NeuroImage 109 (2015) 388–401

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Characterizing nonlinear relationships in functional imaging data using eigenspace maximal information canonical correlation analysis (*emi*CCA)

Li Dong [a], Yangsong Zhang [a], Rui Zhang [a], Xingxing Zhang [a], Diankun Gong [a], Pedro A. Valdes-Sosa [a,b], Peng Xu [a], Cheng Luo [a], Dezhong Yao [a,*]

[a] The Key Laboratory for NeuroInformation of Ministry of Education, Center for Information in BioMedicine, High-Field Magnetic Resonance Brain Imaging Key Laboratory of Sichuan Province, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China
[b] Cuban Neuroscience Center, Havana, Cuba

ARTICLE INFO

Article history:
Accepted 1 January 2015
Available online 12 January 2015

Keywords:
Eigenspace maximal information canonical correlation analysis (*emi*CCA)
Functional magnetic resonance imaging data analysis
Nonlinearity
Unsupervised
Motor execution

ABSTRACT

Many important problems in the analysis of neuroimages can be formulated as discovering the relationship between two sets of variables, a task for which linear techniques such as canonical correlation analysis (CCA) have been commonly used. However, to further explore potential nonlinear processes that might co-exist with linear ones in brain function, a more flexible method is required. Here, we propose a new unsupervised and data-driven method, termed the eigenspace maximal information canonical correlation analysis (*emi*CCA), which is capable of automatically capturing the linear and/or nonlinear relationships between various data sets. A simulation confirmed the superior performance of *emi*CCA in comparison with linear CCA and kernel CCA (a nonlinear version of CCA). An *emi*CCA framework for functional magnetic resonance imaging (fMRI) data processing was designed and applied to data from a real motor execution fMRI experiment.
This analysis uncovered one linear (in primary motor cortex) and a few nonlinear networks (e.g., in the supplementary motor area, bilateral insula, and cerebellum). This suggests that these various task-related brain areas are part of networks that also contribute to the execution of movements of the hand. These results suggest that *emi*CCA is a promising technique for exploring various data.

Introduction

Many important problems in the analysis of neuroimages actually boil down to discover the relationship between two sets of vector variables, X and Y. As an example (see Fig. 1), let us suppose we wish to study the relations between the X matrix comprising the time courses of a spatial independent component analysis (spatial ICA) of task-related fMRI data, and the Y matrix of imposed or observed time courses of relevant variables (the design matrix). More formally, we are considering the two following sets: $X = X' + \boldsymbol{\varepsilon}_x \in R^{n \times m}$ (i.e., time courses of independent components) and $Y = Y' + \boldsymbol{\varepsilon}_y \in R^{n \times q}$ (i.e., fMRI design matrix), where n is the number of time points, m is the number of the ICA time courses, q is the number of imposed or observed time courses related to stimulus, and $\boldsymbol{\varepsilon}$ is a random noise term reflecting our imperfect knowledge of the underlying processes. This type of question is most general since the definitions of X and Y are quite flexible. X could also be the blood oxygenation level-dependent (BOLD) signals in regions of interest (ROI), or even at the voxel level, or other fMRI features, or for this matter, any type of dynamical neuroimages [such as near-infrared spectroscopy (NIRS), electroencephalogram (EEG), and magnetoencephalogram (MEG)]. On the other hand, the matrix Y could be also other types of time varying parameters or even other types of neuroimages.

The problem of comparing two sets of variables was tackled early on by Hotelling, who developed the well-known canonical correlation analysis (CCA) in 1936 (Hotelling, 1936). His paper still stands as a key reference in multivariate statistical literature. CCA (Hotelling, 1936) finds the two vectors of relative weights **a** and **b** such that the linear combinations $X * \mathbf{a}$ and $Y * \mathbf{b}$ maximize the pairwise correlations across the two data sets, that is, the following problem:

$$\begin{aligned} &\underset{\boldsymbol{a},\boldsymbol{b}}{\text{maximize}} \quad C_{X*\mathbf{a},Y*\mathbf{b}} \\ &\text{subject to} \quad \operatorname{var}(X*\mathbf{a}) = \operatorname{var}(Y*\mathbf{b}) = 1 \end{aligned} \tag{1}$$

where $\operatorname{var}(\cdot)$ is the variance and $C_{(\cdot,\cdot)}$ is the Pearson correlation. Furthermore, we can define correlation matrix as follows:

$$C = \begin{pmatrix} C_{X,X} & C_{X,Y} \\ C_{Y,X} & C_{Y,Y} \end{pmatrix} \tag{2}$$

* Corresponding author. Fax: +86 28 83208238.
E-mail address: dyao@uestc.edu.cn (D. Yao).

http://dx.doi.org/10.1016/j.neuroimage.2015.01.006
1053-8119/

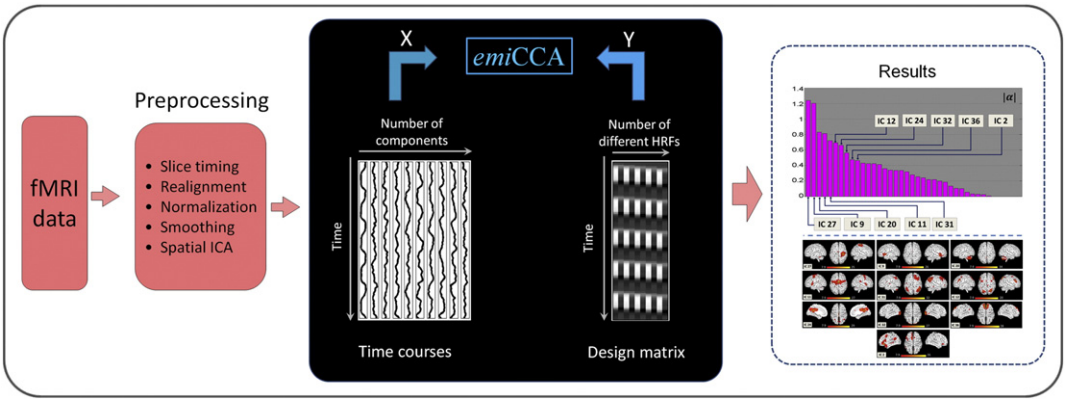

Fig. 1. The framework of fMRI data analysis using *emi*CCA. As an example, let us suppose we wish to study the relations between the X and Y in the analysis of neuroimages. In the *emi*CCA (black background), data set X (time points × number of components) was defined by mean time courses of independent components (ICs) over subjects; data set Y (time points × number of different HRFs) was defined by a design matrix. It is worth noting that the definitions of data sets can be very flexible in *emi*CCA.

where $C_{(\cdot,\cdot)}$ is the correlation between or within data sets. The problem can be solved with a generalized eigenvalue solution of the following form:

$$\begin{cases} C_{X,X}^{-1}C_{X,Y}C_{Y,Y}^{-1}C_{Y,X}\mathbf{a} = \lambda^2\mathbf{a} \\ C_{Y,Y}^{-1}C_{Y,X}C_{X,X}^{-1}C_{X,Y}\mathbf{b} = \lambda^2\mathbf{b} \end{cases} \tag{3}$$

where λ is the canonical correlation coefficient (square root of eigenvalue), and **a** and **b** are canonical coefficients vectors (eigenvectors). This method is now known as linear CCA (*l*CCA). It is not surprising that this well-known statistical method was applied early on to reveal homogeneous brain activity in fMRI data space (Friman et al., 2001). In this paper, X was defined as BOLD signals of a group of neighboring voxels, and Y was defined as the Fourier series. Linear CCA has also been used to achieve feature-based fusion of biomedical imaging modalities (e.g., EEG and fMRI) by determining inter-subject covariations across modalities (Correa et al., 2010; Correa et al., 2008). Furthermore, *l*CCA has been used in stable state visual evoked potential (SSVEP) detection (Zhang et al., 2012). It was also applied to investigate a group of neighboring voxels in fMRI data while adding the spatial constraints. X was defined as time series of neighboring voxels, and Y was defined as the design matrix (Cordes et al., 2012).

However, *l*CCA is based on the analysis of correlation matrices. It is well known that the correlation coefficient excels in detecting linear relationships but may fail in the case of nonlinear ones. However, nonlinearities have been convincingly demonstrated to originate from neural and/or vascular sources (Birn and Bandettini, 2005; Zhang et al., 2008). To complicate matters further, other factors such as methods for quantifying signals may introduce further nonlinearities (He et al., 2011). This argues that it may be necessary to go beyond the usual models, such as the general linear model (Friston, 2007; Friston et al., 1995) and *l*CCA (Friman et al., 2001), to detect activation patterns in fMRI data in order to cope with possible nonlinear effects, in particularly, the relationships between tasks and BOLD/neural activation.

A nonlinear generalization of CCA is the kernel version of CCA or kernel CCA (*k*CCA), which is obtained in a similar fashion as CCA but from the eigenspace corresponding to the kernel matrix. For *k*CCA (Akaho, 2006; Fukumizu et al., 2007), transform functions *f* and *g* project the data X and Y into a higher dimensional feature spaces, that is,

$$\begin{cases} f : X \in R^m \mapsto f(X) \in H_x^L \\ g : Y \in R^q \mapsto g(Y) \in H_y^L \end{cases} \tag{4}$$

where *H* is a higher dimensional Hilbert space. Then the kernel function *K*, which is instead of the inner product of the data matrices *f*(X) and *g*(Y), can be defined as

$$(K_x)_{ij} = k_x\left(\mathrm{X}_i, \mathrm{X}_j\right), \left(K_y\right)_{ij} = k_y\left(\mathrm{Y}_i, \mathrm{Y}_j\right) \tag{5}$$

Finally, the kernel matrix in high-dimensional space is obtained as

$$K = \begin{pmatrix} K_X K_X & K_X K_Y \\ K_Y K_X & K_Y K_Y \end{pmatrix} \tag{6}$$

and Eq. (1) can also be transformed into the following generalized eigenvalue problem in kernel space,

$$\begin{cases} (K_X K_X)^{-1}(K_X K_Y)(K_Y K_Y)^{-1}(K_Y K_X)\mathbf{a}^* = \lambda^2\mathbf{a}^* \\ (K_Y K_Y)^{-1}(K_Y K_X)(K_X K_X)^{-1}(K_X K_Y)\mathbf{b}^* = \lambda^2\mathbf{b}^* \end{cases} \tag{7}$$

where λ is the canonical correlation coefficient (square root of eigenvalue), and $\mathbf{a}^*$ and $\mathbf{b}^*$ are canonical coefficients vectors in kernel space. Kernel CCA has been to detect the nonlinearity in neural imaging (Biessmann et al., 2010; Hardoon et al., 2007). Nevertheless, *k*CCA confronts several practical problems in its use: (a) the practical choice of the regularization coefficient and kernel function (Fukumizu et al., 2007) and (b) the direct interpretation of the *k*CCA weights is also difficult since they are estimated in high-dimensional space and not in the original data space.

Our goal in this study has been to develop a new unsupervised and data-driven method for exploring associations between various data sets in the original data space, in which unknown linear and nonlinear relationships might exist simultaneously. The crucial point is the substitution of the linear correlation coefficients in Eq. (2) by a generalized measure of association based on mutual information, the maximal information coefficient (MIC). The MIC, which builds on entropy and the mutual information of random variables and has good properties of generality and equitability, is a more robust and appropriate measure to explore with various and complex signals (Reshef et al., 2013; Reshef et al., 2011; Speed, 2011). We take advantage of these properties by utilizing the eigenvectors and eigenvalues from the eigenspaces of the MIC matrix as a new measure for assessing the relationships between the two original data sets—the eigenspace maximal information canonical correlation analysis (*emi*CCA).

