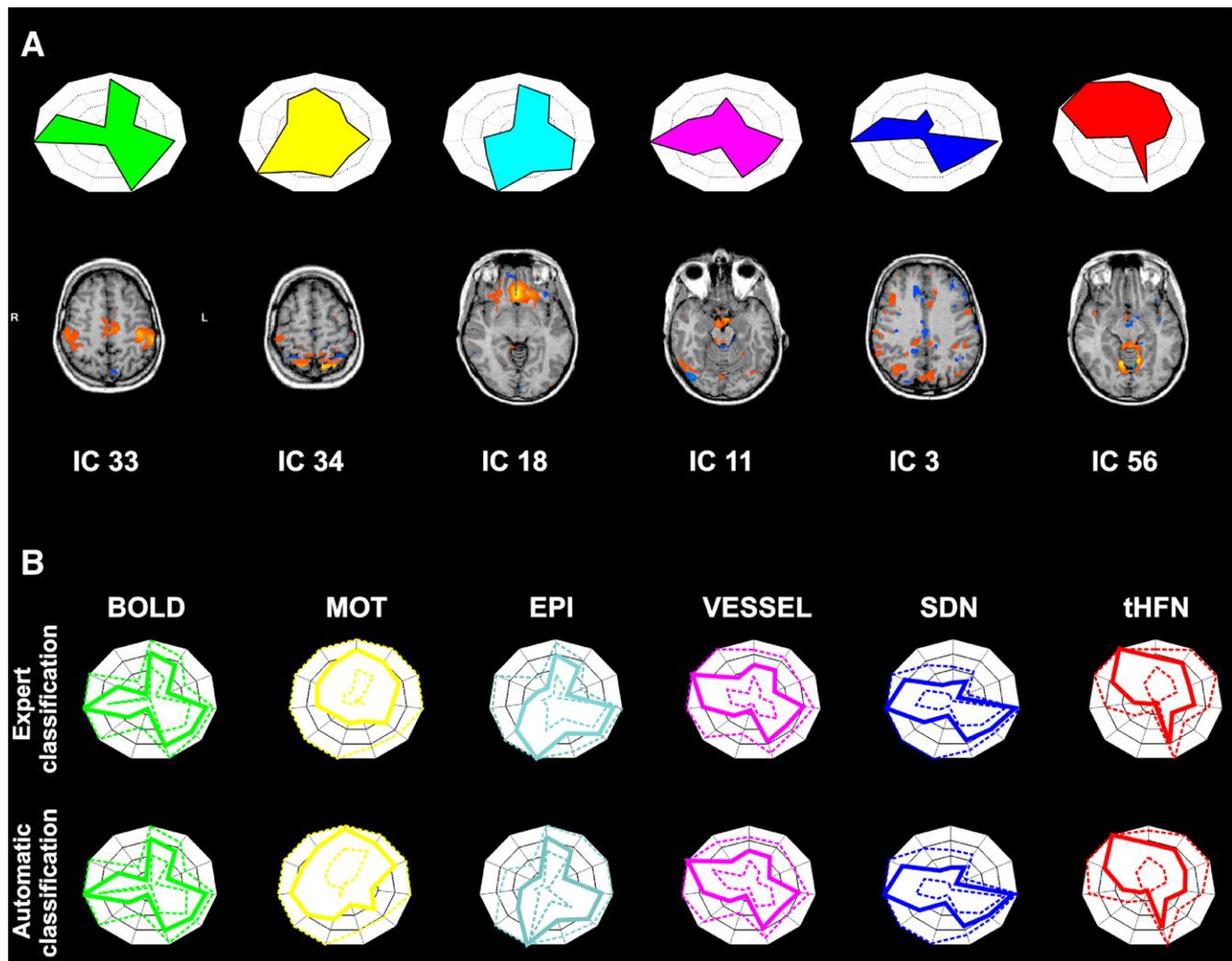
Fig. 6. Representative IC-*fingerprint* of the six classes used in the training. (A) *Fingerprints* and transversal slice of each class-representative IC in the training data set. Numbering of ICs is based on the order of extraction in the ICA decomposition. (B) Median (bold line), 5% and 95% quantiles (dashed lines) IC-*fingerprints*, for each class as obtained in the expert (top row) and automatic (bottom row) classification.

