ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 37 (2007) 855–865

Local Linear Discriminant Analysis (LLDA) for group and region of interest (ROI)-based fMRI analysis

Martin J. McKeown,[a,b,c,*] Junning Li,[d] Xuemei Huang,[e] Mechelle M. Lewis,[e] Seungshin Rhee,[f] K.N. Young Truong,[f] and Z. Jane Wang[c,d]

[a]*Pacific Parkinson's Research Centre, University of British Columbia, Vancouver, Canada*
[b]*Department of Medicine (Neurology), University of British Columbia, Vancouver, Canada*
[c]*Brain Research Centre, University of British Columbia, Vancouver, Canada*
[d]*Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada*
[e]*Department of Neurology, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA*
[f]*Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC 27599, USA*

Received 17 October 2006; revised 18 March 2007; accepted 19 April 2007
Available online 26 May 2007

A post-processing method for group discriminant analysis of fMRI is proposed. It assumes that the fMRI data have been pre-processed and analyzed so that each voxel is given a statistic specifying task-related activation(s), and that individually specific regions of interest (ROIs) have been drawn for each subject. The method then utilizes Local Linear Discriminant Analysis (LLDA) to jointly optimize the individually-specific and group linear combinations of ROIs that maximally discriminates between groups (or between tasks, if using the same subjects). LLDA tries to linearly transform each subject's voxel-based activation statistics within ROIs to a common vector space of ROI combinations, enabling the relative similarity of different subjects' activation to be assessed. We applied the method to data recorded from 10 normal subjects during a motor task expected to activate both cortical and subcortical structures. The proposed method detected activation in multiple cortical and subcortical structures that were not present when the data were analyzed by warping the data to a common space. We suggest that the method be applied to group fMRI data when warping to a common space may be ill-advised, such as examining activation in small subcortical structures susceptible to mis-registration, or examining older or neurological patient populations.

Keywords: Discriminant analysis; fMRI; Group analysis; Regions of interest

Introduction

Group analysis in fMRI is typically done in several consecutive steps. First, fMRI data are corrected for motion, despite the fact that most methods cannot easily distinguish changes in fMRI signal from that induced by motion (Liao et al., 2005, 2006). Data are then spatially transformed to a common space such as the atlas by Talaraich (Talairach and Tournoux, 1988) or the probabilistic space suggested by the Montreal Neurological Institute (Collins et al., 1998) to minimize intersubject differences. However, because of the variability in human brain anatomy, the inter-subject registration is typically imperfect, so spatial low-pass filtering ("smoothing") is performed to de-emphasize anatomical differences (Friston, 1996). Once data have been motion corrected, warped to a common space, and spatially smoothed, the task-related activation of a voxel of a subject k is estimated with linear regression techniques:

$$Y_k = \mathbf{X}_k\beta_k + \varepsilon_k, \text{ and } \mathrm{Cov}(\varepsilon_k) = \sigma_k^2\mathbf{V}_k \quad (1)$$

where Y_k is the $T_k \times 1$ time course of the voxel, $\mathbf{X}_k$ is the $T_k \times D$ design matrix containing the hypothesized activation (often incorporating estimates of the hemodynamic response function) as well as other covariates, ε_k is the $T_k \times 1$ vector of residuals, σ_k^2 is the homogeneous variance of the residuals, and $\mathbf{V}_k$ is the correlation matrix. The subscript k indicates that all the variables are related to subject k.

As fMRI data are typically not temporally white, data are often pre-whitened using a whitening matrix $\mathbf{W}_k$ such that:

$$\mathbf{W}_k\mathbf{V}_k\mathbf{W}_k^T = I \quad (2)$$

(for an excellent summary the reader is referred to: Mumford and Nichols, 2006). If each term in Eq. (1) is pre-multiplied by $\mathbf{W}_k$, we have:

$$Y_k^* = \mathbf{X}_k^*\beta_k + \varepsilon_k^* \quad (3)$$

* Corresponding author. Pacific Parkinson's Research Centre, University of British Columbia (UBC) M31, Purdy Pavilion, University Hospital, UBC Site, 2221 Wesbrook Mall, Vancouver, British Columbia, Canada V6T 2B5. Fax: +1 604 822 7866.
E-mail address: mmckeown@interchange.ubc.ca (M.J. McKeown).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.04.072

where the superscript *denotes the whitened quantities. The whitening matrix $\mathbf{W}_k$ is estimated by the residuals ε_k and $\mathbf{V}_k$ as:

$$\mathbf{W}_k = \mathbf{V}_k^{-\frac{1}{2}}. \tag{4}$$

The regression estimates of Eq. (3) can then be estimated by Ordinary Least Squares (OLS) to give the Generalized Least Squares estimate of Eq. (1):

$$\hat{\beta}_k^{\mathrm{GLS}} = (\mathbf{X}_k^{*T}\mathbf{X}_k*)^{-1}\mathbf{X}_k^{*T}Y_k* \tag{5}$$

$$\mathrm{Cov}\left(\hat{\beta}_k^{\mathrm{GLS}}\right) = \sigma_k^2(\mathbf{X}_k^{*T}\mathbf{X}_k*)^{-1} \tag{6}$$

Contrasts between conditions are of most interest in an experiment, e.g., contrasting BOLD signal during performance of a given task compared to rest. In the study on a single subject, the null hypothesis is that the contrast between the least-squared estimates is zero:

$$H_0 : \boldsymbol{c}\beta_k = 0$$

where $\boldsymbol{c}$ is the contrast row vector. For example, if we are interested in the comparison between task 1 and task 2, $\boldsymbol{c}$ is [1, −1].

Group analyses are usually done using a Summary Statistics method, which is a two-staged approach; first individual models are fit to each subject as described above, and then a second level is applied to make group inferences on the $\boldsymbol{c}\beta_k$ (Mumford and Nichols, 2006). In the usual situation where one is contrasting activation across two groups, the second level is a multivariate regression equation with the design matrix encoded with group inclusion indicators (Fig. 1(a)):

$$\beta_{\mathrm{cont}} = \mathbf{X}_{\mathrm{g}}\beta_{\mathrm{g}} + \varepsilon_{\mathrm{g}} \tag{7}$$

where $\mathbf{X}_{\mathrm{g}} = \begin{pmatrix} 10 \\ \cdots \\ 10 \\ 01 \\ \cdots \\ 01 \end{pmatrix}$ is a binary $K\times 2$ matrix coded to show group inclusion (K is the number of subjects from the two groups), β_{cont} is composed of the contrasts $\boldsymbol{c}\beta_k$ for each individual as defined in the first stage, $\beta_{\mathrm{g}}=[\beta_{\mathrm{g}1}, \beta_{\mathrm{g}2}]^{\mathrm{T}}$ is mean activation of the two groups and $\varepsilon_{\mathrm{g}} \sim N(0,\delta_{\mathrm{g}}^2 V_{\mathrm{g}})$ is the residual with the variance δ_{g}^2 and the correlation matrix $\mathbf{V}_{\mathrm{g}}$ being a diagonal matrix, typically just I. Here the null hypothesis is that the group activations for a given voxel in the common spatially transformed space are not significantly different:

$$H_0 : \beta_{\mathrm{g}1} - \beta_{\mathrm{g}2} = 0.$$

A number of different implementations have been proposed to implement the above analysis in a practical way. The fMRIStat method uses Restricted Maximum Likelihood (ReML) to estimate σ_{g}^2 (Worsley et al., 2002), then smoothes the data to increase its degree of freedom and accuracy and finally tests the hypothesis with t-statistics. The SPM2 package (Friston et al., 2002a,b) estimates the $\delta^2\mathbf{V}+\delta_{\mathrm{g}}^2\mathbf{V}_{\mathrm{g}}$ term with ReML under a simplifying assumption that all the subjects share a common covariance matrix $\delta_k^2\mathbf{V}_k=\delta^2\mathbf{V}$, and then tests the hypotheses with F statistics. The FMRIB software library estimates σ_{g}^2 with the maximum *a posteriori* (MAP) criteria, then screens obviously insignificant voxels with Z-statistics and finally performs a Bayesian inference on the significance of the remaining voxels with a slower but more accurate Markov Chain Monte Carlo (MCMC) simulation (Beckmann et al., 1998).