This paper is organized as follows. The next section describes in more detail our new method. We then go on to demonstrate the face validity of the *emi*CCA by means of simulations that contained various functional relationships. Meanwhile, a framework for the processing of real fMRI data is proposed to illustrate the application to both block

design and event-related fMRI motor execution experiments. Finally, discussions are provided regarding the resultant brain functional activations and the performance of the method.

Methods and materials

Eigenspace maximal information canonical correlation analysis

Here, we formulate the new method that is based on MIC (Reshef et al., 2011) to capture the linear and nonlinear relationships that exist between two data sets. This method is called eigenspace maximal information canonical correlation analysis.

For the same data sets, $X \in R^{n \times m}$ and $Y \in R^{n \times q}$, assuming that an unknown and nowhere-constant function f exists such that

$$Y'_j = f_{ij}(X'_i); i \in D_x \subseteq \{1, \ldots, m\}; j \in D_y \subseteq \{1, \ldots, q\} \tag{8}$$

then we have

$$\phi\left(X_i; Y_j\right) > 0; i \in D_x; j \in D_y \tag{9}$$

where ϕ is the MIC that falls between 0 and 1 (Reshef et al., 2011). Between pairs of variables X_i and Y_j, the MIC can be calculated from

$$\phi\left(X_i; Y_j\right) = \max_{|\mathbb{X}_i||\mathbb{Y}_j| < B} \left\{ \frac{I^*\left(\mathbb{X}_i; \mathbb{Y}_j\right)}{log_2\left\{min\left\{|\mathbb{X}_i|, |\mathbb{Y}_j|\right\}\right\}} \right\}; i \in \{1, \ldots, m\}; j \in \{1, \ldots, q\} \tag{10}$$

where m and q are numbers of data dimensions, $\mathbb{X}_i$ and $\mathbb{Y}_j$ are bins of a rectangular grid on the X_i–Y_j scatter plot, $I^*(.;.)$ is the maximal mutual information achieved by any grid on the data, and $|\mathbb{X}_i||\mathbb{Y}_j| < B$ indicates that the total number of bins is less than some number B. For these two data sets X and Y, similarly, we may have an MIC matrix (instead of correlation matrix and kernel matrix, corresponding to Eqs. (2) and (6), respectively),

$$\Phi = \begin{pmatrix} \Phi_{X,X} & \Phi_{X,Y} \\ \Phi_{Y,X} & \Phi_{Y,Y} \end{pmatrix} \tag{11}$$

where $\Phi_{(\cdot,\cdot)}$ is the MIC matrix between or within data sets calculated from Eq. (10). Then we assume that there are eigenvectors corresponding to the eigenspaces of X and Y ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively) that satisfy

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \quad \boldsymbol{\alpha}^{\mathrm{T}} \Phi_{X,Y} \boldsymbol{\beta} \tag{12}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are also vectors of relative weights that correspond to the eigenspaces of X and Y, respectively. To reduce the freedom of the scaling of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we may further add the following constraint:

$$\mathrm{var}\left(\sum_{i=1}^{m} \boldsymbol{\alpha}_i X_i\right) = \mathrm{var}\left(\sum_{i=1}^{q} \boldsymbol{\beta}_i Y_i\right) = 1 \tag{13}$$

where $\mathrm{var}(\cdot)$ is the variance. Eqs. (12) and (13) are also posed as a constrained optimization problem that uses Lagrange multipliers and is solved with a generalized eigenvalue solution in which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the eigenvectors (corresponding eigenvalues) of the following two matrices, respectively,

$$\begin{cases} \Phi_{X,X}^{-1} \Phi_{X,Y} \Phi_{Y,Y}^{-1} \Phi_{Y,X} \\ \Phi_{Y,Y}^{-1} \Phi_{Y,X} \Phi_{X,X}^{-1} \Phi_{X,Y} \end{cases} \tag{14}$$

Then the vectors of the weights ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; eigenvectors) that represent the weightiness of each piece of dimensional data in the cross MIC matrix and the maximal information eigen coefficient (MIEC, the square root of eigenvalues), which quantifies the relationships between the two data sets, is obtained simultaneously (a proof is provided in Appendix A). Additionally, because MIC is symmetric (e.g., $\phi(x, y) = \phi(y, x)$) (Reshef et al., 2011), *emi*CCA is also symmetric, that is, *emi*CCA(X, Y) = *emi*CCA(Y, X). Furthermore, it is worth noting that, although *emi*CCA can be relative with the *l*CCA conceptually in an unknown space X^* and Y^* in which relationships are linearized, *emi*CCA is a different method instead of a different version of *k*CCA (a proof is provided in Appendix B).

MIC assigns scores that approach 0 to statistically independent variables and depends only on the rank order of the data (Reshef et al., 2011). Therefore, the significance of a given MIC value was established by comparing the MIC scores of the random data. Considering MIC is a measure based on mutual information, the significance of MIC values in Φ were controlled by the false discovery rate (FDR) method, and the non-significant MIC values were rejected before solving the above-mentioned problem (Eqs. (12) and (13)). The p-value of each MIEC was obtained by random permutation tests of the MICs between the weighted sums of X and Y (a proof is provided in Appendix C). Here, 5000 random permutations were employed to generate the MIC null distribution.

Simulation

With $n = 500$ samples in $m = 6$ and $q = 4$ dimensions, where one relevant variable (the first pair, x_1 and y_1) or several relevant variables were present in random variables X and Y (standard uniform distributions), the linear or nonlinear relationships between the variates were set in our simulations. These functional relationships were linear, quadratic, and cosinusoidal, in addition to other types. (Table 1). Independent Gaussian noises with the standard deviation of 0.05 (Akaho, 2006) were added in test samples, and various data dimensions (m, q) were set at (3, 2), (6, 4), (12, 8), (18, 12), (24, 16), and (30, 20) to assess the effects of the dimensionality to *emi*CCA. Furthermore, to assess the effects of sample size and noise, various data sizes (from 60 to 1600) and noises with different standard deviations (from 0 to 0.5, based on noise levels in previous papers) (Akaho, 2006; Balakrishnan et al., 2012) were considered and set in the simulation. The FDR was set at 0.05 to reject non-significant MICs in the MIC matrix, and the maximal MIEC (and weights) of the results was extracted and averaged over 20 repeats of the above-mentioned experiments. To assess the performance of *emi*CCA (*emi*CCA toolbox will be available on http://www.neuro.uestc.edu.cn/*emi*CCA.html with the paper), *l*CCA and *k*CCA (*k*CCA toolbox is available on http://sourceforge.net/projects/kmbox/) were also conducted on these experiments, and the results were compared with the results from the *emi*CCA. The parameters of *k*CCA were fixed and as follows: kernel type was Gaussian kernel, regularization was set at 10^{-5}, and kernel parameter value was set at 1. In addition, considering that fMRI time courses were autocorrelated, the possible impacts of autocorrelation on performances of *emi*CCA, *l*CCA, and *k*CCA were also considered in the simulation (more details can be seen in supplementary material A).

Table 1
Definitions of the functional relationships.

Relationship name	Description (domain is [−1 1])
Linear	$y_1 = x_1$
Parabolic	$y_1 = x_1^2$
Absolute	$y_1 = \lvert x_1 \rvert$
Cosinusoidal	$y_1 = \cos(2\pi x_1)$
Exponential	$y_1 = e^{,\sin(2\pi x_1)}$
Circle	$y_1^2 + x_1^2 = 1$
Nonseparable	$y_1 = x_1 x_2$
Equations	$\begin{cases} y_1 = x_1^2 \\ y_2 = ,\cos(2\pi x_1) \end{cases}$

Real fMRI data

To illustrate the performance of the *emi*CCA with real data, an implementation framework based on *emi*CCA was designed to analyze fMRI data that were gathered during a motor execution paradigm (Fig. 1). Twenty-six healthy volunteers participated in the experiment (17 males and 9 females; mean age = 23 years; standard deviation = 2 years; age range = 19-26 years) after giving written informed consent. An one-task (left hand movement) block design was used in the experiment. In each block, a yellow cross appeared on the center of the screen for 2 s, and the subjects were asked to prepare for the following hand movement during this period. Next, the color of the cross turned to white, and a white left arrow appeared on the screen simultaneously (lasted for 20 s). The subjects were asked to perform the left hand movement during this time; when the left arrow disappeared, the subjects were able to rest but were asked to focus their attention on the white cross. Twenty blocks were scanned for each subject, and the type of left hand movement used in this experiment was moving the palm up and down, such as tapping a ball. The study was approved by the Ethics Committee of the University of Electronic Science and Technology of China.

1) fMRI data acquisition: Data were recorded using a 3.0 T scanner (GE Discovery MR750, USA) at the University of Electronic Science and Technology of China. Axial anatomical T1-weighted images were collected with a 3-dimensional fast spoiled gradient echo sequence (TR/TE = 6.008 ms/1.984 ms, flip angle = 9°, matrix size = 256 × 256, field of view = 25.6 × 25.6 cm^2, slice thickness (no gap) = 1 mm). Functional images were collected with an echo planar sequence (TR/TE = 2000 ms/30 ms, flip angle = 90°, matrix size = 64 × 64, field of view = 24 × 24 cm^2, thickness/gap = 4 mm/0.4 mm), and a total of 405 volumes (32 slices per volumes) were also obtained over an 810 s period.
2) Preprocessing: The first five volumes were first discarded to remove the T1 saturation effects. Slice time correction, realignment, spatial normalization (3 mm × 3 mm × 3 mm), and spatial smoothing (8-mm full-width at half-maximum of an isotropic Gaussian filter) were analyzed using SPM8. Next, to decrease the noise level of the fMRI data and to increase calculation efficiency, group spatial ICA (infomax algorithm) was applied to extract the spatiotemporal features of the fMRI data (Calhoun et al., 2001b). The optimal number of independent components was estimated at 38 based on the minimum description length criteria (Li et al., 2007), and four components associated with the possible artifacts (such as head motion, cerebrospinal fluid, large vessels, and dispersion of clusters) were visually inspected, discarded, and rejected from further investigation.
3) *emi*CCA analysis: Data set X was defined by mean time courses of independent components (ICs) over subjects (m = 34 dimensions). Data set Y was defined by a design matrix in the fMRI data contained the onset times of the stimuli convolved with HRFs of different onset times (−2 s, −1 s, 0 s, 1 s, and 2 s) and their derivatives (finally generating q = 10 dimensions). Then *emi*CCA was applied (The FDR was set at 0.01 to reject non-significant MICs in MIC matrix), and the results corresponding to the maximal MIECs were obtained.
4) *emi*CCA assessing: Additionally, to assess the performance of *emi*CCA, *l*CCA (also conducted on the above-mentioned data sets X and Y) and GLMs (stimulus onsets were convolved with canonical HRF, and the activation was subsequently estimated by SPM) were also investigated and compared with *emi*CCA.

