

# A MULTIVARIATE MODEL FOR COMPARISON OF TWO DATASETS AND ITS APPLICATION TO FMRI ANALYSIS \*

*Yi-Ou Li and Tülay Adalı*

University of Maryland Baltimore County  
Baltimore, MD

*Vince D. Calhoun*

The MIND Institute and the  
University of New Mexico, Albuquerque, NM  
Yale University, New Haven, CT

## ABSTRACT

In this work, we propose a structured approach to compare common and distinct features of two multidimensional datasets using a combination of canonical correlation analysis (CCA) and independent component analysis (ICA). We develop formulations of information theoretic criteria to determine the dimension of the subspaces for common and distinct features of the two datasets. We apply the proposed method to a simulated dataset to demonstrate that it improves the estimation of both common and distinct features when compared to performing ICA on the concatenation of two datasets. We also apply the method to compare brain activation in functional magnetic resonance imaging (fMRI) data acquired during a simulated driving experiment and observe distinctions between the driving and watching conditions revealed in relevant brain function studies.

## 1. INTRODUCTION

The comparison of two sets of multidimensional data, e.g., fMRI data acquired for two different tasks, or, data collected from control and diseased groups, is a frequently encountered problem in data analysis. As a typical approach, multivariate methods such as independent component analysis (ICA) can be applied within each dataset to represent the data as a set of independent features. The estimated features from the two datasets can then be matched by using a suitable metric such as correlation. However, cross-correlations among all the estimated components between the two sets introduces ambiguity in feature matching. As another approach, group ICA [1] can be performed to obtain associated component estimates from the two datasets, but this approach considers only common components between the two datasets. As a third approach, ICA can be applied directly to the concatenation of two datasets, and the profiles of the mixing vectors can be used as an indicator of the

difference between the two datasets with respect to the corresponding estimated independent components. However, since the independence assumption inherent in ICA is not always plausible for the two datasets that might share common attributes, a global ICA decomposition on the concatenated dataset could attenuate the distinct features from each dataset. To improve the estimation of common and distinct features, we propose a decomposition of each dataset into two subspaces containing, respectively, the common and distinct features. Therefore, ICA can be applied within each subspace to avoid the interference of the two types of features in estimation.

For the subspace decomposition, we propose using canonical correlation analysis (CCA) to transform each dataset into a group of orthogonal components where the components between the two datasets correlate only on the associated indices, *i.e.* the two sets of components are bi-orthogonal. These correlation values are used to identify the subspace of common and distinct features. Specifically, the dimension of the subspace for distinct features is determined by information theoretic criteria (ITC) based on the model that those distinct components between the two datasets have zero correlation. ICA is individually applied within each subspace to estimate, respectively, the common and distinct features.

In fMRI data analysis, one is typically interested in comparing brain activations between different conditions, e.g., task and resting state, healthy versus diseased subjects performing the same task. Though, the localization and spatially distributed nature of networks make the independence assumption plausible, the similarity of areas of activation and their nature is likely to affect the results of analysis due to some dependence in the activations in the two conditions. Hence, such a partitioning is expected to improve the results of a subsequent ICA analysis as we demonstrate by using a simulated data set. We also apply the method to analysis of fMRI data obtained during a simulated driving experiment to study differences in components during the driving and watching epochs of the experiment.

In the next section, we outline the application of CCA

---

\*THIS RESEARCH IS SUPPORTED IN PART BY THE NIH UNDER GRANTS 1 R01 EB 000840-02 AND R01 AA015615.

to two multidimensional datasets. We derive the ITC formulations to select the common dimensionality of the two datasets in Section 3. In Section 4, we show experimental results of the proposed method on simulated and true fMRI data. In the last section, we conclude our work with a discussion of several interesting aspects of the proposed method.