90%), 100% for the class tHFN, 81% for the class SDN (75%, 88%), 37% for the class VESSEL (40%, 33%). No component was classified as MOT or EPI.

Discussion

We have illustrated a general approach for the characterization and classification of fMRI independent components. Differently from conventional univariate statistical analyses, in which a small

set of predefined hypotheses is tested, spatial maps (and associated time courses) obtained in fMRI-ICA are determined solely by the intrinsic structure of the data. Such a data-driven analysis provides an attractive opportunity for a blind detection of potentially interesting spatiotemporal patterns (such as networks of functionally connected brain regions) and structured artifacts. At the same time, the application of ICA to fMRI data analysis challenges the experimenter with the problem of selecting a 'meaningful' subset from the large set of obtained components. Whereas this problem

Table 1
Performances of the ls-SVM based classifier on the classification of IC-*fingerprints* when compared with the expert classification

	BOLD		MOT	EPI	VESSEL	SDN	tHFN
True positive rate	94% (89%)		35% (48%)	61% (82%)	61% (60%)	100% (83%)	100% (74%)
	Task-related 100% (100%)	Others 90% (84%)					
False positive rate	0% (0%)	4% (5%)	0% (5%)	2% (3%)	2% (10%)	4% (3%)	3% (1%)

True positive rate is defined as the ratio between the number of components correctly assigned to a class and the total number of components in that class. False positive rate is defined as the ratio between the components incorrectly assigned to a class and the total number of components not belonging to that class. For comparison, performances of a linear discriminant classifier are reported in parenthesis.
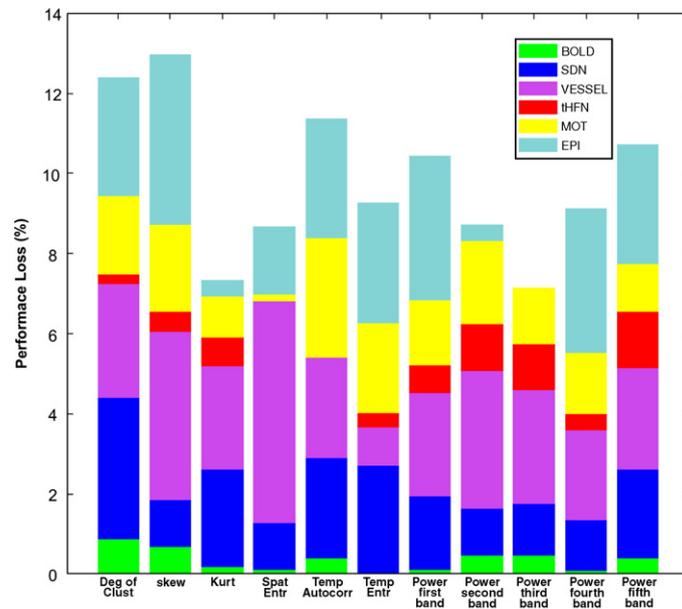
Fig. 7. Percent loss of performance (in terms of correspondence with the expert classification) of the automatic classification after removal of each of the parameters forming the *fingerprint*. Loss values are detailed for all six classes and are referred to the performance achieved by the original classifier.

ultimately requires interpretation – grounded in the knowledge of an expert – in the first part of this article we show that certain computable properties of the components can be utilized to guide the exploration of the fMRI-ICA results. In the second part of the article, we show that expert knowledge on fMRI-ICs can be transferred to a machine learning algorithm, allowing an accurate and fast automatic classification of 'signal' and 'noise' components.

In the analysis of the data used for training, the IC-*fingerprint* proved to be a powerful tool for the inspection of the ICs, independently of the automatic classification. Neurophysiologically plausible (BOLD) ICs were characterized by a distinctive *fingerprint* resulting from a high spatial and temporal structure and a prominent power contribution in the 0.01–0.1 Hz frequency range. This was not surprising in the case of consistently task-related ICs because of their simple relation to the visual stimulation (0.03 Hz) employed in the presented experiments. However, this also held true for other BOLD components which did not correlate strongly with the stimulus, such as the fronto-parietal components, the 'face' motor component and the *default mode* components, which reflected interesting effects not detected in the conventional GLM analysis (Kriegeskorte et al., 2003). These observations are in line with recent work suggesting that intrinsic neurovascular oscillations in functionally connected regions are reflected in this frequency range (Cordes et al., 2001; Raichle et al., 2001; van de Ven et al., 2004; Greicius et al., 2003). Similarly, components reflecting the same type of sources resulted in *fingerprints* with characteristic and recognizable shapes, indicating that visualizing fMRI-ICs using this display is a viable approach for their grouping.