Nevertheless, there are a number of shortcomings with the previously described methods. The above methods work on the voxel level—this assumes that after suitable spatial transformation, there is a perfect correspondence between the same voxel across subjects. While this may be mitigated somewhat by spatial smoothing, such low-pass filtering degrades the spatial resolution of the data. Activation estimates in small, subcortical structures such as the basal ganglia or thalami, which abut functionally different tissues (e.g., the internal capsule), may be particularly affected by mis-registration errors.

One way to partially circumvent the difficulties associated with spatially transforming functional maps to a common space is to manually draw anatomical regions of interest (ROIs) for each subject, and performing analyses at the ROI level—as opposed to the individual voxel level. Using standard atlases, a particular brain region (e.g., the lateral cerebellar hemisphere) is manually circumscribed on the high-resolution structural MRI scans that have been co-registered with the functional data, and the voxels within this region are analyzed. The benefit of this method is that it does not require rigid spatial transformation, preventing possible gross distortion of a particular brain area, as may occur if the anatomy of a given individual differs significantly in size and shape to the homologous area in the exemplar brain. However, drawing ROIs is labor-intensive, subject to human error, and requires the assumption that a functionally active region (the SMA for example) of a given brain will be within an anatomically standardized index (i.e., Broadman's Area 9) which is used to draw the ROI.

In addition to the possibilities of mis-registration, the previously described voxel-based methods do not explicitly model interactions between brain regions. Covarying regions are often of interest, but are not included in the group methods described above. Conceptually, group methods are done in two stages: in the first stage, individually specific regression models are fit to the data; and then the results of these models are used in a group-level analysis. Because the goal of these methods is to test a specific hypothesis, these methods may be conducted sequentially. In contrast, if the goal is to find which combination of brain regions is maximally different between tasks, it is desirable to jointly optimize the individual statistical model and the overall models simultaneously.

In individual-subject fMRI analysis, in addition to hypothesis driven methods, there is a role for data driven methods, such as Independent Component Analysis (ICA) (McKeown et al., 1998; Calhoun et al., 2003), which do not need rigorous *a priori* specification of activation patterns. In an analogous manner, there may be particular interest in discovering the combinations of brain regions (specified by ROIs) that are maximally contrasted during performance of certain tasks (Fig. 1(b)). There is therefore a need for a multivariate, discriminant analysis approach that works at the region of interest (ROI) level as opposed to the individual voxel level.

Previous work has taken individual activations (or the t-statistics associated with them) and used a multivariate discriminant approach (McKeown and Hanlon, 2004). In order to apply a discriminant approach, we first assume that some statistical analysis has been performed to assign a t-statistic, related to

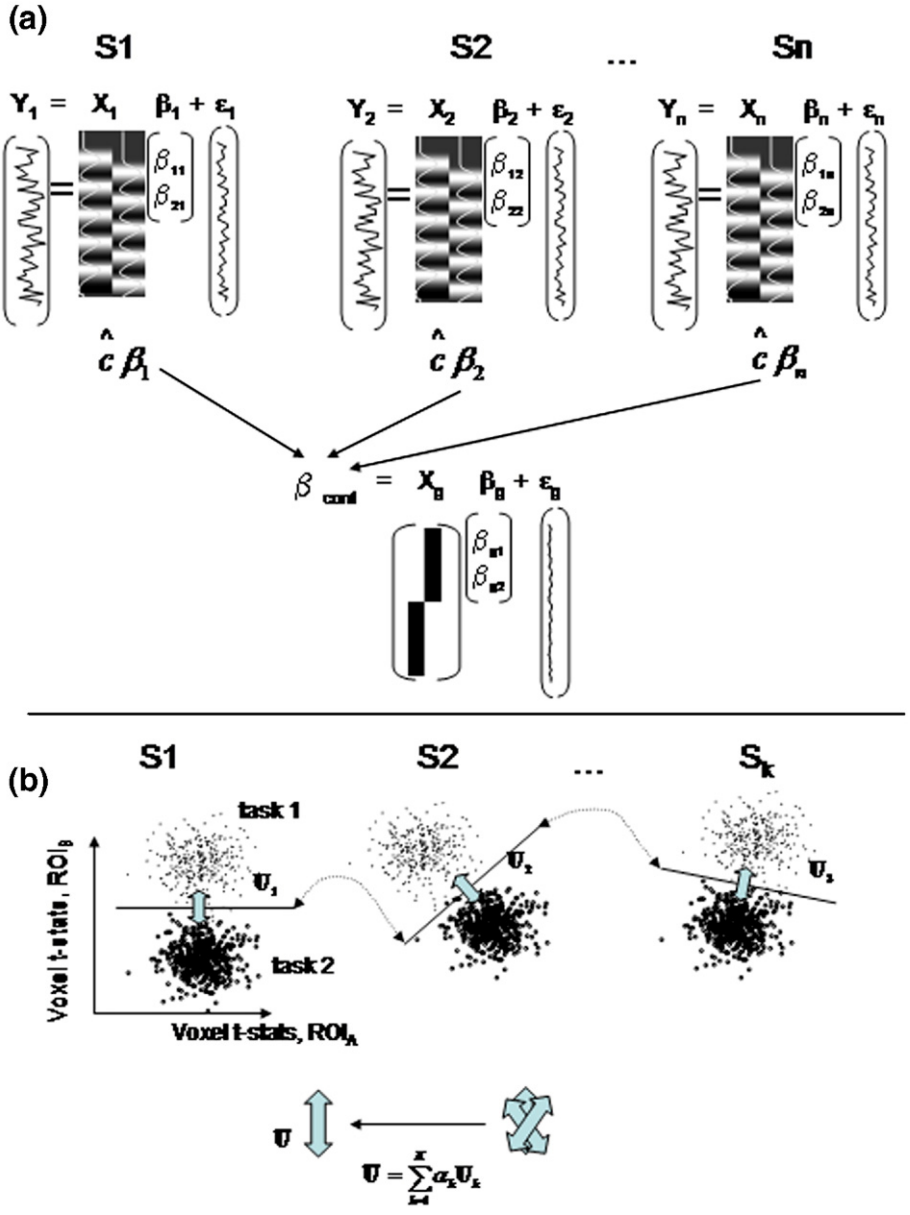

Fig. 1. Comparison of fMRI group analysis methods. (a) With a standard GLM approach, a two step approach is typically employed. An individually specific general linear model is first used to estimate task-related activity. In the second stage, the individual contrasts from each subject are collected with a dummy matrix encoding group membership. Note that this implicitly implies that each voxel across subjects is comparable, i.e., each subject's data has been spatially warped to the same space. (b) With the LLDA approach, voxel-based, activation *t*-statistics from different ROIs are compared for each subject. The optimum combination of ROIs resulting in maximal discriminability (represented by the straight lines and the $\mathbf{U}_i$s) is calculated. Note that a joint optimization is performed (represented by the curved arrows), so that the each $\mathbf{U}_i$ is not calculated in isolation. (lower panel). The individual $\mathbf{U}_i$s can then be (weighted) averaged to get an overall linear combination of ROIs that maximally discriminate between tasks.