## 2. CANONICAL CORRELATION ANALYSIS OF TWO MULTIDIMENSIONAL DATASETS

Given two multidimensional datasets  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , PCA is performed independently to each dataset as the first step. PCA is a typically used preprocessing step in multivariate data analysis to decompose the data into a set of uncorrelated principal components ordered by the variance of each component. For data with low contrast to noise ratio and large dimensionality such as the fMRI time sequences with over 100 time points, dimension reduction is usually performed based on, e.g., variance of the principal components, to remove the insignificant principal components due to the noise effect. The dimension reduced data are then whitened by normalizing each principal component to unit variance given by

$$\mathbf{y}_i = \Lambda_i^{-\frac{1}{2}} \mathbf{B}_i \mathbf{x}_i, \quad i = 1, 2,$$

where  $\Lambda_i$  and  $\mathbf{B}_i$  are, respectively, the diagonal eigenvalue and eigenvector matrix of the covariance matrix  $E\{\mathbf{x}_i \mathbf{x}_i^T\}$  for the  $i^{\text{th}}$  dataset.

Without loss of generality, we assume that the whitened principal components from each dataset,  $\mathbf{y}_1$  and  $\mathbf{y}_2$ , are  $p \times 1$  and  $q \times 1$  vectors respectively with  $p \leq q$ . Although there is no correlation among the components within each dataset, there could be correlations among the principal components across the two datasets. The cross-correlation makes it difficult to perform a direct comparison of the principal components from the two datasets. Therefore, further (linear) transformations of the principal components in each dataset are motivated to remove the between-set cross-correlations while preserving the whiteness of the principal components within each dataset.

To achieve this aim, we seek two orthonormal transformations:  $\mathbf{E}_1$  and  $\mathbf{E}_2$ , and apply the transformations to  $\mathbf{y}_1$  and  $\mathbf{y}_2$  respectively, such that: (i) the transformed variates  $\mathbf{z}_1 = \mathbf{E}_1 \mathbf{y}_1$  and  $\mathbf{z}_2 = \mathbf{E}_2 \mathbf{y}_2$  are bi-orthogonal, i.e.,  $E\{\mathbf{z}_1 \mathbf{z}_2^T\} = \mathbf{D}$  where  $\mathbf{D}$  is a diagonal matrix; and (ii)  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are whitened such that  $E\{\mathbf{z}_i \mathbf{z}_i^T\} = \mathbf{I}$  for  $i = 1, 2$ . The two orthonormal transformations can be obtained using CCA [2] as described below.

Let  $\mathbf{D}$  and  $\mathbf{E}_1$  be the eigenvalue and eigenvector matrix of  $E\{\mathbf{y}_1 \mathbf{y}_2^T\} \cdot E\{\mathbf{y}_2 \mathbf{y}_1^T\}$ , and  $\mathbf{E}_2 = \mathbf{D}^{-1/2} \cdot \mathbf{E}_1 \cdot E\{\mathbf{y}_1 \mathbf{y}_2^T\}$ . It is straightforward to verify that:

- (i)  $E\{\mathbf{z}_1 \mathbf{z}_2^T\} = \mathbf{E}_1 \cdot E\{\mathbf{y}_1 \mathbf{y}_2^T\} \cdot \mathbf{E}_2^T = \mathbf{D}^{1/2}$ ,
- (ii)  $E\{\mathbf{z}_1 \mathbf{z}_1^T\} = \mathbf{E}_1 \cdot E\{\mathbf{y}_1 \mathbf{y}_1^T\} \cdot \mathbf{E}_1^T = \mathbf{I}$ , and

$$(iii) E\{\mathbf{z}_2 \mathbf{z}_2^T\} = \mathbf{E}_2 \cdot E\{\mathbf{y}_2 \mathbf{y}_2^T\} \cdot \mathbf{E}_2^T = \mathbf{I}.$$

Therefore, two new sets of variates  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are both  $p \times 1$  whitened vectors and they correlate only on the same indices with the correlations measured by the canonical correlation coefficients that are the diagonal entries of  $\mathbf{D}^{1/2}$ . The obtained variates are defined as the canonical variates (CVs).