Results

Simulation results

Table 2 shows the results (mean values) of the *emi*CCA that contains maximal information eigen coefficients (MIECs) and solutions for the weights (α and β). From linear, parabolic, absolute, cosinusoidal, exponential, circular, nonseparable, and complex (equations) relationships, the mean *emi*CCA MIECs were 0.9927, 0.9309, 0.9348, 0.8576, 0.9088, 0.5376, 0.3506, and 0.9586, respectively, and the p-values by permutation test for the pairs were all $\ll 2 \times 10^{-4}$. Fig. 2 shows the mean values that contain MIECs and the canonical coefficients of *l*CCA and *k*CCA. Significant MIECs were obtained for all given relationships (linear or nonlinear) but not for independent variables (random). Linear CCA produced significant canonical correlation coefficients for the linear relationship. For nonlinearity that can be approximately linearized to some extent (e.g., exponential relationship), significant coefficients of *l*CCA were also obtained. All canonical coefficients of *k*CCA approximated 1 (see Fig. 2), regardless of whether a relationship existed. By visually inspecting scatter plots of the associations between the canonical variate of *k*CCA and the variable of original data (see Fig. S5), *k*CCA extracted the correct features for linear, parabolic, absolute, circular, and nonseparable relationships, but not for cosinusoidal and exponential relationships. For complex relationships (equations), *k*CCA may have ignored several relationships, even though *k*CCA correctly extracted partial features. For independent data (random), *k*CCA produced false canonical correlation coefficients (also ~1). In addition, the MIECs were barely affected by increases in the dimensionality of the data for these relationships, whereas the canonical correlation coefficients of *l*CCA increased with increasing dimensions for nonlinear relationships (Fig. 3).

To assess the effects of the data sample size and noise, different data sizes and noise levels with different standard deviations were considered; the results of these simulations are shown in Fig. 4. For the linear relationship, *emi*CCA and *l*CCA had similar performances. For the nonlinear relationships, *emi*CCA performed better than *l*CCA. As shown in Fig. 4, the MIEC calculated by *emi*CCA decreased with increases in

Table 2
Results of *emi*CCA performed on data sets of sample size = 500 to which Gaussian noise with a standard deviation of 0.05 was added. R is the mean MIEC, and α and β are solutions for the weights.

	Linear	Parabolic	Absolute	Cosinusoidal	Exponential	Circular	Nonseparable	Equations
R	0.9927	0.9309	0.9348	0.8576	0.9088	0.5376	0.3506	0.9586
α	1.7128	1.7103	1.7143	1.7115	1.7161	1.4200	1.2814	1.7149
	0	0	0	0	0	0	1.0816	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
β	1.7112	3.2929	3.4172	1.4196	1.2033	1.3974	2.9500	2.0658
	0	0	0	0	0	0	0	0.8713
	0	0	0	0	0	0	0	-5.84×10^{-18}
	0	0	0	0	0	0.0268	0	0

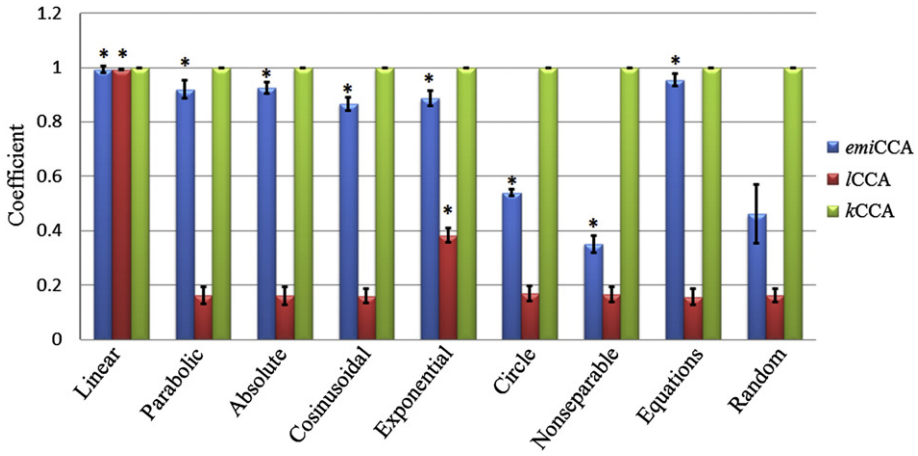

Fig. 2. The mean values (averaged over 20 repeats) that contain MIECs (*emi*CCA) and canonical coefficients of *l*CCA and *k*CCA. Results were obtained from data sets with a sample size = 500 to which Gaussian noise with a standard deviation of 0.05 was added. *Significant.

noise, and small sample sizes with high noise levels may lead to a plausible MIEC (not significant). Using *l*CCA, for the linear relationship, the canonical correlation coefficients also decreased with increasing noise. However, for all nonlinear relationships, the *l*CCA coefficients markedly decreased with increases in data size.

Real data results

The real fMRI data were gathered during a task (motor execution paradigm) that involved moving the palm up and down using the left hand. The analyses of these data using *emi*CCA, *l*CCA, and GLMs (SPM)

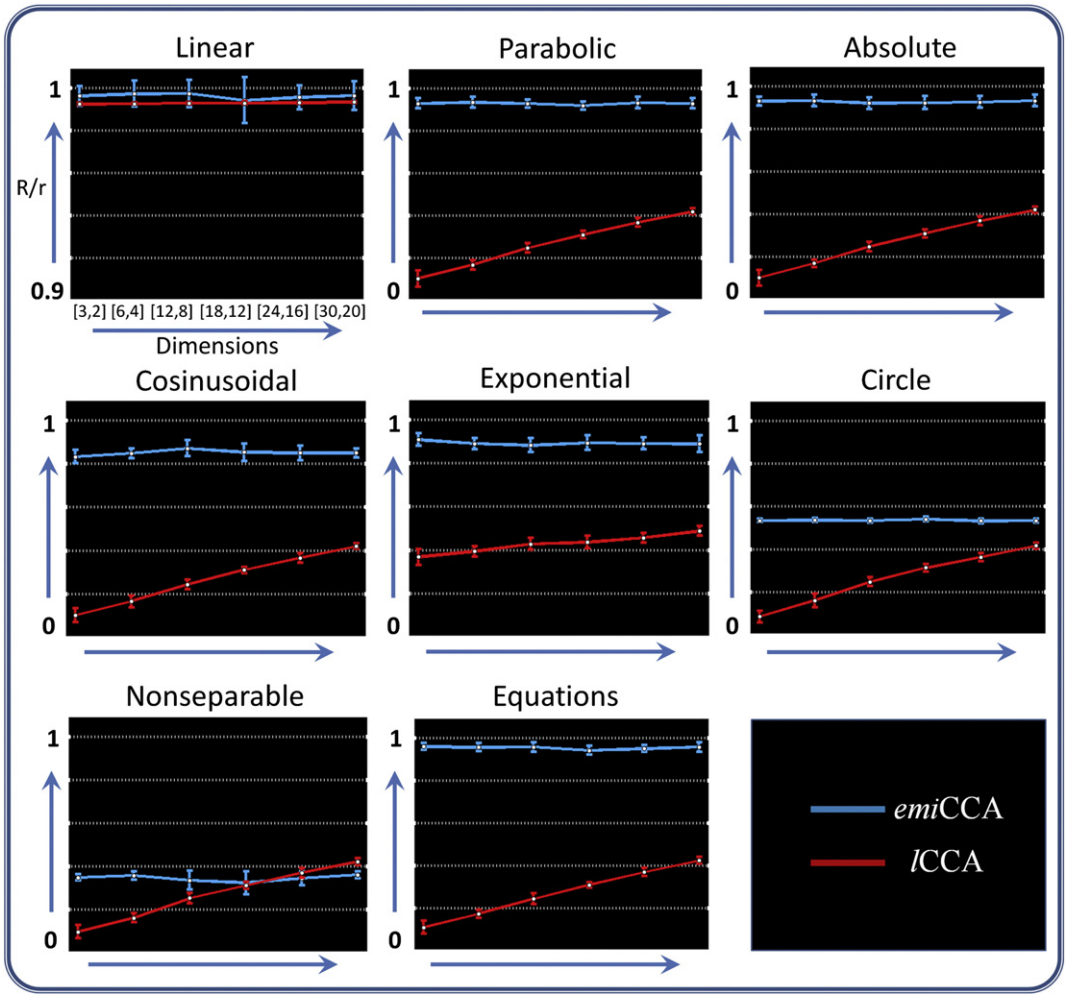

Fig. 3. Mean results (averaged over 20 repeats) of *emi*CCA and *l*CCA in various dimensions with a data sample size set at 500 and Gaussian noises with a standard deviation 0.05 added. The parameters of data dimension (m, q) were set at (3, 2), (6, 4), (12, 8), (18, 12), (24, 16), and (30, 20). R/r is the mean MIEC/mean canonical correlation coefficient.

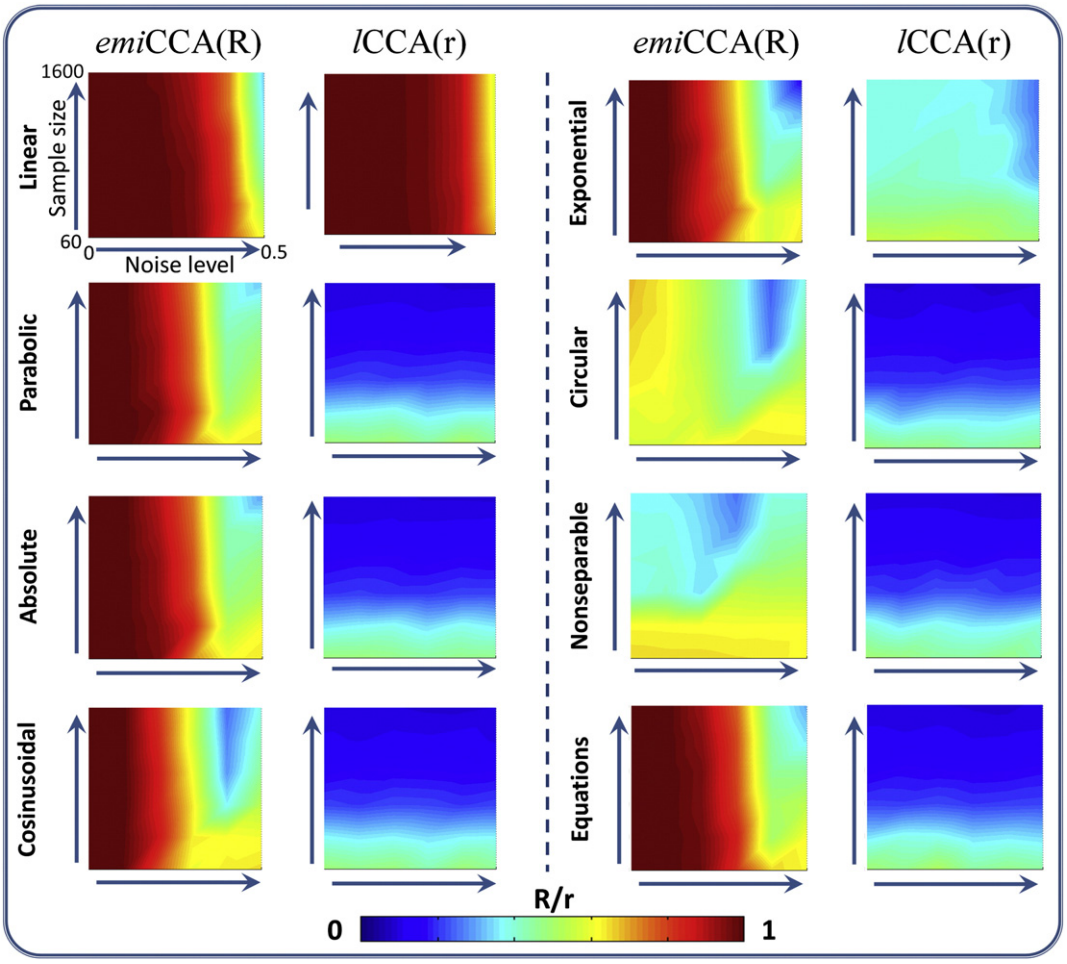

Fig. 4. Results of *emi*CCA and *l*CCA using data sets of various sample sizes and noise levels with different standard deviations. The first and third columns illustrate the mean MIECs (R) of the *emi*CCA, and the second and fourth columns illustrate the mean canonical correlation coefficients (r) of the *l*CCA.

are shown in Fig. 5. The weights of *k*CCA were estimated in high-dimensional space; therefore, *k*CCA was not suited to identify task-related ICs in our framework.