activation, to each voxel, such as simple *t*-test or regression (e.g., *t*-statistic of a voxel comparing mean BOLD signal for a given condition to BOLD signal for that voxel at rest). Since most multivariate discriminant approaches assume Gaussian distributions, we extract features from the *t*-statistics with bootstrap methods to ensure the Gaussian assumption. Consider a subject k, we randomly select a voxel from within each of its N ROIs, and assemble the result into a column *t*-statistic vector:

$$t(k,V) = [t(k,1,v_1),t(k,2,v_2),\ldots,t(k,N,v_N)]^T, \tag{8}$$

where $t(k, r, v_r)$ is the *t*-statistic of the v_rth voxel in the rth ROI of subject k and the voxel index $\mathbf{V}=[v_1, v_2,\ldots,v_N]^T$. The random selection is repeated a number of times, say B_t times, and the ith draw is notated as $t(k, \boldsymbol{V}_i)$. After a reasonable number of draws (e.g., $B_t=30$), the normalized mean of the *t*-statistic vectors is taken as a feature to ensure the data can be modeled as multivariate Gaussian:

$$f_k = \frac{1}{\sqrt{B_t}} \sum_{i=1}^{Bt} t(K,V_i). \tag{9}$$

The above process is repeated B_f times (B_f is several hundreds or thousands) and all feature vectors are then collected:

$$F_k = [f_k(1), f_k(2), \ldots, f_k(B_f)], \tag{10}$$

where $f_k(i)$ is the ith random sample of a feature vector. This process is then repeated for all S subjects in the groups and all the F_k's are concatenated into a big matrix

$$\mathbf{X} = [F_1, F_2, \ldots F_S]. \tag{11}$$

The benefit of formulating the problem in this way is two fold; the F_k's will be asymptotically normally distributed (by the multivariate central limit theorem; Flury, 1997) even if the voxel-based statistics are not, and that p-variate linear analyses can now be performed on $\mathbf{X}$ (Flury, 1997).

The variability of the amplitude of activation across subjects is well known and typically this is dealt with by using a random-effects analysis (Mumford and Nichols, 2006). In random-effects models, the data are assumed to be derived from a hierarchy of different populations whose differences are constrained by the hierarchy. Each subject from a different group (e.g., group of normal subjects) is considered to be representative of the entire population of normal subjects.

In some cases the magnitude of inter-subject differences in fMRI activation can exceed the task-specific differences within individuals. In order to deal with this situation, yet still maintaining the benefits of linear discriminant analysis, we propose using a recently developed solution, Local Linear Discriminant Analysis (LLDA), that was initially designed to solve a somewhat different, but still related, problem (Kim and Kittler, 2005). In discriminant analysis trying to classify images of faces, often the difference in discriminant features of different poses of the same face can greatly exceed the difference in discriminant features between faces. Finding a classifier that is sensitive to images from different subjects, yet insensitive to different poses from the same subject, has been problematic. Kim and Kittler (2005) suggested LLDA as a solution to this problem. We therefore propose using LLDA to sensitively discriminate between task-dependent ROI-based patterns of activity, while being relatively robust to the differences between subjects (Fig. 1(b)).

We apply this method to data derived from a motor paradigm that would be expected to activate cortical and subcortical structures. We show that the proposed method, consistent with prior neuroscience knowledge derived from animal models, detects significant group activation in subcortical structures that was not present when the same group of data were analyzed using standard methods utilizing spatial normalization.

Experimental methods

To demonstrate the proposed method, we utilized fMRI data that would be expected, based on prior knowledge, to activate subcortical structures. We enrolled 10 healthy volunteers, 5 males and 5 females (range 27–45 years). All subjects were right hand dominant, and had normal neurological examinations, had no history of neurological disease, and were not currently using any psychoactive prescription medications. Handedness was determined according to Edinburgh Handedness Inventory. The paradigm consisted of externally guided (EG) or internally guided (IG) movements based on three different finger sequencing movements (FSMs) performed alternatively by either the right or left hand. For FSM #1, subjects had to (a) make finger-to-thumb opposition movements in the specific order of the index, middle, ring and little finger; (b) open and clench the fist twice; (c) complete finger-to-thumb oppositions in the opposite order (i.e., little, ring, middle and index finger); (d) open and clench the fist twice again; and then (e) repeat the same series of movements. The FSM #2 was the same as above except the sequence for (a) changed to index, ring, middle, and little fingers and (c) changed to the reversed order of the revised (a) (i.e., little finger, middle, ring, and index finger). The FSM #3 was the same as above except the sequence for (a) changed to middle, little, index, and ring fingers and (c) changed to the reverse of above the revised (a) (i.e., ring, index, little, middle fingers). The above three sequences (instead only one sequence) were chosen to insure the continuous engagements of the subjects' attention.

The above FSM were performed in two test conditions: *following* (externally guided movements—EG) and *continuation* (internally guided movements—IG). In the EG condition (30 s), subjects followed the finger tapping sequence shown on the screen (tapping frequency of 1 Hz). In the IG condition (30 s), the visual cue discontinued and the subjects were instructed to continue to keep tapping the same sequence as shown on the earlier screen (Fig. 2). The two consecutive conditions were preceded and followed by a rest (R) period (30 s). The EG, IG, and R periods were designated using the visual cues, "FOLLOW," "CONTINUE," and "REST", respectively. The visual cues remained on the screen with a pair of hands labeled as "Right" and "Left" to mirror the subjects' hands throughout the fMRI session. The FOLLOW–CONTINUE–REST cycle was repeated four times during each run (total duration of 6 min). There were total of 4–6 runs performed on each subject (depending upon tolerability of each subject). Subjects practiced the task for about 20 min prior to scanning session, with more than 80% correction rate for the sequence. Subjects were not able to see their own hands at anytime during the scans. All subjects were monitored for performance accuracy throughout the study with the use of a video camera mounted on a tripod. These videos were assessed for accuracy and tabulated for each subject.

FMRI image pre-processing

The fMRI data were preprocessed for each individual independently for motion correction, smoothing, and time realignment using standard parametric mapping software (SPM 5). The time series of functional images were aligned for each slice in order to minimize the signal changes related to small motion of the subject during the acquisition. Temporal filtering of functional time series included removal of the linear drifts of the signal with respect to time from each voxel's time-course and low-pass filtering of each voxel's time-course with a one-dimensional Gaussian filter with FWHM=6 s. The data were not spatially smoothed and were not spatially transformed to a common space.

Eight regions of interests (ROIs) were defined bilaterally (total=16 ROIs) and manually traced for each subject based on anatomical sulcal landmarks and with the guidance of a brain atlas (Damasio, 2005): anterior cingulate cortex (ACC), supplementary motor areas (SMA), primary motor cortex (PMC), dorsal lateral prefrontal cortex (DLPFC), caudate (Caud), globus pallidus/putamen (GP/Put), thalamus, and lateral cerebellum hemisphere.

The proposed method is a post-processing method that utilizes statistical parametric maps. To ensure that any benefits from the

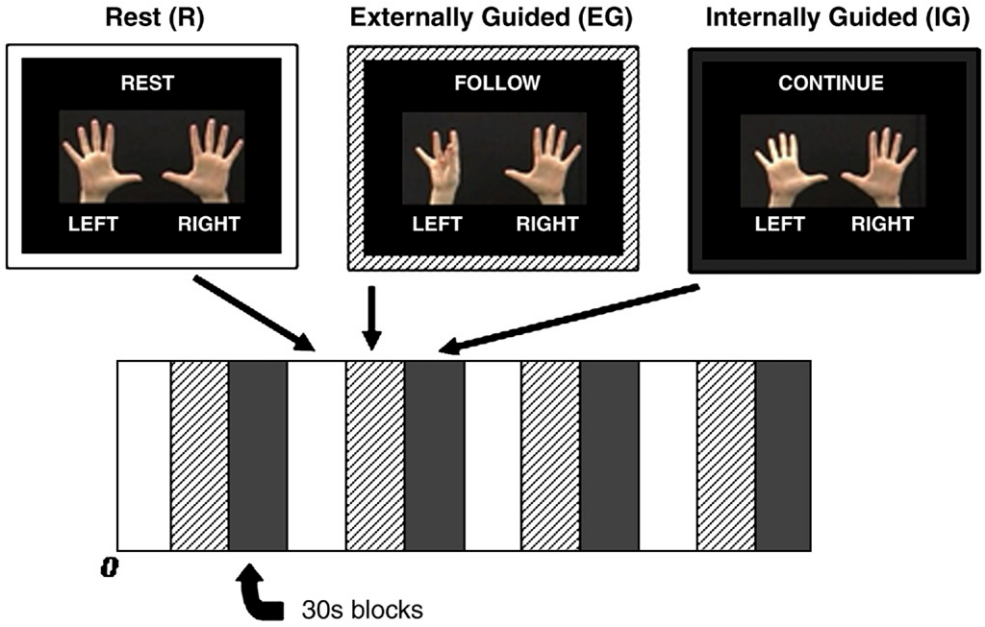

Fig. 2. Activation paradigm for fMRI motor task. A block design paradigm was used wherein subjects held their right hand at rest, followed the sequence on the screen, or generated the previous sequence internally. Each block was 30 s in duration and the rate of finger tapping was 1 Hz.