## 3. SELECTION OF THE COMMON DIMENSION OF TWO MULTIDIMENSIONAL DATASETS

Among different approaches for model order selection, information theoretic criteria (ITC) have proven to be particularly attractive for many signal processing applications. One of the major advantages of the approach is the automation of the order selection process so that no empirical threshold value needs to be specified.

A commonly used ITC for order selection, Akaike's information criterion (AIC), is developed based on the minimization of the Kullback-Leibler divergence between the true model and the fitted model [3]. The closely related minimum description length (MDL) criterion and the Bayesian information criterion (BIC) are developed based on the minimum code length [4] and the Bayes solution to model order selection [5] respectively.

AIC and MDL share the general form given below:

$$\hat{k} = \arg \min_k \{-\mathcal{L}(\Theta_k) + \mathcal{G}(\Theta_k)\}$$

where  $\mathcal{L}(\cdot)$  is the maximum likelihood of the observed data based on the assumed model parameterized by  $\Theta_k$ , and  $\mathcal{G}(\cdot)$  is a penalty term related to the number of free parameters in  $\Theta_k$ , i.e., the intrinsic dimension of the parameter space for the assumed model.

Under the assumption that the two sets of CVs,  $\mathbf{y}_1$  and  $\mathbf{y}_2$ , are jointly Gaussian random vectors, the maximum likelihood can be written as [6]

$$\mathcal{L}(\Theta_{p+q}) = -\frac{1}{2} N \sum_{i=1}^{p+q} \log(\lambda_i) \quad (1)$$

where  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue of the covariance matrix  $\mathbf{R}_y = E\{\mathbf{y} \mathbf{y}^T\}$ ,  $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ , in descending order and  $N$  is the number of independent and identically distributed (i.i.d.) data samples.

When CCA is applied to  $\mathbf{y}_1$  and  $\mathbf{y}_2$  as described in the previous section, since  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are orthonormal transformations, the effective transformation to the aggregated data  $\mathbf{y}$  is orthonormal, which can be written as

$$\mathbf{z} \equiv \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{E}_1 & | & 0 \\ 0 & | & \mathbf{E}_2 \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

Thus, the covariance matrix of  $\mathbf{z}$  has the same eigenvalues as that of  $\mathbf{y}$  due to the orthonormal transformation—

except that, when  $p \leq q$ ,  $E\{\mathbf{y}\mathbf{y}^T\}$  has  $q - p$  additional unit eigenvalues, a result that can be easily shown.

The covariance matrix of the concatenated CVs has four diagonal blocks:

$$E\{\mathbf{z}\mathbf{z}^T\} = \begin{bmatrix} \mathbf{I} & & d_1 & \dots & 0 \\ & & \dots & d_k & \dots \\ & & 0 & \dots & d_{k+1} & \dots \\ & & & & & \dots \\ & & & & & \mathbf{I} \\ d_1 & \dots & 0 & & & d_p \\ \dots & d_k & & & & \\ d_{k+1} & \dots & & & & \\ 0 & & & & & d_p \end{bmatrix}$$

where  $d_i$  is the canonical correlation coefficient of the  $i$ th pair of CVs and  $d_i \in [0, 1], \forall i$ . Given this block diagonal structure of the covariance matrix  $E\{\mathbf{z}\mathbf{z}^T\}$ , it can be shown that its eigenvalues are uniquely determined by the canonical correlation coefficients as

$$\{\boldsymbol{\lambda}\} = \{1 + d_1, 1 + d_2, \dots, 1 + d_p, 1 - d_p, \dots, 1 - d_2, 1 - d_1\}.$$

When the first  $k$  CVs in each dataset are selected to be the common CVs, the rest of the CVs are deemed to be the distinct CVs by ignoring their insignificant canonical correlation coefficients, i.e.,  $d_i = 0, \forall i = k + 1, \dots, p$ . In this case, the eigenspectrum of  $E\{\mathbf{z}\mathbf{z}^T\}$  becomes:

$$\{\boldsymbol{\lambda}\} = \{1 + d_1, 1 + d_2, \dots, 1 + d_k, 1, \dots, 1, 1 - d_k, \dots, 1 - d_2, 1 - d_1\}$$

where there are  $2(p - k)$  “1”s due to the specified zero canonical correlation coefficients for the set of distinct CVs.