These results show that extremely significant relationships may have existed between the assumed brain activations and the fMRI measurement. The *p*-value yielded by a permutation test for the MIECs of *emi*CCA ($R = 0.8005$) was $\ll 2 \times 10^{-4}$, while the *p*-value of the canonical correlation coefficient (*l*CCA, $r = 0.9955$) was ~0. The solutions for the weights (α) corresponding to the time courses of all components were also obtained. For *emi*CCA, the first ten components (the remaining 24 components and the rejected 4 components related to possible artifacts are also showed in the Figs. S6 and S7, respectively) that corresponded to the larger weights of the absolute α values were present as follows (Fig. 5, cyan background and Table 3):

IC27: the spatial distributions consisting of right precentral gyrus [primary motor cortex, Brodmann area (BA) 4] and left anterior lobe.

IC9: the insula network primarily encompassed the bilateral insula and superior temporal gyrus (BA13/BA38).

IC20: the cerebellum network primarily encompassed the cerebellar tonsil, inferior semi-lunar lobule, and culmen.

IC11: the spatial distributions consisting of the supplementary motor area (BA6) and bilateral insula (BA13).

IC31: the right lateral frontoparietal network showed spatial patterns consisting of the inferior parietal lobule (BA39/BA40), superior parietal lobule (BA7), middle frontal gyrus (BA6/BA10), and inferior semi-lunar lobule.

IC12: the spatial patterns primarily encompassed the inferior frontal gyrus (BA9), middle frontal gyrus (BA9), superior parietal lobule (BA7), precentral gyrus (BA9), and pyramis.

IC24: the spatial patterns consisting of the posterior cingulate (BA30), precuneus (BA31), and bilateral angular gyrus (BA39).

IC32: the visual networks including the occipital lobe (BA18), lingual gyrus (BA17), and cuneus (BA18).

IC36: the spatial patterns primarily encompassed the medial frontal gyrus (BA8), superior frontal gyrus (BA8/BA9), and inferior semi-lunar lobule.

IC2: the spatial patterns consisting of the bilateral inferior frontal gyrus (BA47), left superior frontal gyrus (BA6/BA8), and left middle frontal gyrus (BA9).

Furthermore, for *l*CCA, only IC27 was identified. The spatial distributions of these analyses are shown at the bottom of the cyan area in Fig. 5. The results of the SPM analysis (Table 4) are also shown in the bottom area (black background) of Fig. 5, and the activated regions revealed by this analysis consisted of the right precentral gyrus and the left cerebellum, which was in accordance with the IC27 identified by the aforementioned methods. Additionally, to further illustrate that nonlinearity may exist in neural imaging, trend lines (referenced by polynomial curve fitting, the order = 1, 2, or 3) between the time courses of the

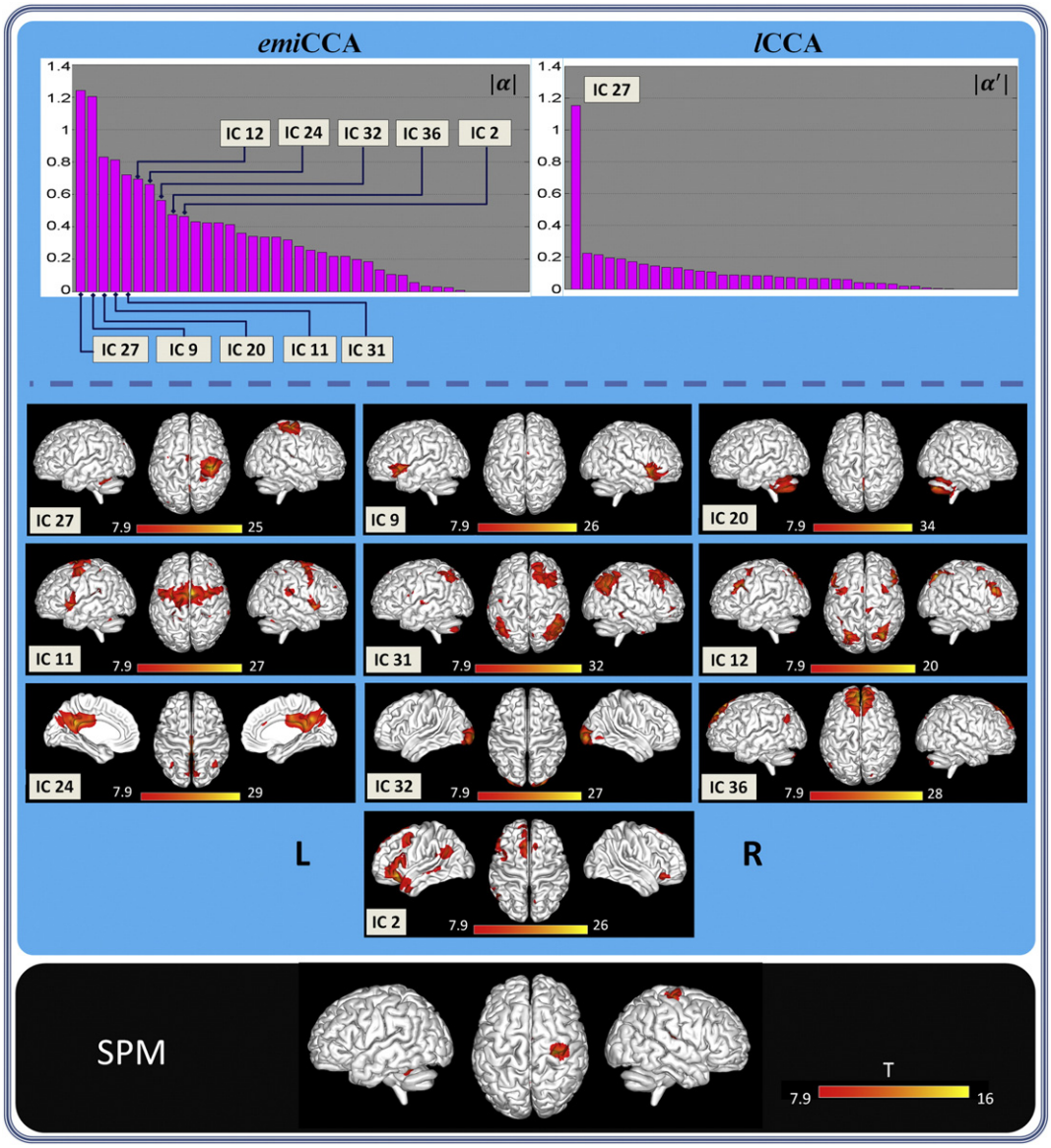

Fig. 5. Real data results. In the upper area of the region with a cyan background, the left panel represents the weights of absolute values estimated by *emi*CCA (|α|), and the right panel represents the weights of absolute values estimated by *l*CCA (|α′|). The T-maps of spatial ICs (family-wise error corrected, $p < 0.001$) corresponding to the larger absolute values of α (first ten) are shown in the bottom area of the region with a cyan background. In the bottom area with the black background, the regions of activation as estimated by GLMs (SPM) are shown, and the color bar represents the T-value (family-wise error corrected, $p < 0.001$). L: left; R: right.

above-mentioned ICs and the assumed BOLD responses that were stimulus onsets convolved with canonical HRF are also displayed (Fig. 6). The relationship between the time course of IC27 and the assumed BOLD response was obviously linear, while other relationships were parabolic or cubic (remaining 9 components).

Discussion

In this paper, we present a novel unsupervised method called *emi*CCA for fMRI data analysis. The simulation was implemented to demonstrate the performance of *emi*CCA, and fMRI data gathered during a motor execution paradigm were also analyzed in this work to illustrate the superiority of *emi*CCA.

Methodological considerations of emiCCA

To illustrate the performance of *emi*CCA, the data sample size (set at 500) and Gaussian noises (standard deviation = 0.05) were initially fixed to obtain the results. For the linear situation, *emi*CCA performed as well as *l*CCA in capturing the features between the two data sets. For the nonlinear situations, *emi*CCA also performed well in detecting the parabolic, absolute, cosinusoidal, exponential, and complex (equations) relations, even the circular and nonseparable relationships were to some extent captured (see Fig. 2 and Table 2). Furthermore, the MIECs of *emi*CCA were barely affected by increases in the dimensionality of the data for these relationships (Fig. 3). To further assess the properties of *emi*CCA, the effects of the data sample size and noise on the performance of *emi*CCA were investigated. As shown in Fig. 4, the MIECs of *emi*CCA generally decreased as the noise increased. For the situation of autoregressive process, *emi*CCA could also tolerate this situation to some degree and performed well in detecting relationships (more details can be seen in supplementary material A). The stability and reliability of the results benefit from large sample sizes and low noise levels in real data sets. Additional stimuli or runs during fMRI recording may, to some extent, increase the accuracy of the results. Moreover, the noise of the original fMRI data acquisition may be produced at the levels of local field potentials, hemodynamics and recording (Roebroeck

Table 3
One-sample *t*-test results of the ten components identified by *emi*CCA. The significance threshold was set at $p < 0.001$ (family-wise error corrected, cluster size > 23 adjacent voxels).