proposed method were not due to the methods of obtaining the statistical parametric maps themselves, we utilized simple *t*-tests based on the BOLD signal changes in all runs between a task and rest (e.g., right hand IG vs. rest). The voxels in the statistical maps were then labeled by the appropriate ROIs drawn on the anatomical image. The labeled statistic parametric maps for different conditions (e.g., Right Hand IG vs. rest contrasted to Right EG vs. rest) were then contrasted with LLDA.

LLDA algorithm

The underlying idea of LLDA (Kim and Kittler, 2005) is that global nonlinear data structures in many cases are locally linear and local structures can be linearly aligned. The LLDA linearly transforms each local structure (called a "cluster") to a common vector space with an individual transformation matrix and optimizes the discriminant between different classes globally in the common space.

Starting with the resampled feature matrix $\mathbf{X}$ in Eq. (11) in a study involving K subjects and two tasks, we regard the data of each subject as a local linear structure/cluster and attempt to find an individual transformation matrix for it such that the transformed data of all the subjects are globally and optimally discriminated between the tasks/classes. Let $\boldsymbol{x}$, a $N\times 1$ vector be a column of $\mathbf{X}$ and it belongs to a subject $k\in\{1,2,\ldots,K\}$ and a task $c\in\{1,2\}$. (Notations $\boldsymbol{x}\in k$ and $x\in c$ will respectively mean that $\boldsymbol{x}$ belongs to subject k and task c). Next, $\boldsymbol{x}$ is transformed to $\boldsymbol{y}$ in the common vector space with Eq. (12)

$$\boldsymbol{y}=\mathbf{U}_k^T(\boldsymbol{x}-\boldsymbol{m}_{\cdot k}) \tag{12}$$

$$\boldsymbol{m}_{\cdot k}=\frac{1}{N_{\cdot k}}\sum_{\boldsymbol{x}\in k}\boldsymbol{x} \tag{13}$$

where $\mathbf{U}_k=[\boldsymbol{u}_{k1},\ldots,\boldsymbol{u}_{kn},\ldots,\boldsymbol{u}_{kN}]$ is the $N\times N$ orthogonal transformation matrix of cluster k with $\boldsymbol{u}_{kn}$ being its nth base, $\boldsymbol{m}_{\cdot k}$, is the mean vector of cluster k, and $N_{\cdot k}$ is number of $\boldsymbol{x}$'s which belong to cluster k. The mean of each cluster is removed in the transformation.

The discriminant after the transformation is scored with Eq. (14):

$$J=\log\left(\frac{|\tilde{\mathbf{B}}|}{|\tilde{\mathbf{W}}|}\right) \tag{14}$$

where $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{W}}$ are the between-class and within-class scatter matrices in the common space. The transformed scatter matrices are defined as:

$$\tilde{\mathbf{B}}=\sum_{c=1}^{C}N_{c\cdot}(\tilde{\boldsymbol{m}}_{c\cdot}-\tilde{\boldsymbol{m}})(\tilde{\boldsymbol{m}}_{c\cdot}-\tilde{\boldsymbol{m}})^T=\sum_{c=1}^{C}N_{c\cdot}\tilde{\boldsymbol{m}}_{c\cdot}\tilde{\boldsymbol{m}}_{c\cdot}^T \tag{15}$$

$$\tilde{\mathbf{W}}=\sum_{c=1}^{C}\sum_{\boldsymbol{x}\in c}(\boldsymbol{y}-\tilde{\boldsymbol{m}}_{c\cdot})(\boldsymbol{y}-\tilde{\boldsymbol{m}}_{c\cdot})^T \tag{16}$$

where $\tilde{\boldsymbol{m}}=\frac{1}{N}\sum_{\boldsymbol{x}}y=0$ and $\tilde{\boldsymbol{m}}_{c\cdot}=\frac{1}{N_{c\cdot}}\sum_{\boldsymbol{x}\in c}y$ are the global mean and the mean of class c respectively after the transformation, and $N_{cs\cdot}$ is the number of x's which belong to class c. Because the mean of a cluster is removed in the transformation, in our case $\tilde{\boldsymbol{m}}$ equals 0. $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{W}}$ can also be written in their matrix form as:

$$\tilde{\mathbf{B}}=[\mathbf{U}_1^T\ldots\mathbf{U}_K^T]*\begin{bmatrix}\mathbf{B}_{11} & \ldots & \mathbf{B}_{1K}\\ \vdots & \ldots & \vdots\\ \mathbf{B}_{K1} & \ldots & \mathbf{B}_{KK}\end{bmatrix}*\begin{bmatrix}\mathbf{U}_1\\ \vdots\\ \mathbf{U}_K\end{bmatrix}=\mathbf{U}^T\mathbf{B}\mathbf{U} \tag{17}$$

where $\mathbf{B}_{ij}=\sum_{c=1}^{C}N_{c\cdot}\boldsymbol{m}_{ci}\boldsymbol{m}_{cj}^T$ and $\boldsymbol{m}_{ck}=\frac{1}{N_{c\cdot}}\sum_{\boldsymbol{x}\in c,\,\boldsymbol{x}\in k}(\boldsymbol{x}-\boldsymbol{m}_{\cdot k})$

$$\tilde{\mathbf{W}}=[\mathbf{U}_1^T\ldots\mathbf{U}_K^T]*\begin{bmatrix}\mathbf{W}_{11} & \ldots & \mathbf{W}_{1K}\\ \vdots & \ldots & \vdots\\ \mathbf{W}_{K1} & \ldots & \mathbf{W}_{KK}\end{bmatrix}*\begin{bmatrix}\mathbf{U}_1\\ \vdots\\ \mathbf{U}_K\end{bmatrix}=\mathbf{U}^T\mathbf{W}\mathbf{U} \tag{18}$$

where

$$\mathbf{W}_{ij}=\begin{cases}-\mathbf{B}_{ij}, & \text{if } \neq j\\ \sum_{\boldsymbol{x}\in k}(\boldsymbol{x}-\boldsymbol{m}_{\cdot k})(\boldsymbol{x}-\boldsymbol{m}_{\cdot k})^T-\mathbf{B}_{kk}, & \text{if } i=j=k\end{cases} \tag{19}$$

LLDA attempts to maximize $J=\log(|\mathbf{U}^T\mathbf{B}\mathbf{U}|/|\mathbf{U}^T\mathbf{W}\mathbf{U}|)$ under the orthogonality and normal constraint $\mathbf{U}_k\mathbf{U}_k^T=\mathbf{U}_k^T\mathbf{U}_k=I$. The constrained nonlinear programming is solved by successively calculating the bases of $\mathbf{U}_k$ from the subspace orthogonal to the already calculated bases. Unlike the original LLDA description by Kim and Kittler, we propose using an overall optimization procedure using two routines, a "subspace" routine, and a "one-base" LLDA routine, which we found more robust and reliable for fMRI data. The "subspace" routine creates subspaces orthogonal to the already calculated bases $\mathbf{u}_{k1},...,\mathbf{u}_{k(n-1)}$ and calculates $\mathbf{u}_{kn}$ in the subspace by calling the "one-base" routine which solves a one-base LLDA problem. The "subspace" routine repeats iteratively until all the bases are calculated. The "one-base" routine solves a one-base LLDA problem that is similar to LLDA but with a different constraint $\mathbf{U}_k^T\mathbf{U}_k=1$ where $\mathbf{U}_k$ is just a column vector and only subject to the normal constraint. The pseudo-code of the two routines is given in the Appendix A.