IC-*fingerprints* were formed using a set of eleven measures (dimensions). These measures were chosen on the basis of previous experience (Formisano et al., 2002) and of theoretical and heuristic considerations on global statistical and spatiotemporal properties of the fMRI signal (see Appendix 1). In the fMRI data sets we analyzed, the selected set of measures proved to be informative with respect to the problem of differentiating the various source

types. The informative nature of the measures forming the IC-*fingerprints* is supported not only by a qualitative inspection of the *fingerprint* shapes but also by the results of the re-evaluation of the classification performances after removal of each of the measures. For each individual measure, the observed drop of performances can be taken as an indication of its relevance to the classification problem.

While showing that the IC-*fingerprint* representation is of practical relevance and utility in the analysis of fMRI data, our work does not allow concluding that the proposed representation is 'optimal' in the sense of classification performance. IC-*fingerprints* can be easily extended (and possibly improved) so as to incorporate a larger set of general features (such as stationary index of the time series or spatial smoothness of the maps). The inclusion of a new (set of) measures requires a new training of the classifier and a new analysis of the classification performance.

The ls-SVM algorithm learned expert knowledge on ICs from a very small training data set and was able to generalize this knowledge across runs and subjects, even with different order of stimulation conditions. The algorithm achieved very high levels of correspondence with the expert in detecting task-related and other BOLD ICs in general. In each subject, the ICs labeled as 'BOLD' included expected (i.e., consistently stimulus-related) responses in early visual areas, hMT+ and temporal regions, ventral occipito-temporal and inferior parietal regions. Interestingly, this ventro-parietal component indicates co-modulation of activity in regions of the 'what' and 'where' processing streams during the encoding of complex surfaces. Unexpectedly, but with high intra- and inter-subject reproducibility, the 'BOLD' class also included two separate ICs consisting of two spatially segregated bilateral fronto-parietal networks and one component consisting of the face region of the sensorimotor cortex and adjacent precentral sulcus (see Fig. 4). Although a conclusive interpretation of these findings is not possible without confirmatory experiments, it is interesting to relate these observations with results from recent neuroscientific literature.
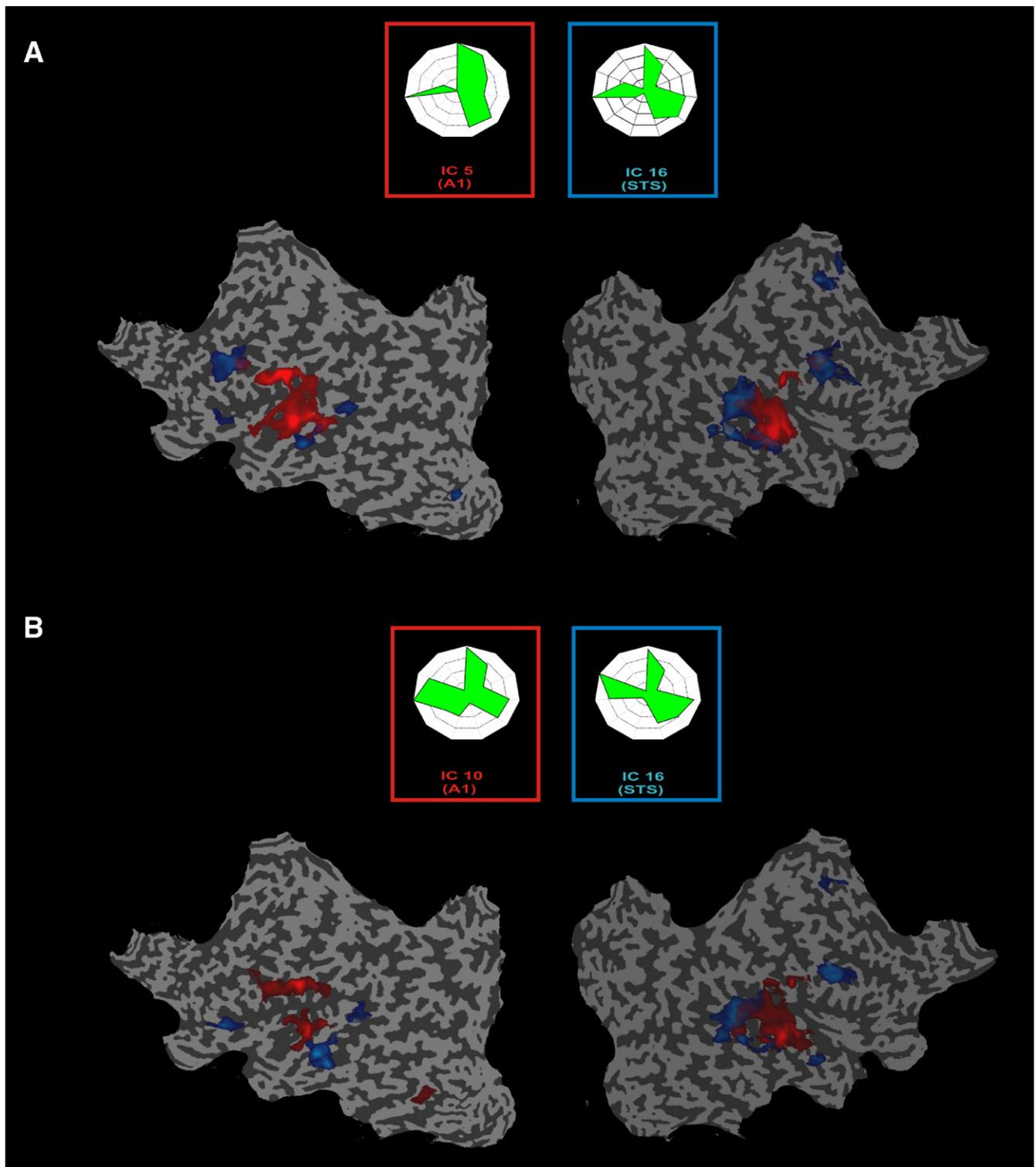
Fig. 8. Task-related ICs and corresponding *fingerprints* of the time series collected during auditory presentation of sentences. ICs were automatically identified by the ls-SVM classifier trained on the SFM data. (A) Block design experiment. (B) Event-related experiment. Red: auditory component; blue: temporo-frontal component.