	MNI coordinates			L/R	Lobe	Brodmann area	*T*-value	Voxels
	x	*y*	*z*					
IC 27	39	−24	66	R	Precentral gyrus	BA 4	25.24	1262
	−3	−6	60	L	Medial frontal gyrus	BA 6	14.53	
	−9	−51	−24	L	Anterior lobe		19.96	406
IC 9	−45	12	−9	L	Insula/superior temporal gyrus	BA13/BA38	26.28	824
	−36	18	−6	L	Insula/inferior frontal gyrus	BA13/BA47	22.96	
	33	24	−12	R	Insula/inferior frontal gyrus	BA47	23.96	1165
	42	12	−9	R	Insula	BA13	22.69	
IC 20	33	−72	−36	R	Cerebellar tonsil		34.31	5265
	−15	−66	−45	L	Inferior semi-lunar lobule		27.87	
	24	−66	−27	R	Culmen		25.65	
IC 11	6	0	72	R	Medial frontal gyrus	BA6	27.03	4936
	6	12	48	R	Superior frontal gyrus	BA6	26.08	
	6	12	39	R	Cingulate gyrus	BA24	25.20	
	−42	6	0	L	Insula	BA13	22.61	670
	66	−33	21	R	Superior temporal gyrus	BA42	13.01	169
	36	45	27	R	Middle frontal gyrus	BA9	11.96	21
	−57	−39	21	L	Superior temporal gyrus	BA13	10.89	95
	−30	48	24	L	Superior frontal gyrus	BA9	10.18	25
IC 31	45	−60	39	R	Inferior parietal lobule	BA39/BA40	32.02	1190
	33	−69	48	R	Superior parietal lobule	BA7	18.54	
	42	21	48	R	Middle frontal gyrus	BA8/BA9	20.03	1146
	27	18	54	R	Superior frontal gyrus	BA8	19.82	
	−45	−63	48	L	Inferior parietal lobule	BA39/BA40	18.15	345
	−36	−75	45	L	Superior parietal lobule	BA7	11.41	
	−15	−6	57	L	Medial frontal gyrus	BA6	13.39	63
	36	51	3	R	Middle frontal gyrus	BA10	12.61	93
	6	−36	33	R	Cingulate gyrus	BA31/BA23	12.00	119
	−33	−72	−42	L	Inferior semi-lunar lobule		11.79	82
	−30	−63	−36	L	Cerebellar tonsil		9.34	
IC 12	45	15	30	R	Middle/inferior frontal gyrus	BA9	20.31	574
	48	−3	30	R	Precentral gyrus	BA6	12.64	
	33	−72	48	R	Superior parietal lobule	BA7	19.50	782
	−48	21	30	L	Middle frontal gyrus	BA9	19.46	543
	−45	12	33	L	Precentral gyrus	BA9/BA6	17.71	
	−27	−69	45	L	Superior parietal lobule	BA7	19.05	723
	−27	−81	30	L	Middle occipital gyrus	BA19	14.34	
	−9	−81	−30	L	Pyramis		14.70	117
	9	−87	−39	R	Inferior semi-lunar lobule		9.37	
	30	21	−9	R	Inferior frontal gyrus	BA47	11.90	60
	33	9	57	R	Middle frontal gyrus	BA6	11.46	119
	−3	6	72	L	Superior frontal gyrus	BA6	11.08	96
IC 24	−6	−48	27	L	Posterior cingulate/precuneus	BA30/BA31	29.60	2226
	6	−60	33	R	Cingulate gyrus	BA31	24.76	
	6	−27	27	R	Posterior cingulate	BA23	24.68	
	42	−66	48	R	Inferior parietal lobule	BA40	13.44	55
	−39	−63	39	L	Angular gyrus	BA39	10.63	79
IC 32	−21	−96	−6	L	Occipital lobe	BA18	27.44	674
	−18	−105	−6	L	Lingual gyrus	BA17	24.82	
	−12	−108	6	L	Cuneus	BA18	10.85	
	27	−99	−3	R	Inferior occipital gyrus	BA17	25.54	790
	42	−63	−6	R	Occipital lobe	BA37	9.05	
IC 36	3	48	48	R	Medial frontal gyrus	BA8	27.88	2108
	−6	51	42	L	Superior frontal gyrus	BA8/BA9	25.47	
	0	57	30		Medial frontal gyrus	BA9/BA10	22.33	
	−33	−84	−36	L	Inferior semi-lunar lobule		12.61	119
	−30	−87	−24	L	Uvula		9.43	
	33	−84	−36	R	Pyramis		11.41	181
	15	−90	−18	R	Declive		9.72	
	30	−15	48	R	Precentral gyrus	BA4	11.12	82
IC 2	−45	24	−9	L	Inferior frontal gyrus	BA47	26.02	1434
	−45	39	−3	L	Inferior frontal gyrus	BA46	16.75	
	−9	27	57	L	Superior frontal gyrus	BA6/BA8	17.36	869
	−6	54	30	L	Medial frontal gyrus	BA10	16.57	
	−39	9	45	L	Middle frontal gyrus	BA9	15.41	204
	−51	−54	24	L	Supramarginal gyrus	BA40	15.29	342
	45	33	−12	R	Inferior frontal gyrus	BA47	13.40	145

et al., 2005), and traditional preprocessing, such as spatial smoothing with a Gaussian kernel (Friston, 2007) can always be used to effectively suppress noise. Furthermore, principal component analysis (PCA) and ICA have also been widely applied to remove artifacts or extract data features and decrease the noise in neural imaging data sets (Calhoun et al., 2001b; Calhoun et al., 2009). In brief, to analyze real data with *emi*CCA, preprocessing steps seeking to remove artifacts are, to some extent, preferred to better capture the

Table 4
Results of SPM analysis ($p < 0.001$, family-wise error corrected, cluster size > 23 adjacent voxels).

MNI coordinates			L/R	Lobe	Brodmann area	*T*-value	Voxels
x	*y*	*z*					
33	−21	60	R	Precentral gyrus	BA4	16.06	324
−12	−51	−24	L	Anterior lobe		12.55	228
−21	−48	−24	L	Culmen		12.02	
54	−18	18	R	Postcentral gyrus	BA43	10.93	75

relationships, and sample sizes should be sufficiently large so that information regarding the potential relationships is effectively provided.

Because the MIC is an univariate method that can identify interesting relationships between pairs of variables in large data sets (Reshef et al., 2011), the elements of MIC matrix, Φ in Eq. (11), are calculated from all pairs of variables between or within data sets X and Y, respectively (a total of $(m+q)(m+q-1)/2$ MICs will be calculated). Therefore, the time cost of *emi*CCA may mainly depend on the MIC calculation. However, the time costs of *emi*CCA calculation are acceptable when the data dimensions and sample sizes are increasing, whereas the time cost of *k*CCA calculation mainly depends on the sample size of data set (Fig. S4). Reshef et al. (2013) have suggested that the modified parameters and improved approximation algorithm in calculation of MIC can gain a significant decrease in runtime without significant loss of MIC performance. Therefore, the calculation of *emi*CCA can benefit not only from the development of computer (e.g., parallel computing) but also from the development of MIC algorithm and modification in the future.

In addition, it is worth noting that *emi*CCA can be relative with the *l*CCA conceptually in an unknown space (a proof is provided in Appendix B); however, the objective of *emi*CCA is, to some extent, distinct from *l*CCA (Eqs. (1) and (12)). Because, considering two data sets X and Y, we have

$$\begin{cases} \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \ \boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} \neq \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}}\, \Phi_{\mathrm{X}*\boldsymbol{\alpha},\mathrm{Y}*\boldsymbol{\beta}} \\ \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \ \boldsymbol{\alpha}^{\mathrm{T}}\mathrm{C}_{\mathrm{X,Y}}\boldsymbol{\beta} = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}}\, \mathrm{C}_{\mathrm{X}*\boldsymbol{\alpha},\mathrm{Y}*\boldsymbol{\beta}} \end{cases} \quad (15)$$

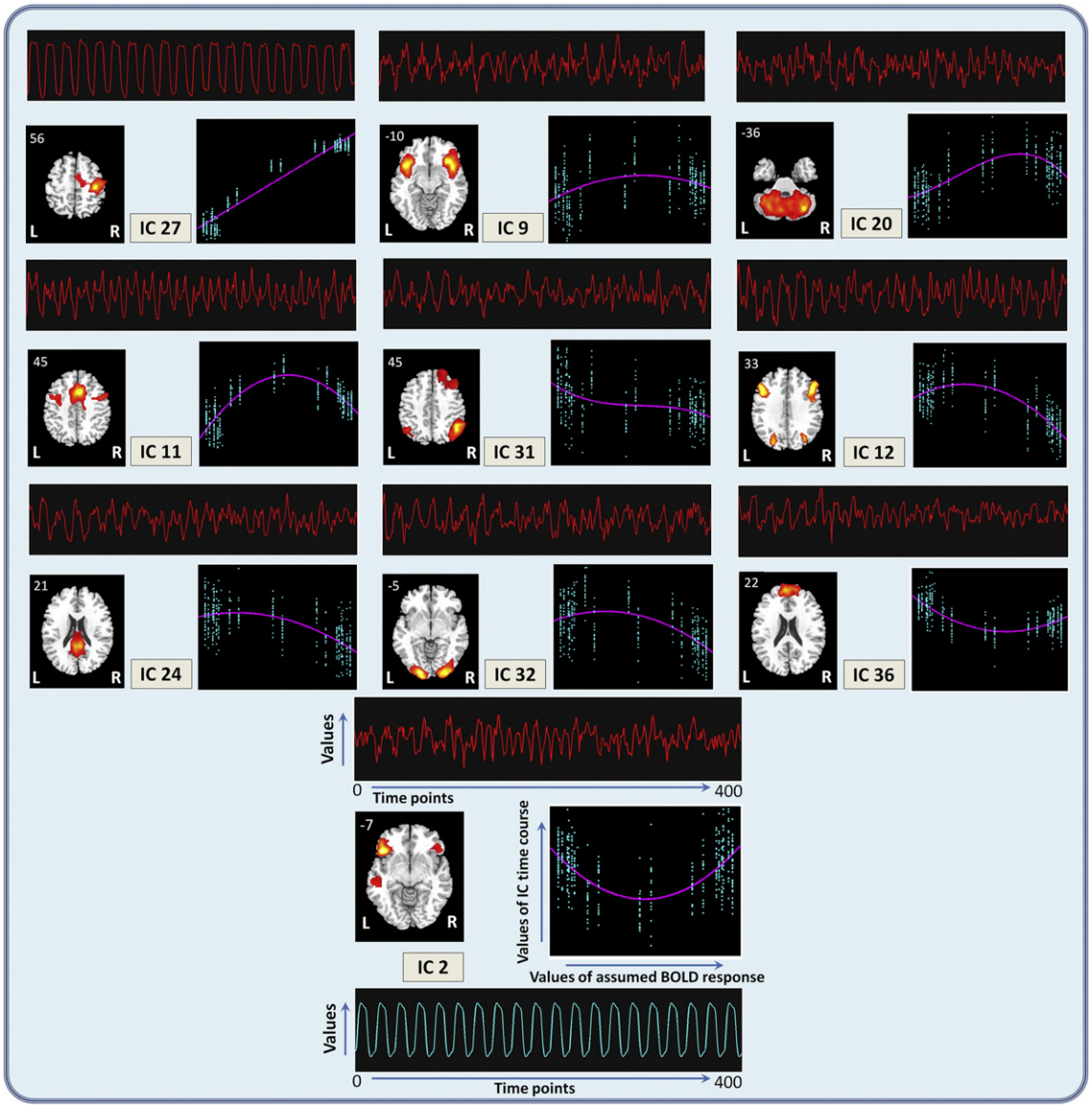

Fig. 6. Linear and nonlinear relationships between IC time courses (red lines) and the assumed BOLD response (the cyan line on the bottom of the figure; stimulus onsets convolved with canonical HRF). Time courses for the ten ICs (identified by *emi*CCA) and their spatial distributions in one slice are also shown. The polynomial functions (the order = 1, 2, or 3) were utilized to fit the trend lines (magenta lines) between the time courses and the assumed BOLD response, and they are displayed on the right of the corresponding spatial distributions. L: left; R: right.

where $\Phi_{(\cdot,\cdot)}$ is the MIC matrix, $C_{(\cdot,\cdot)}$ is the correlation matrix, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are weights in original data sets. Therefore, the involved assumption and computational implementation of *emi*CCA are to some degree different from *l*CCA.

lCCA vs. emiCCA

For the linear relationship, *l*CCA produced canonical correlation coefficients that were in accordance with the performance of *emi*CCA. However, for nonlinear relationships, the coefficients of the *l*CCA cannot assess these relationships well (see Figs. 2 and 4). The linear nature of *l*CCA (Hotelling, 1936) is limited in its potential to uncover more complex relationships. In Fig. 4, the coefficients of *l*CCA markedly decreased with increases in data size, which may have been caused by *l*CCA approximating linear situations when the information about the relationships is insufficient (i.e., when the sample size is too small). With increases in data sizes, the true relationships (e.g., nonlinear relationships) became more stable and obvious, and *l*CCA was not suitable for these situations. In addition, considering that the coefficients of *l*CCA increase with increasing dimensions for nonlinear relationships, CCA was not suitable for nonlinearity. Due to the generality of MIC in that a wide range of associations (linear or nonlinear) can be captured (Reshef et al., 2011), *emi*CCA would have better generality and performance than *l*CCA.