Therefore, the maximum likelihood in Eq. 1 takes the form:

$$\mathcal{L}(\boldsymbol{\Theta}_k) = -\frac{1}{2}N \sum_{i=1}^k (\log(1 + d_i) + \log(1 - d_i)), \quad (2)$$

and  $\boldsymbol{\Theta}_k = \{d_1, d_2, \dots, d_k, \mathbf{E}_{1(1 \sim k)}, \mathbf{E}_{2(1 \sim k)}\}$  where  $\mathbf{E}_{1(1 \sim k)}$  and  $\mathbf{E}_{2(1 \sim k)}$  represent the first  $k$  eigenvectors in each transformation matrix.

Taking into account the orthonormality of the transformation matrices  $\mathbf{E}_1$  and  $\mathbf{E}_2$ , the penalty term is calculated as

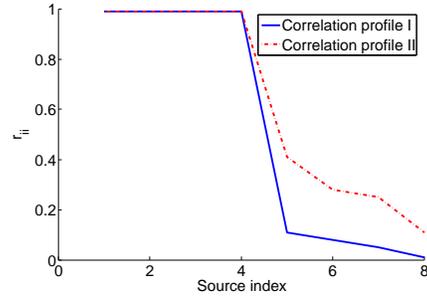
$$\mathcal{G}(\boldsymbol{\Theta}_k) = k + 2(pk - \frac{1}{2}k(k + 1)). \quad (3)$$

In case the data samples are dependent in the sample space, an i.i.d. subsampling scheme can be used to identify the set of independent samples to avoid over-estimation of the dimension [7].

## 4. SIMULATIONS

### 4.1. Experiment on simulated datasets

Eight sources are generated to simulate the independent features. To control the similarity of the features between the

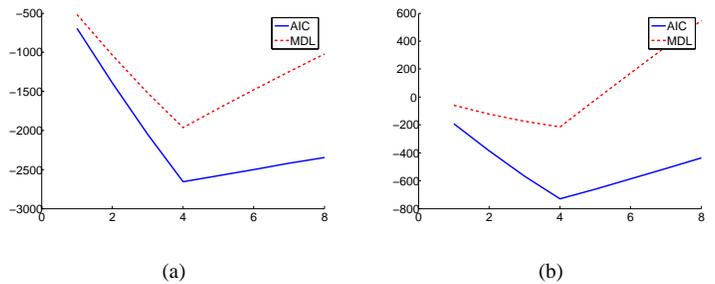


**Fig. 1.** Correlation profiles of the two sets of simulated sources

two datasets, the sources are altered for each dataset differently by adding Gaussian noise at controlled levels. The altered sources are, thus, used as the true sources to generate the datasets and to evaluate the estimation results.

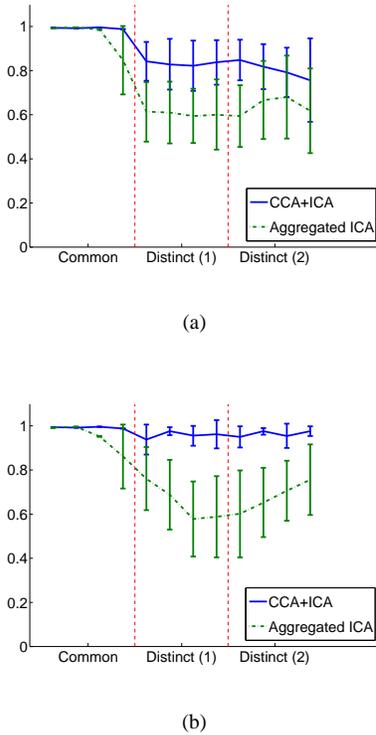
Four of the sources are slightly altered for each dataset to simulate the common features, while the other four sources are significantly altered for each dataset to simulate the distinct features. The sources for each dataset are mixed by different random mixing matrices, resulting in two observed multidimensional datasets. Two cases with different correlations settings between the corresponding features are simulated with the profiles of the correlations shown in Figure 1.