The systematic segregation of the clusters in the two fronto-parietal components with a high power contribution in the low-frequency bands suggests that these sets of regions belong to functionally connected networks similar to those previously reported in studies of resting state with fMRI (Cordes et al., 2001; van de Ven et al., 2004; Greicius et al., 2003; Raichle et al., 2001), electroencephalography (Tucker et al., 1986) and direct neuronal recording (Leopold et al., 2003). In the present case, however, subjects were involved in a perceptual task and thus it cannot be excluded that these ICs reflect neuronal activity with a

more complex relation to the stimulus (e.g., control or suppression of eye movements, changes in attentional state of the subjects) and which is not captured by the linear relation of the IC time courses with the stimulation protocol. The interpretation of these ICs may be aided, in future studies, by simultaneous collection of additional data (e.g., recording of eye movement data and electroencephalographic activity).

The consistent presence in all subjects of a component with high values clustered around the 'face' sensorimotor regions suggests a transient involvement of these regions in the SFM perceptual task. At the beginning of each stimulation block, subjects were asked to indicate, by pressing one of two buttons, whether the stimulus presented was either a face or a non-face control, as soon as enough evidence was extracted from the moving dot stimuli. It is plausible that subjects solved such a task by reverting to the use of implicit motor (or somatosensory) imagery (Parsons et al., 1995), i.e., they imagined their face moving (or they imagined dots moving on their face) in order to recognize whether the SFM stimulus was a moving face or not. Alternatively, this effect may be interpreted as an automatic response, generating the interesting hypothesis that, in addition to or as part of the mirror system, somatotopic regions of the primary motor and somatosensory cortex are involved in the recognition of a moving body part (Rizzolatti and Craighero, 2004). To test these hypotheses and, more in general, to elucidate the role of primary sensorimotor regions in the perception, recognition and processing of faces, we are currently designing fMRI experiments that exploit parametric manipulations of SFM stimuli.