kCCA vs. emiCCA

As a nonlinear version of *l*CCA, *k*CCA extends *l*CCA using the kernel method to transform the original data into high-dimensional Hilbert space, such that their correlation is maximized (Akaho, 2006). Based on previous studies (Akaho, 2006; Fukumizu et al., 2007; Hardoon et al., 2004), we used a fixed kernel function, Gaussian kernel, in the *k*CCA. In our simulation, all canonical coefficients of *k*CCA approximated 1 (see Fig. 2). By visual inspection (see Fig. S5), the *k*CCA performance worked well for the linear, parabolic, absolute, circle, and nonseparable relationships yet failed in the cosinusoidal and exponential relationships. Furthermore, *k*CCA may have also ignored several complex (equations) relationships. One crucial problem of *k*CCA is that false canonical coefficients (also ~1) are obtained from random data where no relationships exist because the kernel method is likely to overfit data. Another problem is that it is difficult to distinguish the true from false (or plausible) results of *k*CCA. Several regularization techniques have been utilized to avoid the overfitting of *k*CCA (Akaho, 2006); however, *k*CCA also confronts the problems regarding how to choose the optimal regularization coefficient in practice and optimal kernel functions (Fukumizu et al., 2007). Meanwhile, the workspace is the transformed high-dimensional data space (Akaho, 2006); therefore, the weights estimated by *k*CCA did not directly represent the weightiness of each piece of original dimensional data and may have reduced its applicability. In brief, the above-mentioned factors may have limited the potential of *k*CCA for applications. The *emi*CCA directly investigates the eigenspace of the MIC matrix that is generated from original data sets and does not use the kernel method. The *emi*CCA may avoid these problems of *k*CCA and produces acceptable results. While in the present paper we found that *emi*CCA performed better than *k*CCA in our simulations, it is conceivable that there might be version of *k*CCA (e.g., implemented other kernel functions) that are comparable to our method. Further work is required to answer this issue.

Application to real data

In the real data set, the above-mentioned methods were utilized to characterize brain activation during a motor execution paradigm (Fig. 5). Using *l*CCA, one consistently task-related component that consisted of the right precentral gyrus and left cerebellum was identified, and this relationship was strongly linear. Due to the weights of *k*CCA that did not directly represent the weightiness of each independent component, *k*CCA was not suitable for our framework of fMRI data analysis.

For *emi*CCA, linearity and nonlinearity between the assumed BOLD responses (related to task) and the ICA time courses (related to fMRI recordings) were found in various independent components (IC27, IC9, IC20, IC11, IC31, IC12, IC24, IC32, IC36, and IC2). The relationships between the time course of IC27 and the assumed BOLD response were obviously linear, while nonlinearities were present in the other 9 components (Fig. 5). In previous studies, various cortical networks have been found during motor execution task (Kansaku et al., 2005; Sharma and Baron, 2013). In our work, the *emi*CCA identified IC27 (in primary motor cortex) as consistently related to the motor execution task, which corroborates previous studies (Gao et al., 2011; Halder et al., 2011; Kansaku et al., 2005). The supplementary motor cortex implicated in motor planning and learning (Halsband and Lange, 2006; Kansaku et al., 2005) was identified by IC11. Furthermore, vast networks (such as IC9, IC20, IC11, IC12, IC32, and IC2) may reflect a transient and rapid process that is associated with IC27 and may play an important role in supporting IC27 during the task. The precuneus (IC24) and medial frontal gyrus (IC36) may have exhibited task-induced deactivation (TID) that represents the reallocation of processing resources from areas in which TIDs occurred to areas that have been identified in task performance (McKiernan et al., 2006; McKiernan et al., 2003). It is not surprising that the execution of movements of the hand involves many cognitive components, such as perception, attention/working memory, response selection, and execution as well as including harmonizing of brain regions in vast networks, such as the visual (Calhoun et al., 2001a; Damoiseaux et al., 2006), attention (Fox et al., 2006a), and sensorimotor networks (Fox et al., 2006b). In addition, it is worth noting that some of the remaining 24 components may not be irrelevant (or meaningless) components. Descending absolute α values of all 34 components demonstrated that the different components perhaps had different contributions to the execution of movements of hand and might depict the results of harmonizing of brain regions in vast networks.

Linearity and nonlinearity in neural imaging

It is traditionally assumed that the neuronal electrophysiological response is linearly correlated with the BOLD fMRI signal. However, a certain extent of nonlinearity should be acknowledged in this relationship, and these nonlinearities can be present in the associations between tasks and neural activation, between neural activity and the BOLD response, or between tasks and the BOLD response (He and Liu, 2008; He et al., 2011). Many potential factors lead to nonlinearity in neural imaging, for example, nonlinearity may be derived from neural and/or vascular sources that reflect neurophysiological mechanisms (Birn and Bandettini, 2005; Zhang et al., 2008). In addition, some nonlinearity may be due to non-neurophysiological sources such as the vastly different temporal and spatial scales of hemodynamic and electrophysiological responses and even the variety of methods used to quantify signals (He et al., 2011). Using traditional *l*CCA, one might find that the linear relationships are involved in the fMRI data (see Fig. 5 for *l*CCA); however, one might lose important information related to nonlinearity that may be useful and meaningful. In this work, our *emi*CCA method, which is generalizable because it directly detected both linearity and nonlinearity in original data space (see Fig. 5 for *emi*CCA and Fig. 6), was utilized to detect novel relationships between the assumed BOLD responses (related to task) and ICA time courses (related to fMRI recording) in a motor execution paradigm.

Additionally, it is well established that HRFs vary across brain regions (Handwerker et al., 2004; Miezin et al., 2000), and this variation may potentially lead to nonlinearity (see Fig. S8). However, our method can, to some extent, perform well in this situation and further optimize the assumed BOLD response during calculation. In other words, *emi*CCA

tolerates several variably shaped HRFs due to its nonlinear processing ability. Considering that convolving neural activity with canonical HRFs of varying shapes (e.g., different onset times) is necessary in fMRI data analysis, defining the data set when applying *emi*CCA is good practice. Furthermore, based on the aforementioned evidence, the concepts of linearity and nonlinearity are represented in a Venn diagram using information theoretic quantities (Fig. 7). The relationships between assumed BOLD responses (related to task/stimulus) and fMRI data (related to recording) are linear or/and nonlinear; thus, it is better to use a more generalized method to explore these potential relationships. Because the MIC measure finds the maximal mutual information between two variates (Reshef et al., 2011), *emi*CCA may optimize the mutual information between assumed BOLD responses (related to task/stimulus) and fMRI recordings. Thus, the *emi*CCA method may extend our perspectives from linearity to nonlinearity and help us to better understand the potential process of brain activity.

Virtues of emiCCA and its applicability in the future

Currently, supervised methods always require that the experimentally controlled stimulation can be modeled. However, because of the complex neurovascular coupling mechanism, brain activity is always difficult to model. Thus, "noise" (so-called only because we do not have a good model) that is not directly reflected by, or related to, the stimulus may play a constructive role in neural activity (Biessmann et al., 2011; Ermentrout et al., 2008). The *emi*CCA can directly assesses various relationships in the original data space (no need to find optimal kernel functions to transform data); therefore, this method may have better general applicability than *k*CCA and *l*CCA and may help to explore and identify interesting and novel patterns in various data. Furthermore, in view of different measurements (e.g., EEG and fMRI) that have respective strengths and weaknesses that are complementary, multimodal fusion has been widely researched as a means to achieve more accurate information regarding brain activity (Dong et al., 2014; Huster et al., 2012). Various fusion techniques have been developed, and the three most influential approaches for EEG-fMRI integration are (He et al., 2011; Huster et al., 2012; Rosa et al., 2010) fMRI-informed EEG (Lei et al., 2011), EEG-informed fMRI (Debener et al., 2005; Lei et al., 2010; Luo et al., 2010), and symmetric EEG-fMRI fusion (Valdes-Sosa et al., 2009). As a data-driven method, *emi*CCA symmetrically analyzes both modalities simultaneously and avoids any possible bias and has the potential to uncover the underlying mechanisms of brain activity in multimodal fusion studies. Thus, *emi*CCA could constitute a counterpart to model-driven neuro-generative approaches. In brief, because the definitions of data sets can be very flexible on *emi*CCA, *emi*CCA is likely to provide important information that will extend our understanding of brain functions and dysfunctions and represents a promising technique for various areas of research, such as target detection in cognitive processing (Calhoun et al., 2006), interictal epileptic discharges (Luo et al., 2010; Marques et al., 2009), and resting-state processes (Damoiseaux et al., 2006).

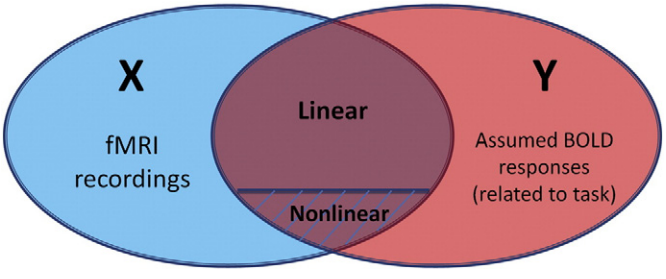

Fig. 7. Illustration of the linear and nonlinear aspects of neural imaging. Information theoretic quantities are displayed as areas in a Venn diagram. The intersection between the assumed BOLD responses (related to task) and the fMRI recordings may contain both linearity and nonlinearity. Linearity may be primary, and nonlinearity that directly reflects nonlinear neurophysiological mechanisms may also exist. However, nonlinearity that is due to non-neurophysiological sources (such as vastly different spatial/temporal scales of BOLD and electrophysiological response, unsuitable HRFs, and various methods of quantifying multimodal signals) can also be considered.

Limitations

Several limitations and suggestions for the further development of the approach should be considered. First, Kinney and Atwal have argued the "equitability" property of MIC is not strictly satisfied (Kinney and Atwal, 2013); therefore, it is necessary to pay attention to this issue when using *emi*CCA to examine the relationships between data sets. However, more evidence and applications are needed to assess our method in the future. Second, because MIC does not detect for various complex relationships (e.g., circular and nonseparable relationships) very well (not ~1), *emi*CCA may not capture these features perfectly (see Fig. 4). However, we suggest that preprocessing to remove artifacts, to extract features (e.g., using PCA and ICA), and to develop a sufficient sample sizes can alleviate this problem and can also generate reliable results. Additionally, we were only concerned about the maximal MIEC in our work, which according to the generality of MIC, may be reasonable in practical applications; however, other MIECs may be useful depending on the experimental context. Despite these potential limitations, the further development and application of *emi*CCA will be the focus of future research.

Conclusion

In conclusion, the novelty of this work encompasses that, in the process termed *emi*CCA, we utilized the MIC approach to directly identify potential relationships between data sets to increase the generalizability of various data analyses (e.g., potential nonlinearities, different data sets) that were demonstrated with simulation data. Furthermore, an example framework based on *emi*CCA was proposed for the analysis of fMRI data, and this example performed well in revealing underlying brain functions of a real fMRI data set. This method has the potential for applications involving unimodal analyses and multimodal fusion and will likely provide important information that will further our understanding of various cognitive processes.

Acknowledgments

All of the authors have no conflict of interest. This work was supported by the 973 Project (2011CB707803), the National Nature Science Foundation of China (Nos. 91232725, 81271547, 81330032, 81471638), the 863 project (No. 2012BAI16B02), the "111" Project (B12027), and the Program for Changjiang Scholars and Innovative Research Team (IRT 0910).