The procedure proposed here has several advantages over the original one proposed by Kim and Kittler for LLDA (Kim and Kittler, 2005). First, the orthogonalization is performed only once for every $\mathbf{u}_{kn}$ in our procedure (at the "subspace" Routine's step 4) because the orthogonality constraint is implemented in advance of solving the one-base LLDA problem by projecting $\mathbf{u}_{kn}$ (at the "subspace" Routine steps 3 and 5) to the subspaces bases A_k which is orthogonal to the already calculated $\mathbf{u}_{k1}\Lambda\mathbf{u}_{k(n-1)}$. In contrast, the orthogonality constraint in the original LLDA implementation (Kim and Kittler, 2005) is carried out in solving the one-base LLDA problem and thus the orthogonalization must be performed at every iteration, which is less computationally efficient. Second, in the one-base LLDA problem, we do not randomize the start point as Kim and Kittler did but estimate the initial starting point by solving a generalized eigenvector problem, which results in much faster and more stable convergence.

As pointed out by Kherif et al. (2003), averaging of fMRI data across individuals is only prudent when the mean is a good representation of the group. They suggested a way to look for selecting subjects with "similar" activation patterns. In the current situation, since the activation statistics from each individual have been transformed to a common vector space, defined by the *y*s, we can now selectively weight each subject so that the weighted *y*s are maximally discriminable in the transformed vector space. Specifically, we can weight each subject, k, by a small positive factor α_k,

$$\Psi^* = \alpha_k \boldsymbol{y} = \alpha_k \mathbf{U}_k^T(\boldsymbol{x} - \boldsymbol{m}_k), \tag{20}$$

subject to:

$$\sum_{k=1}^{K} \alpha_k = 1, \alpha_k \in [0,1] \tag{21}$$

where k is the subject and there are K total subjects and the means of the y's are maximally discriminable by, e.g., a standard t-test. The α_k's can then be estimated by constrained non-linear optimization methods.

To estimate the overall transformation, i.e., the linear combination of ROIs that maximally discriminate between groups, we take advantage of the individual subject α_k's computed with Eq. (20) and summarize all the individual $\mathbf{U}_k$'s by:

$$\overline{\mathbf{U}} = \sum_{k=1}^{K} \alpha_k \mathbf{U}_k \tag{22}$$

In Appendix A, we explicitly demonstrate how $\bar{\mathrm{U}}$, at least for the two-class problem, can be jointly estimated during the LLDA optimization. This joint optimization procedure gave results virtually identical to that given by Eq. (22).

Backward step-wise discrimination

To determine which ROIs should actually be included in the discriminant analysis, a backward step-wise procedure can be employed. This method first assumes that all ROIs significantly contribute to the discrimination. The least insignificant element of $\bar{\mathrm{U}}$ is then removed from the analysis, and the procedure is repeated. The fewest number of overall included ROIs that result in the maximum number of significant ROIs is then retained.

Contrasting multiple tasks

In fMRI experiments there is frequently interest in contrasting multiple tasks.

For example, given the model:

$$\mathbf{Y}_k = \mathbf{X}_k\beta_k + \varepsilon_k$$

the contrast of choice for a single task and rest is: [1 −1 0 0 …], where the first two columns of the design matrix represent the anticipated activation and rest respectively. In the current case, we are interested in contrasting activations resulting from right or left hand movement, independent of whether or not the task was externally guided (EG) or internally guided (IG), resulting in a contrast [½ ½ −1 0 0 …] if the design matrix represented [EG, IG, rest, …]. However, for the LLDA implementation, we can proceed two ways:

1. We can pool the *t*-statistics from the EG and IG tasks—in effect placing the same ROI mask over both statistical maps from the EG vs. rest and IG vs. rest to create an overall group of voxel-based statistics representing that ROI. We note that this is equivalent to creating a contrast such as [½ ½ −1 0 0 …].
2. We can treat each *task* as a separate cluster and calculate the $\mathbf{U}_k$ for *each task of each subject*.

Although standard approaches employ method (1), we have had much better empirical results with method (2). While it is difficult to analytically determine why this would be so, a possible and intuitive explanation is the "Simpson paradox" (Simpson, 1951). The paradox shows that the conditional association between two factors A and B given another factor C may be completely altered compared to their marginal association, without knowing C. In our situation, right (or left) hand movement and ROIs activities are the two factors A and B, while the condition EG or IG is the third factor. Method (1) pools the EG and IG data together and thus studies the marginal association, possibly missing meaningful interactions. Method (2) considers EG and IG separately and thus studies the conditional association.

Significance of discrimination

A necessary but not sufficient condition of assessing the role of individual ROIs in the overall transformation, $\bar{\mathbf{U}}$, results in

separation that is statistically significant. Specifically, we determine the separation on a subject by subject basis by evaluating the differences with *t*-tests (Press et al., 1992):

$$t_i \sim \frac{\overline{y}_{A,i} - \overline{y}_{B,i}}{\sqrt{\frac{\sigma^2_{y_{A,i}}}{\eta_{y_{A,i}}} + \frac{\sigma^2_{y_{B,i}}}{\eta_{y_{B,i}}}}}, \quad (25)$$

where y is the transformed data using the mean transformation, $\bar{\mathbf{U}}$, A, and B refer to different activation maps being compared, and i is the subject index. We can test the null hypothesis that the different t_is from all subjects are not significantly different from 0.

In order to determine whether or not the overall discriminant, $\bar{\mathbf{U}}$ (Eq. (22)) resulted in false positives, ROC curves were drawn to compare the distributions of $\bar{\mathbf{U}}$ under two hypotheses:

H0. None of the ROIs are differently activated when the subjects perform the two tasks.

H1. Some of ROIs are differently activated when the subjects perform the two tasks.

The distribution of $\bar{\mathbf{U}}$ is estimated from data with two resampling techniques, bootstrapping and permutation. Boostrapping resamples S subjects from the real data with replacement and simulates the random recruitment of subjects from a population. If the data of the same subject are sampled more than once, they are considered as different subjects, for the purposes of LLDA, in the resampled data. Permutation randomly shuffles the task labels of the data matrices of each subject.

The distribution of $\bar{\mathbf{U}}$ under H0 is estimated with two steps: (1) generate many new data sets with bootstrapping followed by permutation. (2) apply LLDA to the new data sets and calculate a $\bar{\mathbf{U}}$ for each data set. We performed 2000 bootstraps to estimate $\bar{\mathbf{U}}$ under H0. The distribution of $\bar{\mathbf{U}}$ under H0 was estimated similarly, but only bootstrapping was employed in the resampling. With the estimated distribution of $\bar{\mathbf{U}}$ under H0 and H1, ROC curves were drawn.

SPM methodology

We also performed SPM 5 analysis on our data. The FMRI data were pre-processed (motion corrected, spatially normalized to a common stereotactic space and smoothed with a Gaussian kernel with 6-mm full width at half maximum (FWHM)). The first level comparisons were made between right hand and left hand finger sequential movements, and the individual activation map (right-left; and left>right) for each subject was first generated by a fixed-effect model with the voxels that exceeded a probability threshold of $p=0.05$ FDR (False Discovery Rate-corrected). In order to provide a reasonable comparison, the first level analysis was restricted to those voxels within the ROIs, using the PickAtlas software package (Maldjian et al., 2003). The second level analysis was made based on the results of the individual activation maps generated in the first level comparison. That is, the contrast images, one from each of 10 subjects from the first level comparison (for example, right>left) were assessed using one sample *t*-test by a random-effect model. This can be expressed in the equation, $C\hat{\beta} = Z = \mathbf{X}_g\beta_g + \varepsilon_g$, where $C\hat{\beta}$ is the contrast result from first level analysis, $\mathbf{X}_g$ is a design matrix that is simply a single column of 1′s, β_g is a second level parameter, and $\varepsilon_g \sim N(0, \sigma^2_g)$. The regions that have clusters with at least 5 contiguous voxels exceeding a probability threshold of $p=0.001$ (uncorrected) were identified as activated regions.