Both AIC and MDL scores are calculated according to Eqs. 2 and 3, with the i.i.d. sampling scheme to address the sample dependence. The curves of AIC and MDL criteria at different candidate orders are shown in Figure 2.



**Fig. 2.** Calculated AIC and MDL curves for simulated datasets (a) dataset with correlation profile I in Figure 1, and (b) dataset with correlation profile II in Figure 1

From the order selection result, it can be seen that both AIC and MDL criteria select four CVs as the common dimensionality of the two datasets, which agrees with the gen-



**Fig. 3.** Mean and standard deviation of the correlation between true and estimated sources for (a) dataset with correlation profile I in Figure 1 and (b) dataset with correlation profile II in Figure 1

erative model.

We further perform ICA on the concatenated common CVs to estimate four independent components and apply individual ICA on each set of distinct CVs to estimate the independent components distinct to each dataset. Twenty Monte Carlo trials are performed with different realizations of the altered sources. For comparison, ICA is applied on the concatenation of two datasets without CCA and order selection on the common dimensions. For both methods, the correlation between the true and estimated sources are evaluated with the mean and standard deviation values plotted in Figure 3. It can be observed that performing ICA in the properly partitioned common and distinct subspaces using the proposed method significantly improves the estimation of the common and distinct features in each dataset.

#### 4.2. Experiment on fMRI data from a simulated driving experiment

The fMRI data is acquired during a simulated driving experiment. Three tasks, asterisk fixation, driving, and watching are alternatively repeated during three epochs with 20 sec-

onds for fixation, 60 seconds for driving, and 40 seconds for watching. The experiment starts with a 15 seconds of fixation.

To study the difference of the brain activity during driving and watching, fMRI data of 40 time points is selected from each of the driving and watching epochs for analysis. Based on the order selection results for the common dimension of the driving and watching datasets using ITC formulations, common features are estimated from the concatenation of the first 18 highly correlated CVs of driving and watch epochs. Accordingly, 22 distinct features are estimated from the corresponding distinct CVs within each epoch. ICA on concatenation of the two datasets without CCA and order selection is also performed for comparison.

It is observed that three types of brain activations are estimated by both methods: (i) the activation on the posterior cingulate (related to the attentional tasks), (ii) the occipital lobe (primary visual area), and (iii) the cerebellar and motor cortex (motion related). Among those activations, (i) and (ii) are estimated as the common features while (iii) is estimated as a distinct feature of the driving epoch by the proposed method. The results agree with the recent study on a similar fMRI experiment [8].

As examples, the estimated activations of (ii) and (iii) are shown in Figure 4. For (ii), similar activation and time course in the visual cortex are estimated by the two different methods. For (iii), although both methods obtain activations at the motor cortex, the activation regions estimated by ICA without CCA contains spurious activation regions that are irrelevant to motor cortex (Figure 4.g). Since the activation at the motor cortex is estimated as a distinct feature of the driving epoch using the proposed method, its time course (Figure 4.f) assumes a similar trend as the first 40 time points of the time course estimated by ICA without CCA (Figure 4.h). The last 40 time points of the time course in Figure 4.h shows less variance, which indicates that the motor cortex is less active during the watching epoch compared to the driving epoch.

The observation suggests that the proposed CCA-ICA method with the selection of common dimensionality by ITC provides an explicit representation of the common and distinct features of two datasets. The proposed approach better addresses the comparison of two datasets that assume both common and distinct features than performing ICA on a simple concatenation of two datasets.