This transient activation in the face–motor region was undetected in a conventional GLM analysis, in which stimulus-related BOLD responses were modeled as sustained responses. In general, the use of more flexible models (e.g., Fourier basis functions, FIR models) or the explicit inclusion of a transient predictor may also have lead to a similar result. However, relevant differences between ICA and GLM-based approaches remain and are to be noted. First, ICA sensitivity is not influenced by the trial-by-trial variability of its time course, which is (conventionally) not modeled in GLM-based approaches (Duann et al., 2002). Second, depending on the correspondence between hemodynamics and models employed, GLM sensitivity may be different in different regions of the brain. Third, the detection of weak, transient effects with ICA may be favored by the 'automatic' separation from confounds and noise sources, which are not known a priori and thus would not be included in the GLM model as confounds.

Automatic classification was less accurate for some of the artifactual classes, especially in the case of residual motion signal effects. The lower performances for these classes are most likely due to the small number of samples employed in the training. An additional explanation is that motion-related source processes, while being systematically extracted by spatial ICA because of their spatial structure (Kochiyama et al., 2005), are much less stationary in the temporal domain (see Esposito et al., 2003). This non-stationary behavior may reduce the effectiveness of the temporal measures and of the proposed representation in general. Especially in these cases, the performances of the classifier may be enhanced by adding new parameters that take these aspects properly into accounts, the dimensionality of the IC-*fingerprints* not being a relevant problem.

The classifier trained on the SFM data was able to successfully select 'meaningful' components in the block and event-related FIAC data set. Notably, these successful selections were obtained without further training and on time series collected using a different MR scanner and magnetic field strength (3 versus 1.5 T), a different TR (2.5 s versus 2.0 s) and different stimulation timing and modality (auditory versus visual) compared to the data used for training. This result suggests that the proposed approach can be directly applied to a great variety of data sets. More generally, however, it should be considered that larger differences in acquisition parameters (especially TR), in the stimulation timing and/or pre-processing choices (e.g., degree of spatial/temporal smoothing) may alter significantly the spatiotemporal properties of the fMRI time series, thus affecting significantly the shape of the IC-*fingerprint* and, consequently, the performance of the classifier. In such cases, it is necessary to perform a new training of the classifier using the strategy described for our training data set. Additional empirical work is ultimately required to examine to what extent and under which circumstances this extra training is required.

When limited to the detection of 'BOLD' ICs, our approach leads to a major reduction of the number of components to be inspected and interpreted. A reduction of the number of ICs can be also achieved by focusing on ICs that are most common across the sample subjects, as is the case in recently proposed group ICA approaches. An important difference, however, is that these methods rely on spatial (Calhoun et al., 2001a,b; Esposito et al., 2005), temporal (Svensen et al., 2002) or spatiotemporal (Beckmann and Smith, 2005) consistency across subjects and do not make an explicit distinction between common 'signal' (e.g., BOLD) and common 'noise' (e.g., EPI susceptibility artifact). Conversely, the reduction achieved with the proposed method is at the level of the single-subject and allows detection of subject-specific ICs independently of their spatiotemporal matching with other subjects and independently of their contribution to the variance–covariance of the group data set. This may be a very favorable option in, e.g., clinical cases (see van de Ven et al., 2005). In multisubject studies, the proposed approach can be usefully combined with a recently proposed algorithm for grouping ICA components across runs and subjects (Esposito et al., 2005). In fact, a first-level reduction of the whole set of estimated components to the class of neurophysiologically 'meaningful' components can be used to increase the power of the subsequent grouping by reducing the cross-subject interference of signal and noise components.

A final methodological consideration concerns the fact that IC-*fingerprints* are derived, in the present implementation, by post hoc measures of the fMRI-ICs. It may be interesting, in future developments, to tailor fMRI-ICA algorithms by including such global expectations on the temporal, spatial and distributional properties of the 'meaningful' components directly in the principle of estimation (Calhoun et al., 2005) or in the contrast function (Valente et al., 2005). This may allow increasing the sensitivity of current ICA algorithms in detecting spatiotemporally structured fMRI sources and sorting the interesting components during the extraction.