Results

Using SPM 5, we found activation in the left primary motor cortex during right hand movement, and similarly left primary motor cortex activation during right hand movement. Activation in the left cerebellar hemisphere, right somatosensory cortex, and right thalamus was detected using left hand movement only (Fig. 3(a)).

Similarly, during right hand movement, significant activation was detected with LLDA in the left primary cortex. With left hand movement LLDA detected significant activity in the right primary motor cortex, the right supplementary motor area, the right striatum, the right thalamus, and the left cerebellar hemisphere (Fig. 3(b)). The ROCs demonstrated large deviations from the diagonal, suggesting that the null hypothesis of a region not being active and the hypotheses that it is active have little overlap and a threshold can easily be set to classify them well. The ROC curves also demonstrated superior performance of LLDA (solid line) over LDA (dotted line).

In comparing the SPM and LLDA results contrasting EG vs. IG, there were also distinct differences (Fig. 4). SPM analysis did not identify any clusters as having significant activity for either the EG>IG comparison or the IG>EG comparison. In contrast, for the EG vs. IG comparison LLDA detected significant activity in the left primary motor cortex and the left SMA and when comparing IG>EG tasks, a structure previously found to display increased activity during IG tasks (Strick et al., 1998). When comparing EG>IG, significant activity was found in the dorsolateral prefrontal cortex, bilaterally consistent with previous results (Otani, 2003).

The weighting across subjects for the joint optimization was relatively consistent (Fig. 5). In all cases the separation across each subject (cf. Eq. (25)) was significantly different from zero (Fig. 5).

Discussion

We have utilized a modification of the local linear discriminant (LLDA) algorithm to perform group-wise analysis on activation maps that have been manually labeled with individually drawn Regions of Interest (ROIs). In addition to finding the contralateral primary motor cortical activation and ipsilateral cerebellar activation similar to the SPM approach, we found a number of regions that were significantly active, including bilateral dorsolateral prefrontal cortex during externally guided movement. We also found differences between the use of the right and left hand, such that the right SMA was activated only during left hand performance. This result most likely reflects differences due to hand dominance.

In contrast to linear discriminant analysis (LDA), which treats each subject independently, LLDA tried to simultaneously discriminate combinations of ROIs differing between tasks within individuals, while still trying to determine a common transformation across individuals (Fig. 1). We are particularly interested in the vector space of ROI-based activation statistics that is invariant across individuals, but differs by task activation. A permutation/bootstrapping approach has been employed to prevent false positives.

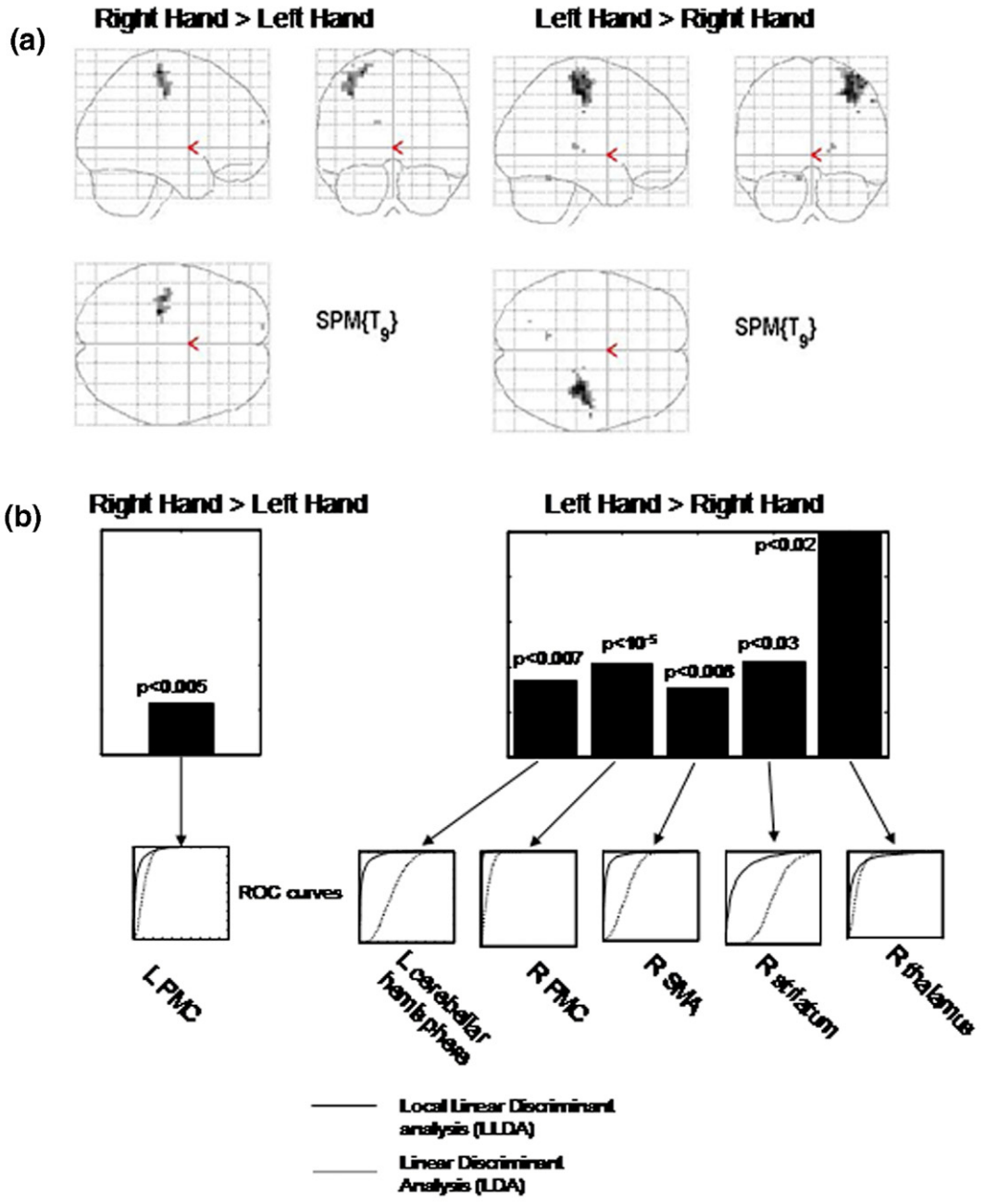

Fig. 3. Comparison of group results. Note that this is a combination of EG and IG movements. (top panel) SPM results; (bottom panel) LLDA results.

A useful feature of LLDA is that it transforms the activation statistics from each subject (or each tasks of each subject) into a common vector space with respect to relative activation of ROIs during task performance. This allows us to meaningfully compare how similar the combined activation, specified by $\bar{U}$, is across subjects (Fig. 5) providing a check for subjects whose activations represent outliers.

The proposed method may be computationally expensive. To run 10 subjects with 16 ROIs, using full backward step-wise discrimination, with LLDA at each step, took approximately 30 min total on a P4 PC. As this was implemented in Matlab for programming convenience, it is expected that compilation could improve the computational performance.