## 5. DISCUSSION

In this work, we propose a framework to combine PCA, CCA and ICA for the analysis of two multidimensional datasets and develop the ITC formulations to select the common dimensionality of the two datasets. The practical formulations of ITC, for a multidimensional dataset, are devel-

oped in the context of detecting the number of signals in noise [9], where both the signals and the noise are modeled by multidimensional complex stationary Gaussian random processes and the noise are assumed to have homogeneous variance. We derive the ITC formations based on the jointly multivariate Gaussian assumption of the CVs and calculate the likelihood for a given model order by ignoring the insignificant canonical correlations. The proposed method optimizes the multivariate analysis tools such as PCA and ICA in comparison of two datasets with common and distinct features.

The relationship between the canonical correlation coefficients and the eigenspectrum of the aggregated CVs provides an insight into the function of PCA on the concatenation of two datasets. That is, the common attributes among the datasets are augmented by PCA, resulting in a set of large eigenvalues and a set of small eigenvalues of the covariance matrix for the concatenated data; the distinct attributes remain as individual principal components with eigenvalues close to unity due to the whiteness of principal components.

The proposed model can be extended to the group analysis of  $M$  datasets for comparison of common and distinct features. In general, CCA can not be directly applied in this case to achieve a joint diagonalization on the cross-covariance matrices of all the pairwise combinations from  $M$  datasets. Though, the method proposed in Section 2 can be extended to achieve an approximate diagonalization, given a proper measure on the diagonality of the cross covariance matrices.

Lukic *et al.* propose a similar data model that incorporates common and specific components from different datasets and apply the ICA algorithm utilizing the temporal autocorrelations of fMRI time sequences to achieve source separation [10]. The dimensionality of common components is determined from the estimated linear minimum mean square error prediction matrix of the fMRI time sequences. However, threshold value is required to determine this dimensionality when data do not strictly follow the specified common-specific component model. Furthermore, all the sources are required to have different autocorrelations at the selected time lags in order to obtain the unique separation matrix in the assumed model. This requirement is, however, not always satisfied in real fMRI data analysis.

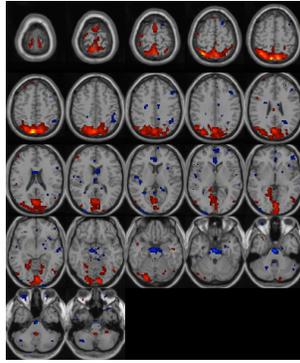
A similar model is defined in the work by Akaho *et al.* [11], but without a partition of common and distinct components. In their method, the separation of the sources is solved by an augmented objective function that maximizes the mutual information (for dependence of corresponding sources from two datasets) as well as minimizes Kullback-Leibler divergence between joint and marginal source densities (for independence of sources within each dataset). In contrast, we take the advantage of CCA for subspace par-

tion and use ICA to compensate for the source separation ambiguity by CCA due to identical canonical correlation coefficients.

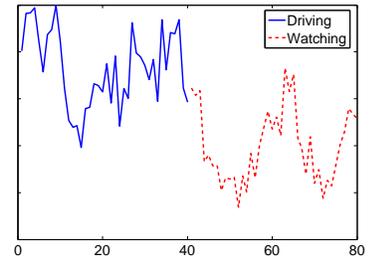
It is worth noting that the data structure assumed in the subspace ICA method [12] is similar but different from the model proposed in this work. Subspace ICA assumes dependence within each dataset and independence among the datasets. The objective is to obtain sparse feature-space representation of the given data such as natural image. In our proposed framework, the features are assumed to be independent within each dataset and dependent between the datasets due to common features.