Appendix A

For two data sets, $X = X' + \boldsymbol{\varepsilon_x} \in R^{n \times m}$ and $Y = Y' + \boldsymbol{\varepsilon_y} \in R^{n \times q}$, where $\boldsymbol{\varepsilon}$ is a random noise term, we define the MIC matrix,

$$\Phi = \begin{pmatrix} \Phi_{X,X} & \Phi_{X,Y} \\ \Phi_{Y,X} & \Phi_{Y,Y} \end{pmatrix} \tag{A.1}$$

where $\Phi_{(\cdot,\cdot)}$ is the MIC matrix between or within data sets. Here, we use Lagrange multipliers to solve the following problem:

$$\begin{aligned} &\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \quad \boldsymbol{\alpha}^{T}\Phi_{X,Y}\boldsymbol{\beta} \\ &\text{subject to} \quad \text{var}(X*\boldsymbol{\alpha}) = \text{var}(Y*\boldsymbol{\beta}) = 1 \end{aligned} \tag{A.2}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of relative weights corresponding to data sets X and Y, respectively. First, considering the constraint in Eq. (A.2), generally, we can suppose that

$$\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,X}}\boldsymbol{\alpha} = \mathrm{C}_1 \neq 0 \tag{A.3}$$

$$\boldsymbol{\beta}^{\mathrm{T}}\Phi_{\mathrm{Y,Y}}\boldsymbol{\beta} = \mathrm{C}_2 \neq 0 \tag{A.4}$$

where C_1 and C_2 are constant. Then the Lagrangian equation can be given as

$$\mathcal{L}(\lambda, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} - \frac{1}{2}\lambda\sqrt{\frac{\mathrm{C}_2}{\mathrm{C}_1}}\left(\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,X}}\boldsymbol{\alpha} - \mathrm{C}_1\right) - \frac{1}{2}\mu\sqrt{\frac{\mathrm{C}_1}{\mathrm{C}_2}}\left(\boldsymbol{\beta}^{\mathrm{T}}\Phi_{\mathrm{Y,Y}}\boldsymbol{\beta} - \mathrm{C}_2\right) \tag{A.5}$$

Taking derivatives in respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we obtain:

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\alpha}} = \Phi_{\mathrm{X,Y}}\boldsymbol{\beta} - \lambda\sqrt{\frac{\mathrm{C}_2}{\mathrm{C}_1}}\Phi_{\mathrm{X,X}}\boldsymbol{\alpha} = 0 \tag{A.6}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\beta}} = \Phi_{\mathrm{Y,X}}\boldsymbol{\alpha} - \mu\sqrt{\frac{\mathrm{C}_1}{\mathrm{C}_2}}\Phi_{\mathrm{Y,Y}}\boldsymbol{\beta} = 0 \tag{A.7}$$

Subtracting $\boldsymbol{\beta}^{\mathrm{T}}$ times Eq. (A.7) from $\boldsymbol{\alpha}^{\mathrm{T}}$ times Eq. (A.6), we have

$$\begin{aligned} 0 &= \boldsymbol{\beta}^{\mathrm{T}}\Phi_{\mathrm{Y,X}}\boldsymbol{\alpha} - \mu\sqrt{\frac{\mathrm{C}_1}{\mathrm{C}_2}}\boldsymbol{\beta}^{\mathrm{T}}\Phi_{\mathrm{Y,Y}}\boldsymbol{\beta} - \boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} + \lambda\sqrt{\frac{\mathrm{C}_2}{\mathrm{C}_1}}\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,X}}\boldsymbol{\alpha} \\ &= \lambda\sqrt{\frac{\mathrm{C}_2}{\mathrm{C}_1}}\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,X}}\boldsymbol{\alpha} - \mu\sqrt{\frac{\mathrm{C}_1}{\mathrm{C}_2}}\boldsymbol{\beta}^{\mathrm{T}}\Phi_{\mathrm{Y,Y}}\boldsymbol{\beta} \end{aligned} \tag{A.8}$$

which together with the constraints (Eqs. (A.3) and (A.4)) implies that

$$\lambda = \mu = \frac{1}{\sqrt{\mathrm{C}_1\mathrm{C}_2}}\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} \tag{A.9}$$

Assuming $\Phi_{\mathrm{X,X}}$ and $\Phi_{\mathrm{Y,Y}}$ are invertible, we have

$$\Phi_{\mathrm{X,X}}^{-1}\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} = \lambda\sqrt{\frac{\mathrm{C}_2}{\mathrm{C}_1}}\boldsymbol{\alpha} \tag{A.10}$$

$$\Phi_{\mathrm{Y,Y}}^{-1}\Phi_{\mathrm{Y,X}}\boldsymbol{\alpha} = \mu\sqrt{\frac{\mathrm{C}_1}{\mathrm{C}_2}}\boldsymbol{\beta} \tag{A.11}$$

substituting Eqs. (A.11) and (A.10) into each other gives

$$\Phi_{\mathrm{X,X}}^{-1}\,\Phi_{\mathrm{X,Y}}\,\Phi_{\mathrm{Y,Y}}^{-1}\,\Phi_{\mathrm{Y,X}}\boldsymbol{\alpha} = \lambda\mu\boldsymbol{\alpha} = \lambda^2\boldsymbol{\alpha} \tag{A.12}$$

$$\Phi_{\mathrm{Y,Y}}^{-1}\,\Phi_{\mathrm{Y,X}}\,\Phi_{\mathrm{X,X}}^{-1}\,\Phi_{\mathrm{X,Y}}\boldsymbol{\beta} = \lambda\mu\boldsymbol{\beta} = \lambda^2\boldsymbol{\beta} \tag{A.13}$$

We are therefore left with an eigenproblem of the form $A\boldsymbol{\nu} = \lambda^*\boldsymbol{\nu}$. The vectors of weights ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) that represent the weightiness of each dimension of data in the cross MIC matrix $\Phi_{\mathrm{X,Y}}$ are the eigenvectors of the following two matrices, respectively:

$$\Phi_{\mathrm{X,X}}^{-1}\,\Phi_{\mathrm{X,Y}}\,\Phi_{\mathrm{Y,Y}}^{-1}\,\Phi_{\mathrm{Y,X}} \tag{A.14}$$

$$\Phi_{\mathrm{Y,Y}}^{-1}\,\Phi_{\mathrm{Y,X}}\,\Phi_{\mathrm{X,X}}^{-1}\,\Phi_{\mathrm{X,Y}} \tag{A.15}$$

The maximal information eigen coefficient (MIEC) is $\lambda > 0$, which is the square root of the eigenvalues (λ^*).

Appendix B

For two data sets, $\mathrm{X} = \mathrm{X}' + \boldsymbol{\varepsilon_x} \in R^{n\times m}$ and $\mathrm{Y} = \mathrm{Y}' + \boldsymbol{\varepsilon_y} \in R^{n\times q}$, where $\boldsymbol{\varepsilon}$ is a random noise term, assuming there exists an unknown and nowhere-constant function f that

$$\mathrm{Y}'_j = f_{ij}(\mathrm{X}'_i);\ i\in D_x\subseteq\{1,\ldots,m\};\ j\in D_y\subseteq\{1,\ldots,q\} \tag{B.1}$$

Then we have

$$\phi\left(\mathrm{X}_i;\mathrm{Y}_j\right) > 0;\ i\in D_x;\ j\in D_y \tag{B.2}$$

where ϕ is the maximal information coefficient (MIC) that falls between 0 and 1 (Reshef et al., 2011). Between stochastic variables x and y, the MIC can be given as

$$\phi(x;y) = \max_{|\mathbb{X}||\mathbb{Y}|<B}\left\{\frac{I^*(\mathbb{X};\mathbb{Y})}{log_2\{min\{|\mathbb{X}|,|\mathbb{Y}|\}\}}\right\} \tag{B.3}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are bins of a rectangular grid on the x–y scatter plot, and $I^*(.;.)$ is the maximal mutual information achieved by any grid on the data. $|\mathbb{X}||\mathbb{Y}|<B$ means that the total number of bins is less than some number B. With the same function f_{ij}, generally, we can suppose that

$$\mathrm{corr}\left(f_{ij}\left(\mathrm{X}'_i+\boldsymbol{\varepsilon}_x^i\right),\mathrm{Y}'_j+\boldsymbol{\varepsilon}_y^j\right)\geq 0;\ i\in D_x;\ j\in D_y \tag{B.4}$$

where corr(.,.) is the Pearson correlation. Due to the property of MIC that it roughly equals the squared correlation (coefficient of determination) of the data relative to the noiseless function, it increases with the increase of the squared correlation in the domain [0, 1] (and vice versa) (Reshef et al., 2013; Reshef et al., 2011). Thus, we have

$$\mathrm{corr}^2\left(f_{ij}\left(\mathrm{X}'_i+\boldsymbol{\varepsilon}_x^i\right),\mathrm{Y}'_j+\boldsymbol{\varepsilon}_y^j\right)\cong\Phi_{\mathrm{X}_i,\mathrm{Y}_j};\ i\in D_x;\ j\in D_y \tag{B.5}$$

where $\Phi_{\mathrm{X,Y}}$ is the MIC matrix between X and Y. Furthermore, we can suppose two new data sets, X^* and Y^*, let

$$\mathrm{corr}\left(\mathrm{X}_i^*,\mathrm{Y}_j^*\right) = \Phi_{\mathrm{X}_i,\mathrm{Y}_j}\cong\mathrm{corr}^2\left(f_{ij}\left(\mathrm{X}'_i+\boldsymbol{\varepsilon}_x^i\right),\mathrm{Y}'_j+\boldsymbol{\varepsilon}_y^j\right);\ i\in D_x;\ j\in D_y \tag{B.6}$$

which indicates that relationships between X and Y are approximately represented in data sets X^* and Y^*. Then considering weights (linear combinations)

$$\mathrm{U}^* = \sum_{i=1}^{m}\mathbf{a}_i\mathrm{X}_i^* = \mathrm{X}^*\mathbf{a}\in R^{n\times m} \tag{B.7}$$

$$\mathrm{V}^* = \sum_{j=1}^{q}\mathbf{b}_j\mathrm{Y}_j^* = \mathrm{Y}^*\mathbf{b}\in R^{n\times q} \tag{B.8}$$

such that the pairwise correlation across the two new data sets (X^* and Y^*) is maximized, where $\mathbf{a}, \mathbf{b} \neq \mathbf{0}$. According to Eqs. (B.5) and (B.6), we have

$$\max\ \mathrm{corr}(\mathrm{U}^*,\mathrm{V}^*) = \max\ \mathbf{a}^{\mathrm{T}}\mathrm{corr}(\mathrm{X}^*,\mathrm{Y}^*)\mathbf{b} = \max\ \mathbf{a}^{\mathrm{T}}\Phi_{\mathrm{X,Y}}\mathbf{b} \tag{B.9}$$

To reduce the freedom of the scaling of $\mathbf{a}$ and $\mathbf{b}$, we may add a further constraint:

$$\mathrm{var}(\mathrm{X}^*\mathbf{a}) = \mathrm{var}(\mathrm{Y}^*\mathbf{b}) = 1 \tag{B.10}$$

where var(·) is the variance. Noting that X^* and Y^* are unknown, the optimization can be equivalent to solving the following problem:

$$\underset{\boldsymbol{a},\boldsymbol{b}}{\text{maximize}} \quad \mathbf{a}^{\mathrm{T}}\Phi_{X,Y}\mathbf{b} = \underset{\boldsymbol{a},\boldsymbol{b}}{\text{maximize}} \quad \mathrm{corr}(U^*, V^*) \\ \text{subject to} \quad \mathrm{var}(X\mathbf{a}) = \mathrm{var}(Y\mathbf{b}) = 1 \tag{B.11}$$

where $\Phi_{X,Y}$ is the MIC between X and Y, and *var*(·) is the variance. Problem (B.11) means that *emi*CCA may be related to a generalized *l*CCA in which the eigenvectors **a** and **b** in *emi*CCA are the generalized eigenvectors of *l*CCA in its unknown linearized space. However, this is merely of conceptual relevance, as the requirement in Eq. (B.6) is very complex (also not necessary) to generate the two new data sets (X^* and Y^*) from the original data (X and Y). In general, it is also hard to find the optimal kernel functions or family of kernel functions in the *k*CCA; therefore, we consider *emi*CCA as a different method instead of a different version of *k*CCA.