Conclusions

We illustrated and validated a technique for the automatic classification and the selection of relevant ICA components in fMRI data. Its most important feature is that it matches the hallmark of ICA, i.e., blind detection of unexpected, yet plausible

and interesting, neural (BOLD) activation patterns. The proposed solution facilitates the use of ICA for the explorative analysis of complex fMRI data sets. In combination with an appropriate choice of specific measures and heuristics, a similar approach to the selection of the components could be extended to other applications of ICA, such as the analysis of electro- and magnetoencephalography data.

Acknowledgments

Appendix A

A.1. Measures derived from IC map values

Components extracted using commonly employed ICA algorithms exhibit a non-Gaussian (typically super-Gaussian) distribution of voxel values. The inclusion in the IC-*fingerprint*s of measures that estimate sparseness, asymmetry and information content (kurtosis, skewness, entropy) of the IC distributions was suggested by previous related work (Formisano et al., 2002; Suzuki et al., 2001) and by the empirical observation that sources of the same type may exhibit similar distributions of map values. As these measures do not convey any information regarding the spatial structure of obtained maps (i.e., the spatial proximity of voxel values is ignored), we also included a measure (degree of clustering) that exploits the fact that meaningful processes tend to have a well-defined spatial structure (Formisano et al., 2002).

Kurtosis is a measure of the sparseness of a distribution; it is zero for Gaussian distributions (Suhir, 1997). For each IC, *kurtosis* was estimated as:

$$\text{kurt}_i = \frac{\sum_{k=1}^{N} z_{i,k}^4}{N} - 3$$

where $z_{i,k}$ represents the value of the kth voxel of the ith component and N is the number of voxels. The normalized vector entering the *fingerprints* was obtained by linear scaling transform of $|\ln(\mathbf{kurt})|$.

Skewness is a measure of the asymmetry of the distribution; it is zero for Gaussian distributions (Suhir, 1997) and has been used for tailoring ICA decompositions in fMRI (Suzuki et al., 2001). For each IC (after mean removal and variance normalization), skewness was estimated as:

$$\text{skew}_i = \frac{\sum_{k=1}^{N} z_{i,k}^3}{N}$$

where $z_{i,k}$ represents the value of the kth voxel of the ith component and N is the number of voxels. The normalized vector entering the *fingerprints* was obtained by linear scaling transform of $|\ln(\mathbf{skew})|$.

Spatial entropy is a measure of the information content of a spatial distribution. Information content (and thus spatial entropy)

is expected to be higher for components with widely distributed values compared to components with a narrow distribution. For each IC, spatial entropy was estimated as:

$$H_i = \sum_{b=1}^{N_b} \text{hs}_i(b) \cdot \log_2(\text{hs}_i(b))$$

where hs_i represents voxel values histogram of the ith component computed over N_b bins. The normalized vector was obtained by linear scaling transform of $|\ln(H)|$.

Degree of clustering is a measure of the spatial structure of the component. For each IC, the number of voxels (N_{tot}) exceeding a threshold value in the z-normalized map ($|z| > 2.5$, see Methods) and the size of the subset of these voxels (N_{clu}) belonging to a 3D cluster of minimum extension (270 mm^3) were computed. The degree of clustering was then defined as $\text{CLU}_i = N_{\text{clu}}/N_{\text{tot}}$ with values in the interval [0, 1].

Measures derived from IC time course and frequency spectrum

The inclusion of measures of temporal structure (one-lag autocorrelation, temporal entropy) in the IC-*fingerprint*s and spectral decomposition was motivated by the following considerations. In spatial ICA no explicit constraint is posed on the time course of the sources. Nevertheless, due to the nature of the hemodynamic response, components reflecting neurophysiologically interesting sources are expected to present higher temporal structure than components reflecting noise and a concentration of spectral power at the low frequencies (0.01–0.1 Hz, Cordes et al., 2001; van de Ven et al., 2004). Conversely, frequencies above 0.1 Hz can reflect the effect of aliasing of cardiac and respiration artifacts while the typical susceptibility artifact due to the EPI sequence presents an effect at very low frequencies below 0.01 Hz.