## 6. REFERENCES

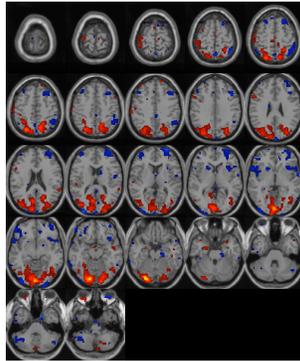
- [1] V. D. Calhoun, T. Adalı, J. J. Pekar, and G. D. Pearlson, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, pp. 140–151, 2001.
- [2] A. C. Rencher, *Methods of Multivariate Analysis (Second Edition)*, Wiley-Interscience, 2002.
- [3] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd International Symposium on Information Theory 1973*, Tsahkadsor, Armenian SSR; Hungary.
- [4] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [5] G. Schwartz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [6] T. W. Anderson, "Asymptotic theory for principal component analysis," *The Annals of Mathematical Statistics*, vol. 34, 1963.
- [7] Y.-O. Li, T. Adalı, and V. D. Calhoun, "Estimating the number of independent components in fMRI data," In press, *Human Brain Mapping*, 2007.
- [8] V. D. Calhoun, J. J. Pekar, V. B. McGinty, T. Adalı, T. D. Watson, and G. D. Pearlson, "Different activation dynamics in multiple neural systems during simulated driving," *Human Brain Mapping*, vol. 16, pp. 158–167, 2002.
- [9] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 387–392, 1985.
- [10] A. S. Lukic, M. N. Wernick, L. K. Hansen, J. Anderson, and S. Strother, "A spatially robust ICA algorithm for multiple fMRI data sets," in *Proc. IEEE International Symposium on Biomedical Imaging (ISBI) 2002*, Arlington, VA.
- [11] S. Akaho, Y. Kiuchi, and S. Umeyama, "MICA: multimodal independent component analysis," in *Proc. IJCNN 1999*, Washington DC.
- [12] A. Hyvarinen and P. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, pp. 1705–20, 2000.



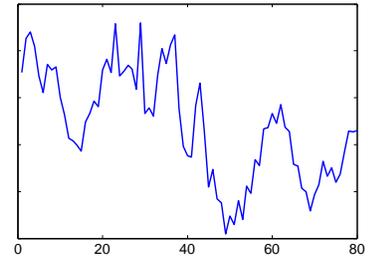
(a)



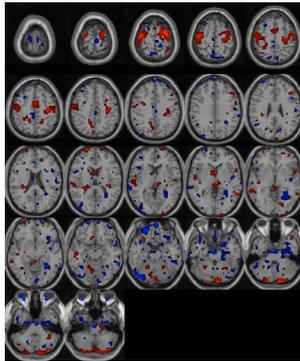
(b)



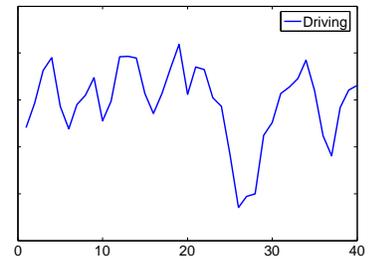
(c)



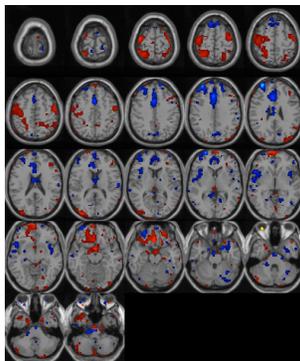
(d)



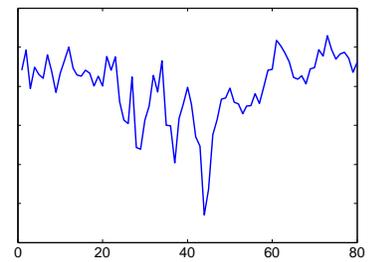
(e)



(f)



(g)



(h)

**Fig. 4.** Comparison of feature estimation in fMRI data from a simulated driving experiment: (a,b) Visual activation estimated as the common feature of driving and watching epochs by the proposed method; (c,d) Visual activation estimated by ICA on the concatenation of driving and watching epochs; (e,f) Motor activation estimated as the distinct feature of the driving epoch by the proposed method; (g,h) Motor activation estimated by ICA on the concatenation of driving and watching epochs.