Appendix C

Here, we assume that the significance of each MIEC can be obtained through random permutation tests of MIC between the weighted sum of X and Y. The direct hypothesis test of the *emi*CCA is as follows:

$$\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{X,Y}\boldsymbol{\beta} = 0; \boldsymbol{\alpha}, \boldsymbol{\beta} \neq 0 \tag{C.1}$$

where $\Phi_{(\cdot,\cdot)}$ is the MIC matrix. Now we use apagoge to prove the following proposition: if $\phi(U; V) = 0$, then $\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{X,Y}\boldsymbol{\beta} = 0$, where ϕ is MIC, and U and V are the weighted sum of X and Y, respectively, which are derived from the following equations:

$$U = \sum_{i=1}^{m} \boldsymbol{\alpha}_i X_i \in R^{n\times m} \tag{C.2}$$

$$V = \sum_{j=1}^{q} \boldsymbol{\beta}_j Y_j \in R^{n\times q} \tag{C.3}$$

Supposing that $\exists\ \boldsymbol{\alpha}, \boldsymbol{\beta} \neq 0$ makes $\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{X,Y}\boldsymbol{\beta} \neq 0$, where $\boldsymbol{\alpha}, \boldsymbol{\beta} = \mathit{argmax}\ \boldsymbol{\alpha}^{\mathrm{T}}\Phi_{X,Y}\boldsymbol{\beta}$. Obviously, we have

$$\boldsymbol{\alpha}^{\mathrm{T}}\Phi_{X,Y}\boldsymbol{\beta} \neq 0 \Rightarrow \Phi_{X,Y} \neq 0 \tag{C.4}$$

Because MIC assigns scores that approach 0 to the statistically independent variables (Reshef et al., 2011), it yields U is not independent of V and $\phi(U; V) \neq 0$. This conflicts with our assumption, and the proposition is proved. Therefore, because MIC depends only on the rank ordering of the data (Reshef et al., 2011), the *p*-value of each MIEC can be obtained through random permutation tests of MIC, $\phi(U; V)$.

Appendix D. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.01.006.

References

Akaho, S., 2006. A kernel method for canonical correlation analysis (arXiv, preprint cs/0609071).

Balakrishnan, S., Puniyani, K., Lafferty, J., 2012. Sparse Additive Functional and Kernel CCA. 29th International Conference of Machine Learning, Edinburgh.

Biessmann, F., Meinecke, F.C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N.K., Muller, K.R., 2010. Temporal kernel CCA and its application in multimodal neuronal data analysis. Mach. Learn. 79, 5–27.

Biessmann, F., Plis, S., Meinecke, F.C., Eichele, T., Muller, K.R., 2011. Analysis of multimodal neuroimaging data. IEEE Rev. Biomed. Eng. 4, 26–58.

Birn, R.M., Bandettini, P.A., 2005. The effect of stimulus duty cycle and "off" duration on BOLD response linearity. NeuroImage 27, 70–82.

Calhoun, V.D., Adali, T., McGinty, V.B., Pekar, J.J., Watson, T.D., Pearlson, G.D., 2001a. fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. NeuroImage 14, 1080–1088.

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, 140–151.

Calhoun, V.D., Adali, T., Pearlson, G.D., Kiehl, K.A., 2006. Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. NeuroImage 30, 544–553.

Calhoun, V.D., Liu, J., Adali, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage 45, S163–S172.

Cordes, D., Jin, M., Curran, T., Nandy, R., 2012. Optimizing the performance of local canonical correlation analysis in fMRI using spatial constraints. Hum. Brain Mapp. 33, 2611–2626.

Correa, N.M., Li, Y.O., Adali, T., Calhoun, V.D., 2008. Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. IEEE J. Sel. Top. Signal. Process. 2, 998–1007.

Correa, N.M., Adali, T., Li, Y.O., Calhoun, V.D., 2010. Canonical correlation analysis for data fusion and group inferences: examining applications of medical imaging data. IEEE Signal Process. Mag. 27, 39–50.

Damoiseaux, J.S., Rombouts, S.A., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. U. S. A. 103, 13848–13853.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D.Y., Engel, A.K., 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J. Neurosci. 25, 11730–11737.

Dong, L., Gong, D., Valdes-Sosa, P.A., Xia, Y., Luo, C., Xu, P., Yao, D., 2014. Simultaneous EEG-fMRI: trial level spatio-temporal fusion for hierarchically reliable information discovery. NeuroImage 99, 28–41.

Ermentrout, G.B., Galan, R.F., Urban, N.N., 2008. Reliability, synchrony and noise. Trends Neurosci. 31, 428–434.

Fox, M.D., Corbetta, M., Snyder, A.Z., Vincent, J.L., Raichle, M.E., 2006a. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc. Natl. Acad. Sci. U. S. A. 103, 10046–10051.

Fox, M.D., Snyder, A.Z., Zacks, J.M., Raichle, M.E., 2006b. Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. Nat. Neurosci. 9, 23–25.

Friman, O., Cedefamn, J., Lundberg, P., Borga, M., Knutsson, H., 2001. Detection of neural activity in functional MRI using canonical correlation analysis. Magn. Reson. Med. 45, 323–330.

Friston, K.J., 2007. Statistical parametric mapping: the analysis of functional brain images. 1st ed. Elsevier/Academic Press, Amsterdam; Boston.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. NeuroImage 2, 45–53.

Fukumizu, K., Bach, F.R., Gretton, A., 2007. Statistical consistency of kernel canonical correlation analysis. J. Mach. Learn. Res. 8, 361–383.

Gao, Q., Duan, X., Chen, H., 2011. Evaluation of effective connectivity of motor areas during motor imagery and execution using conditional Granger causality. NeuroImage 54, 1280–1288.

Halder, S., Agorastos, D., Veit, R., Hammer, E.M., Lee, S., Varkuti, B., Bogdan, M., Rosenstiel, W., Birbaumer, N., Kubler, A., 2011. Neural mechanisms of brain–computer interface control. NeuroImage 55, 1779–1790.

Halsband, U., Lange, R.K., 2006. Motor learning in man: a review of functional and clinical studies. J. Physiol. Paris 99, 414–424.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651.

Hardoon, D.R., Szedmak, S., Shawe-Taylor, J., 2004. Canonical correlation analysis: an overview with application to learning methods. Neural Comput. 16, 2639–2664.

Hardoon, D.R., Mourao-Miranda, J., Brammer, M., Shawe-Taylor, J., 2007. Unsupervised analysis of fMRI data using kernel canonical correlation. NeuroImage 37, 1250–1259.

He, B., Liu, Z., 2008. Multimodal functional neuroimaging: integrating functional MRI and EEG/MEG. IEEE Rev. Biomed. Eng. 1, 23–40.

He, B., Yang, L., Wilke, C., Yuan, H., 2011. Electrophysiological imaging of brain activity and connectivity-challenges and opportunities. IEEE Trans. Biomed. Eng. 58, 1918–1931.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28, 321–377.

Huster, R.J., Debener, S., Eichele, T., Herrmann, C.S., 2012. Methods for simultaneous EEG-fMRI: an introductory review. J. Neurosci. 32, 6053–6060.

Kansaku, K., Muraki, S., Umeyama, S., Nishimori, Y., Kochiyama, T., Yamane, S., Kitazawa, S., 2005. Cortical activity in multiple motor areas during sequential finger movements: an application of independent component analysis. NeuroImage 28, 669–681.

Kinney, J.B., Atwal, G.S., 2013. Equitability, mutual information, and the maximal information coefficient (arXiv, preprint arXiv:1301.7745).

Lei, X., Qiu, C., Xu, P., Yao, D., 2010. A parallel framework for simultaneous EEG/fMRI analysis: methodology and simulation. NeuroImage 52, 1123–1134.

Lei, X., Xu, P., Luo, C., Zhao, J., Zhou, D., Yao, D., 2011. fMRI functional networks for EEG source imaging. Hum. Brain Mapp. 32, 1141–1160.

Li, Y.O., Adali, T., Calhoun, V.D., 2007. Estimating the number of independent components for functional magnetic resonance imaging data. Hum. Brain Mapp. 28, 1251–1266.

Luo, C., Yao, Z., Li, Q., Lei, X., Zhou, D., Qin, Y., Xia, Y., Lai, Y., Gong, Q., Yao, D., 2010. Imaging foci of epileptic discharges from simultaneous EEG and fMRI using the canonical HRF. Epilepsy Res. 91, 133–142.

Marques, J.P., Rebola, J., Figueiredo, P., Pinto, A., Sales, F., Castelo-Branco, M., 2009. ICA decomposition of EEG signal for fMRI processing in epilepsy. Hum. Brain Mapp. 30, 2986–2996.
McKiernan, K.A., Kaufman, J.N., Kucera-Thompson, J., Binder, J.R., 2003. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. J. Cogn. Neurosci. 15, 394–408.
McKiernan, K.A., D'Angelo, B.R., Kaufman, J.N., Binder, J.R., 2006. Interrupting the "stream of consciousness": an fMRI investigation. NeuroImage 29, 1185–1191.
Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 11, 735–759.
Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C., 2011. Detecting novel associations in large data sets. Science 334, 1518–1524.
Reshef, D., Reshef, Y., Mitzenmacher, M., Sabeti, P., 2013. Equitability Analysis of the Maximal Information Coefficient, with Comparisons (arXiv, preprint arXiv:1301.6314).
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242.
Rosa, M.J., Daunizeau, J., Friston, K.J., 2010. EEG-fMRI integration: a critical review of biophysical modeling and data analysis approaches. J. Integr. Neurosci. 9, 453–476.
Sharma, N., Baron, J.C., 2013. Does motor imagery share neural networks with executed movement: a multivariate fMRI analysis. Front. Hum. Neurosci. 7, 564.
Speed, T., 2011. Mathematics. A correlation for the 21st century. Science 334, 1502–1503.
Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Sotero, R.C., Iturria-Medina, Y., Aleman-Gomez, Y., Bosch-Bayard, J., Carbonell, F., Ozaki, T., 2009. Model driven EEG/fMRI fusion of brain oscillations. Hum. Brain Mapp. 30, 2701–2721.
Zhang, N., Zhu, X.H., Chen, W., 2008. Investigating the source of BOLD nonlinearity in human visual cortex in response to paired visual stimuli. NeuroImage 43, 204–212.
Zhang, Y., Xu, P., Liu, T., Hu, J., Zhang, R., Yao, D., 2012. Multiple frequencies sequential coding for SSVEP-based brain–computer interface. PLoS ONE 7, e29519.