One-lag serial autocorrelation is a measure of temporal structure. A "smooth" signal such as the BOLD responses is expected to have high values of autocorrelation (Baumgartner et al., 2000). White noise conversely is characterized by an autocorrelation function that is one at zero lag and zero everywhere else. For each IC, one-lag autocorrelation was estimated as:

$$r_i = \frac{\frac{1}{T-1}\sum_{t=1}^{T-1} a_i(t) \cdot a_i(t+1)}{\frac{1}{T}\sum_{t=1}^{T} a_i(t)^2}$$

where a_i is the time course of the ith component and T is the number of time points. The normalized vector was obtained considering $|r|$.

Temporal entropy is a measure of information content of the time course of a component. White noise is expected to have higher temporal entropy than time structured (periodic) signals. For each IC, temporal entropy was estimated as:

$$H_i = \sum_{b=1}^{N_b} ht_i(b) \cdot \log_2(ht_i(b))$$

where ht_i represents time course values histogram of the ith component computed over N_b bins. The normalized vector was obtained by linear scaling transform of $\exp(|\mathbf{H}|)$.

Power contribution: The power spectrum density for each IC time course was computed using Welch's method (see Childers, 1978). Five measures were obtained by considering the relative

contribution of 5 different frequency bands ([0, 0.008 Hz], [0.008, 0.02 Hz], [0.02–0.05 Hz], [0.05–0.1 Hz] and [0.1–0.25 Hz]) to the overall spectral power. The matrix containing the spectral measures of all the ICs sorted by rows was normalized by scaling each row to its maximum value and then scaling each column to its maximum value.

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.

Bartels, A., Zeki, S., 2005. Brain dynamics during natural viewing conditions: a new guide for mapping connectivity in vivo. NeuroImage 24 (2), 339–349.

Baumgartner, R., Somorjai, R., Summers, R., Richter, W., Ryner, L., 2000. Novelty indices: identifiers of potentially interesting time-courses in functional MRI data. Magn. Reson. Imaging 18 (7), 845–850.

Beckmann, C.F., Smith, S.M., 2005. Tensorial extensions of independent component analysis for multisubject fMRI. NeuroImage 25 (1), 294–311.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001a. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum. Brain Mapp. 13 (1), 43–53.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001b. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14 (3), 140–151.

Calhoun, V.D., Adali, T., Stevens, M.C., Kiehl, K.A., Pekar, J.J., 2005. Semi-blind ICA of fMRI: a method for utilizing hypothesis-derived time courses in a spatial ICA analysis. NeuroImage 25 (2), 527–538.

Castelo-Branco, M., Formisano, E., Backes, W., Zanella, F., Neuenschwander, S., Singer, W., 2002. Activity patterns in human motion-sensitive areas depend on the interpretation of global motion. Proc. Natl. Acad. Sci. U. S. A. 99, 13914–13919.

Childers, D.G., 1978. Modern Spectral Analysis. IEEE Press, New York.

Cristianini, Shawe-Taylor, 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.

Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., Meyerand, M.E., 2001. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. Am. J. Neuroradiol. 22 (7), 1326–1333.

Dehaene-Lambertz, G., Dehaene, S., Anton, J., Campagne, A., Ciuciu, P., Dehaene, G.P., Denghien, I., Jobert, A., LeBihan, D., Sigman, M., Pallier, C., Poline, J., 2006. Functional segregation of cortical language areas by sentence repetition. Hum. Brain Mapp. 27 (5), 360–371.

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

Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., Sejnowski, T.J., 2002. Measuring the variability of event-related BOLD signals. NeuroImage 15, 823–825.

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification, 2nd ed. Wiley.

Esposito, F., Seifritz, E., Formisano, E., Morrone, R., Scarabino, T., Tedeschi, G., Cirillo, S., Goebel, R., Di Salle, F., 2003. Real-time independent component analysis of fMRI time series. NeuroImage 20 (4), 2209–2224.

Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., Tedeschi, G., Goebel, R., Seifritz, E., Di Salle, F., 2005. Independent component analysis of fMRI group studies by self-organizing clustering. NeuroImage 25 (1), 193–205.

Formisano, E., Esposito, F., Kriegeskorte, N., Tedeschi, G., DiSalle, F., Goebel, R., 2002. Spatial independent component analysis of functional magnetic resonance imaging time series: characterization of the cortical components. Neurocomputing 49 (1–4), 241–254.

Formisano, E., Esposito, F., DiSalle, F., Goebel, R., 2004. Cortex-based independent component analysis of fMRI time series. Magn. Reson. Imaging 22 (10), 1493–1504.

Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum. Brain Mapp. 27 (5), 392–401.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. U. S. A. 100, 253–258.

Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc. Natl. Acad. Sci. U. S. A. 101, 4637–4642.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10 (3), 626–634.

Kochiyama, T., Morita, T., Okada, T., Yonekura, Y., Matsumura, M., Sadato, N., 2005. Removing the effects of task-related motion using independent-component analysis. NeuroImage 25 (3), 802–814.

Kriegeskorte, N., Sorger, B., Naumer, M., Schwarzbach, J., vandenBoogert, E., Hussy, W., Goebel, R., 2003. Human cortical object recognition from a visual motion flowfield. J. Neurosci. 23 (4), 1451–1463.

Leopold, D.A., Murayama, Y., Logothetis, N.K., 2003. Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. Cereb. Cortex 13, 422–433.

Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., Grodd, W., 2000. fMRI evaluation of somatotopic representation in human primary motor cortex. NeuroImage 11 (5), 473–481.

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 (3), 160–188.

McKeown, M.J., Hansen, L.K., Sejnowski, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? Curr. Opin. Neurobiol. 13 (5), 620–629.

Mitchell, T.M., 1997. Machine Learning. McGraw Hill International Editions.

Moritz, C.H., Rogers, B.P., Meyerand, M.E., 2003. Power spectrum ranked independent component analysis of a periodic fMRI complex motor paradigm. Hum. Brain Mapp. 18 (2), 111–122.

Moritz, C.H., Carew, J.D., McMillan, A.B., Meyerand, M.E., 2005. Independent component analysis applied to self-paced functional MR imaging paradigms. NeuroImage 25 (1), 181–192.

Parsons, L.M., Fox, P.T., Downs, J.H., Glass, T., Hirsch, T.B., Martin, C.C., Jerabek, P.A., Lancaster, J.L., 1995. Use of implicit motor imagery for visual shape discrimination as revealed by PET. Nature 375 (6526), 54–58.

Poline, J., Strother, S.C., Dehaene-Lambertz, G., Egan, G.F., Lancaster, J.L., 2006. Motivation and synthesis of the FIAC experiment: reproducibility of fMRI results across expert analyses. Hum. Brain Mapp. 27 (5), 351–359.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A.. 98 (2), 676–682.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Schmithorst, V.J., Brown, R.D., 2004. Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions. NeuroImage 22 (3), 1414–1420.

Suykens, J.A.K., VanGestel, T., DeBarbanter, J., DeMoor, B., Vanderwalle, J., 2002. Least Squares Support Vector Machines. World Scientific Publishing.

Suhir, E., 1997. Applied Probability for Engineers and Scientists. McGraw-Hill.

Suzuki, K., Kiryu, T., Nakada, T., 2001. Fast and precise independent component analysis for high field fMRI time series tailored using prior information on spatiotemporal structure. Hum. Brain Mapp. 15, 54–66.

Svensen, M., Kruggel, F., Benali, H., 2002. ICA of fMRI group study data. NeuroImage 16 (3), 551–563.

Thomas, C.G., Richard, A., Menon, R.S., 2002. Noise reduction in BOLD-based fMRI using component analysis. NeuroImage 17 (3), 1521–1537.

Tucker, D.M., Roth, D.L., Bair, T.B., 1986. Functional connections among cortical regions: topography of EEG coherence. Electroencephalogr. Clin. Neurophysiol. 63, 242–250.

Valente, G., DeMartino, F., Balsi, M., Formisano, E., 2005. Optimising ICA using generic knowledge of the sources. IEEE Prime 2005 Conference, Lausanne (Switzerland).

van de Ve n, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Hum. Brain Mapp. 22 (3), 165–178.

van de Ven, V.G., Formisano, E., Röder, C.H., Prvulovic, D., Bitter, R.A., Dietz, M.G., Hubl, D., Dierks, T., Federspiel, A., Esposito, F., et al., 2005. The spatiotemporal pattern of auditory cortical responses during verbal hallucinations. NeuroImage 27 (3), 644–655.

Vapnik, V., 1979. Estimation of Dependences Based on Empirical Data [in Russian]. Nauka, Moscow. (English translation: 1982, Springer Verlag, New York).