By repetitively resampling voxel-based statistics to ensure a multivariate Gaussian distribution of the feature vectors (Eq. (9)) there is a risk that our bootstrap samples may no longer be *i.i.d.* Often researchers restrict the ROIs to the intersection between anatomically derived ROIs and voxels deemed "active" for a given task. However, by restricting subsequent analyses to these active voxels, the distribution of activation statistics will necessarily be highly non-Gaussian. A balance must therefore be obtained between making the data more Gaussian and the risk of using non-*i.i.d.* bootstrap samples. Since recent work has suggested that the *i.i.d.* assumption can be violated without disrupting the results excessively (e.g., Politis, 2003), we suggest the current approach is reasonable.

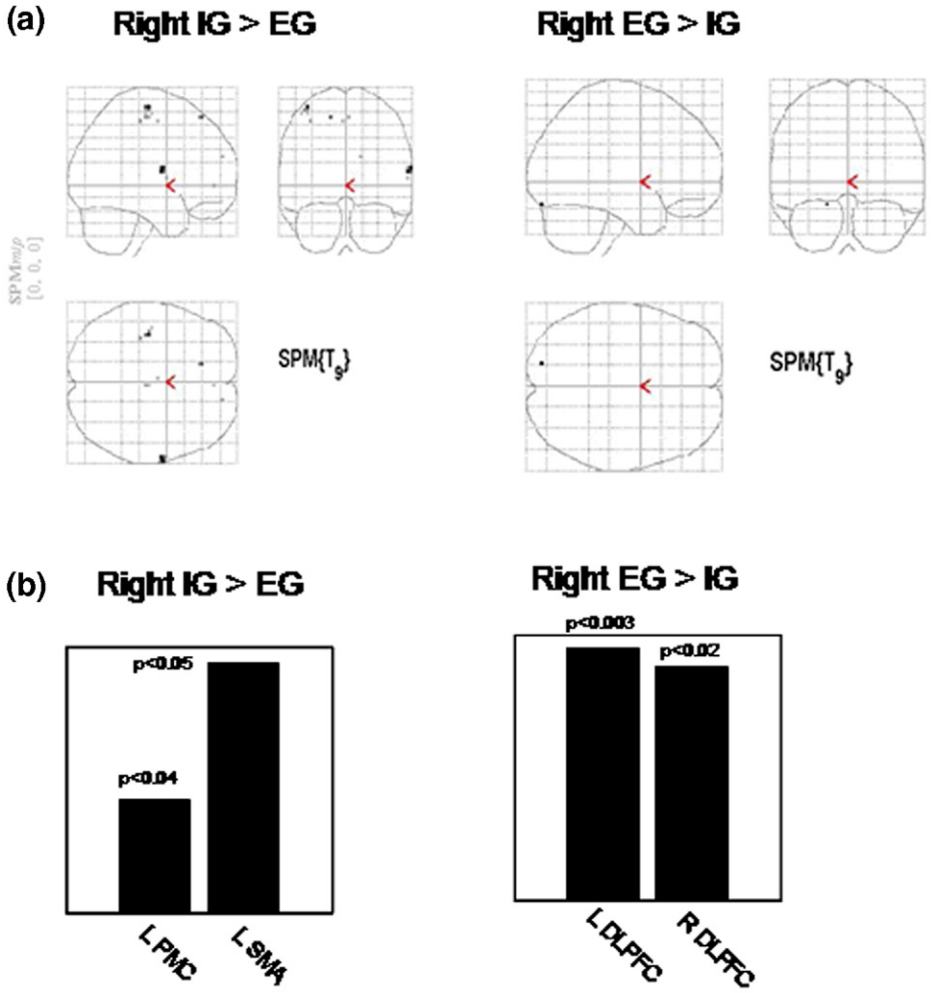

Fig. 4. Comparison of group results, examining movement type (EG vs. IG). (top panel) SPM results; (bottom panel) LLDA results. ROC curves are shown for significantly active regions. A comparison between LLDA (solid line) and regular LDA (dotted line) is shown.

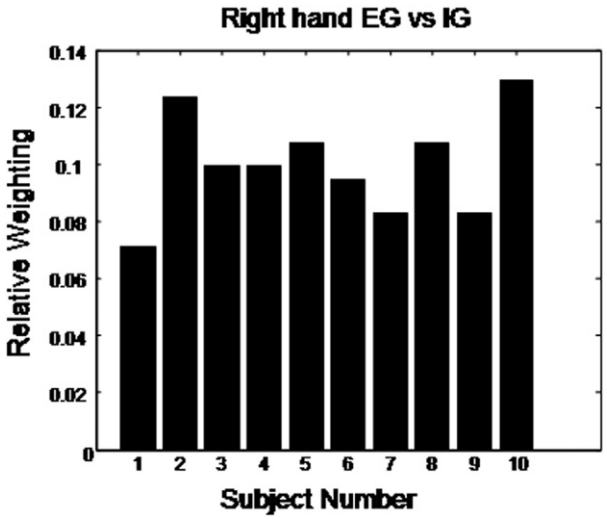

Fig. 5. Discriminability of overall discriminant function, $\tilde{\mathbf{U}}$, across subjects.

Since we are interested in subcortical structures, some consideration must be given to ROIs that have few voxels. We suggest that this will result in discriminant functions that are overly conservative. For example, consider just two ROIs, ROI_{small} with relatively few voxels and ROI_{large} with many voxels. The effective oversampling of ROI_{small} to create a $2 \times n$ bootstrap matrix (n being the number of bootstrap samples, cf. Eq. (11)) will result in conservative estimates of the discriminant function, since the heavily oversampled ROI_{small} voxels will be distributed over the entire range of voxels in ROI_{large}.

There is growing recognition that warping of individual subjects' brains to a common space may cause particular problems with registration, especially with small subcortical structures (Nieto-Castanon et al., 2003; Thirion et al., 2006). Despite widespread evidence that subcortical structures such as the thalamus are an integral part of the network used for motor control, only the LLDA approach, when applied to unwarped data, was able to detect significant activation differences between right- and left-handed task performance. We suggest that experimenters employing tasks expected to activate subcortical structures should use caution when warping individual subjects' brains to a common space.

Acknowledgments

This work was supported by a National Parkinson's Foundation Center of Excellence grant (MJM), NSERC Grant CHRPJ 323602-06 (MJM), CIHR grant CPN-80080 (MJM), NSERC grant RGPIN 312490-05 (ZJW), a Parkinson's fellowship grant from His Excellency Sheik Abdulmohsin Al-Sheikh (MML) and NIH grants AG21491 (XH) and RR00046 (GCRC).

Appendix A. Technical appendix

A.1. "Subspace" routine: incrementally calculate $\boldsymbol{u}_{kn}$

1. Initialize: $n=1$, $\mathbf{A}_{\boldsymbol{k}}=I_{N\times N}$.
   $\mathbf{A}_k$, a $N\times(N+1-n)$ matrix is the subspace which $\boldsymbol{u}_{kn}$ is projected to.
2. Call the "One-base" Routine to calculate $\boldsymbol{u'}_{kn}$.
   $\boldsymbol{u'}_{kn}$ a column vector of $N+1-n$ elements, is the projection of $\boldsymbol{u}_{kn}$ in $\mathbf{A}_{\boldsymbol{k}}$.
3. Convert to $\boldsymbol{u'}_{kn}$ to $\boldsymbol{u}_{kn}$: $\boldsymbol{u}_{kn}=\mathbf{A}_k\boldsymbol{u'}_{kn}$.
   $\boldsymbol{u}_{kn}$ is a column vector of N elements.
4. Create a subspace $\mathbf{S}_{\boldsymbol{k}}$ which is orthogonal to $\boldsymbol{u'}_{kn}$.
   $\mathbf{S}_k$ is a $(N+1-n)\times(N-n)$ matrix.
5. Prepare for the next iteration:
   $\mathbf{A}_k = \mathbf{A}_k\mathbf{S}_k$,
   $\mathbf{S} = \mathrm{diag}(\mathbf{S}_1 \ldots \mathbf{S}_K)$,
   $\mathbf{B} = \mathbf{S}^T\mathbf{B}\mathbf{S}$,
   $\mathbf{W} = \mathbf{S}^T\mathbf{W}\mathbf{S}$,
   $n = n + 1$.
6. If $n>N$, stop; else go back to step 2.

A.2. "One-base" routine: solve the one-base LLDA problem

The non-linear programming with the quadratic constraint is solved with a sequential quadratic programming (SQP) method implemented in Matlab function *fmincon*. The start point $\mathbf{U}_k^{\mathrm{int}}$ is initialized by solving a generalized eigenvector problem as shown below.

1) $\mathbf{B}^*=\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{B}_{ij}$, $\mathbf{W}^*=\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{W}_{ij}$
2) $\mathbf{B}^*\boldsymbol{u}_{\mathrm{int}}=\lambda\mathbf{W}^*\boldsymbol{u}_{\mathrm{int}}$ where λ is the largest generalized eigenvalue.
3) $\mathbf{U}_k^{\mathrm{int}}=\boldsymbol{u}_{\mathrm{int}}$

A.3. Joint optimization of mean $\mathbf{U}$ *for the two-class problem*

The ultimate goal of the optimization is to maximize the discriminant function of y^{mean}, i.e., the weighted y. As $y^{\mathrm{mean}} = \alpha_k y = \alpha_k\mathbf{U}_k^T(\boldsymbol{x}-\boldsymbol{m}_{\cdot k})$ the discriminant function of y^{mean} is

$$J = \log\frac{|\mathbf{B}^*|}{|\mathbf{W}^*|},$$

where

$$\mathbf{B}^* = [\alpha_1\mathbf{U}_1^T \ldots \alpha_K\mathbf{U}_K^T]^*\begin{bmatrix}\mathbf{B}_{11} & \ldots & \mathbf{B}_{1K}\\ \vdots & \ldots & \vdots\\ \mathbf{B}_{K1} & \ldots & \mathbf{B}_{KK}\end{bmatrix}^*\begin{bmatrix}\alpha_1\mathbf{U}_1\\ \vdots\\ \alpha_K\mathbf{U}_K\end{bmatrix} = \mathbf{U}^{*T}\mathbf{B}\mathbf{U}^*$$

$$\mathbf{W}^* = [\alpha_1\mathbf{U}_1^T K\alpha_K\mathbf{U}_K^T]^*\begin{bmatrix}\mathbf{W}_{11} & \Lambda & \mathbf{W}_{1K}\\ M & \mathrm{O} & M\\ \mathbf{W}_{K1} & \Lambda & \mathbf{W}_{KK}\end{bmatrix}^*\begin{bmatrix}\alpha_1\mathbf{U}_1\\ M\\ \alpha_K\mathbf{U}_K\end{bmatrix} = \mathbf{U}^{*T}\mathbf{B}\mathbf{U}^*$$

In the case of a two-class problems, $\mathbf{U}^*$ is a column vector composed of several column vectors $\alpha_i\mathbf{U}_i^T$. Though $\mathbf{U}_i$ and α_i are subject to the constraints $|\mathbf{U}_i|=1$, $\sum\alpha_1=1$ and $\alpha_i\geq 0$, the optimization can be solved without the constraints by the following. First, we solve

$$\max_{\mathbf{U}^*} J = \log\frac{|\mathbf{B}^*|}{|\mathbf{W}^*|} = \log\frac{|\mathbf{U}_*{}^T\mathbf{B}\mathbf{U}^*|}{|\mathbf{U}^{*T}\mathbf{W}\mathbf{U}^*|}$$

$$\mathbf{U}^* = \begin{bmatrix}\mathbf{U}_1{}^*\\ \vdots\\ \mathbf{U}_K\end{bmatrix}$$

without the constraints by calculating the generalized eigenvectors of $\mathbf{B}$ and $\mathbf{W}$. Then, we normalize each $\mathbf{U}_i{}^*$ to $\mathbf{U}_{\boldsymbol{i}}$:

$$\boldsymbol{b}_i = |\mathbf{U}_i{}^*|$$

$$\mathbf{U}_i = \mathbf{U}_i{}^*/\boldsymbol{b}_i$$

Finally, we derive the weight α_i from $\boldsymbol{b}_i$:

$$\alpha_i = \frac{\boldsymbol{b}_i}{\sum\boldsymbol{b}_i}$$

Because $\mathbf{U}_i$ and α_i solved in this way satisfy the constraints and they optimize the unconstrained problem whose solution is not worse than that of the constrained problem, they are also the solution of the constrained problem.

References

Beckmann, C.F., Jenkinson, M., et al., 1998. General multilevel linear modeling for group analysis in FMRI. NeuroImage 20 (2), 1052–1063.
Calhoun, V., Adali, T., et al., 2003. ICA of Functional MRI Data: An Overview. Proc. 4th Int. Symp. ICA2003.
Collins, D.L., Zijdenbos, A.P., et al., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imag. 17 (3), 463–468.
Damasio, H., 2005. Human Brain Anatomy in Computerized Images.
Flury, B., 1997. A First Course in Multivariate Statistics. Springer-Verlag, New York.
Friston, K.J., 1996. Statistical parametric mapping and other analyses of functional imaging data. In: Toga, A.W., Mazziotta, J.C. (Eds.), Brain Mapping, The Methods. Academic Press, San Diego, pp. 363–396.
Friston, K.J., Glaser, D.E., et al., 2002a. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16 (2), 484–512.
Friston, K.J., Penny, W., et al., 2002b. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16 (2), 465–483.
Kherif, F., Poline, J.B., et al., 2003. Group analysis in functional neuroimaging: selecting subjects using similarity measures. NeuroImage 20 (4), 2197–2208.
Kim, T.-K., Kittler, J., 2005. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. IEEE Trans. Pattern Anal. Mach. Intell. 27 (3), 318–327.
Liao, R., Krolik, J.L., et al., 2005. An information-theoretic criterion for intrasubject alignment of FMRI time series: motion corrected independent component analysis. IEEE Trans. Med. Imag. 24 (1), 29–44.
Liao, R., McKeown, M.J., et al., 2006. Isolation and minimization of head motion-induced signal variations in fMRI data using independent component analysis. Magn. Reson. Med. 55 (6), 1396–1413.

Maldjian, J.A., Laurienti, P.J., et al., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage 19 (3), 1233–1239.

McKeown, M.J., Hanlon, C.A., 2004. A post-processing/region of interest (ROI) method for discriminating patterns of activity in statistical maps of fMRI data. J. Neurosci. Methods 135, 137–147.

McKeown, M.J., Makeig, S., et al., 1998. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 6 (3), 160–188.

Mumford, J.A., Nichols, T., 2006. Modeling and inference of multisubject fMRI data. IEEE Eng. Med. Biol. Mag. 25 (2), 42–51.

Nieto-Castanon, A., Ghosh, S.S., et al., 2003. Region of interest based analysis of functional imaging data. NeuroImage 19 (4), 1303–1316.

Otani, S., 2003. Prefrontal cortex function, quasi-physiological stimuli, and synaptic plasticity. J. Physiol. (Paris) 97 (4–6), 423–430.

Politis, D.N., 2003. The impact of bootstrap methods on time series analysis. Stat. Sci. 18, 219–230.

Press, W.H., Teukolsky, S.A., et al., 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge Univ. Press, Cambridge.

Simpson, E.H., 1951. The interpretation of interaction in contingency tables. J. R. Stat. Soc., Ser. B 13, 238–241.

Strick, P.L., Dum, R.P., et al., 1998. Motor areas on the medial wall of the hemisphere. Novartis Found. Symp. 218, 64–75 (discussion 75–80, 104–8).

Talairach, J., Tournoux, M., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers Inc., New York.

Thirion, B., Roche, A., 2006. Improving sensitivity and reliability of fMRI group studies through high level combination of individual subjects results. MMBIA2006, New York.

Worsley, K.J., Liao, C.H., et al., 2002. A general statistical analysis for fMRI data. NeuroImage 15 (1), 1–15.