

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Tsatsishvili, Valeri; Cong, Fengyu; Toiviainen, Petri; Ristaniemi, Tapani

**Title:** Combining PCA and multiset CCA for dimension reduction when group ICA is applied to decompose naturalistic fMRI data

**Year:** 2015

**Version:**

**Please cite the original version:**

Tsatsishvili, V., Cong, F., Toiviainen, P., & Ristaniemi, T. (2015). Combining PCA and multiset CCA for dimension reduction when group ICA is applied to decompose naturalistic fMRI data. In IJCNN 2015 : International Conference on Neural Networks. Institute of Electrical and Electronic Engineers. Proceedings of International Joint Conference on Neural Networks. <https://doi.org/10.1109/IJCNN.2015.7280722>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Combining PCA and Multiset CCA for Dimension Reduction when Group ICA is Applied to Decompose Naturalistic fMRI Data

Valeri Tsatsishvili<sup>1</sup>, Fengyu Cong<sup>1,2</sup>, Petri Toivainen<sup>3</sup>, Tapani Ristaniemi<sup>1</sup>

<sup>1</sup>Department of Mathematical Information Technology  
University of Jyväskylä  
P.O. Box 35, FI-40014, University of Jyväskylä, Finland  
[valeri.v.tsatsishvili@jyu.fi](mailto:valeri.v.tsatsishvili@jyu.fi)

<sup>2</sup>Department of Biomedical Engineering  
Faculty of Electronic Information and Electrical Engineering  
Dalian University of Technology, China

<sup>3</sup>Department of Music  
University of Jyväskylä  
P.O. Box 35, FI-40014, University of Jyväskylä, Finland

**Abstract**—An extension of group independent component analysis (GICA) is introduced, where multi-set canonical correlation analysis (MCCA) is combined with principal component analysis (PCA) for three-stage dimension reduction. The method is applied on naturalistic functional MRI (fMRI) images acquired during task-free continuous music listening experiment, and the results are compared with the outcome of the conventional GICA. The extended GICA resulted slightly faster ICA convergence and, more interestingly, extracted more stimulus-related components than its conventional counterpart. Therefore, we think the extension is beneficial enhancement for GICA, especially when applied to challenging fMRI data.

**Keywords**—Group ICA; temporal concatenation; naturalistic fMRI; dimension reduction; Multiset CCA; PCA

## I. INTRODUCTION

Naturalistic neuroimaging experiments are increasingly utilized by neuroscience community to study cognitive brain function [1-6]. In such experiments real-world experiences are reproduced in laboratory conditions, enabling more ecologically valid stimulation than in common controlled experimental paradigms. From the methodological perspective, analysis of naturalistic functional magnetic resonance imaging (fMRI) data involves new challenges associated with less control over stimulus and response timing, and lack of predefined models for such complex brain responses. That makes data-driven analysis approaches preferable over traditional hypothesis-driven ones due to their flexibility and few assumptions on the nature of data and noise. Independent component analysis (ICA) is arguably the most widely applied

data-driven method to fMRI data. The ICA model assumes the observed fMRI data to be a linear mixture of several independent and non-Gaussian sources, and extracts them from observations. Various extensions of ICA exist nowadays.

Group ICA (GICA) is commonly applied ICA variant on fMRI due to its advantage in drawing group inferences from multi-subject datasets. Within Group ICA, several approaches with different assumptions and data grouping strategies have been developed [1]. In this study, GICA with temporal concatenation was selected [7], since it was shown to outperform other strategies through the simulation study in [8]. Conventional GICA method [7] consists of three main stages: first, the dimensionality of data from each subject is reduced using principal component analysis (PCA). Second, the reduced datasets are concatenated across time dimension and PCA is applied again to further reduce dimensions. Third, the thus reduced dataset is subjected to ICA decomposition. As a result, one mixing matrix is estimated containing partitions corresponding to each subject, and one set of independent components (i.e. spatial maps) - that are common to all subjects. Spatial maps for each subject can also be reconstructed for exploring between-subject differences within each aggregate component. Although the described model has been successfully applied on fMRI before (see e.g. [1] for review), it does not always perform well, as demonstrated on our naturalistic fMRI data in the present study as well as in [9].

In this work we extended the described conventional GICA model by integrating multiset canonical correlation

analysis (MCCA) as an additional pre-processing step before subjecting data to ICA decomposition. Multiset CCA is the multiset generalization of standard two-set CCA introduced in [10], which finds correlated subspace from different but related datasets using second-order statistics [11]. In fMRI experiments finding common brain activation patterns among different subjects is important. Therefore, finding correlated subspace in the data before applying ICA is well justified for our purposes. The extension of blind source separation (BSS) by first finding correlated subspace was proposed in [12] and was reported to be significantly better than ICA alone in separating simulated mixtures. It should be noted, however, that the model we introduce here is not identical with the one applied in [12]. The main difference is in the ICA approaches employed. We selected group ICA with temporal concatenation strategy, whereas in [12] ICA was applied on each dataset separately. Furthermore, MCCA implementation in the present study is from [13], which in turn follows the original publication [11]. On the other hand, authors in [12] offered their own implementation of MCCA (or generalized CCA, as referred in the article). Nevertheless, the main principle of finding the correlated subspace in the data prior subjecting to ICA decomposition is the same. Interestingly, even though Karhunen et al. also included experiments on real fMRI data in [12], only two-set CCA was applied as source separation method rather than as a preprocessing of ICA. In this study, the developed GICA extension was applied on challenging fMRI data acquired during task-free continuous music listening experiment and the results were compared with outcome of conventional GICA.

A crucial question for any ICA-based approach is how many sources should be extracted from the observed data. Estimation of number of sources has tremendous influence on the success of the decomposition. The authors in [7] recommended commonly applied heuristics (such as selecting number of components explaining most of the variance in the data) as well as information-theoretic criteria for estimation of number of sources from the aggregate dataset. However, in the recent study Cong et al. [14] showed through simulations that popular information-theoretic methods, such as Akaike’s information criterion [15] and minimum distance length [16] fail to accurately estimate number of sources when the SNR is low. In fact, the suboptimal performance of estimators was later demonstrated in practice, where model order selection methods failed to estimate reasonable number of sources for GICA with temporal concatenation on the same naturalistic fMRI data as here [9]. To address this issue, we took a simple empirical approach for estimating the number of sources. Specifically, we examined GICA results for different number of sources to find the most appropriate number. Two general criteria were used for evaluating and comparing the results from different trials: first, the stability and reliability of ICA decomposition, and second, quality of the produced independent components. We devised six parameters, which provide concise description of ICA decomposition results according to the two criteria. The parameters, detailed in the section II.D, are simple and straightforward to compute from ICA decomposition results. The proposed evaluation heuristics can be adapted for testing and comparison of different ICA parametrizations in any application.

To summarize, we developed an extension of conventional GICA, where correlated subspace from all subject data is found prior concatenating and separating sources. The model was applied to challenging naturalistic fMRI data, and the results were compared to conventional GICA. In addition, we proposed a set of parameters that enables evaluation and comparison of different parametrizations in terms of the two criteria.

## II. METHOD

### A. Data description

The dataset analyzed in this study consists of fMRI scans of eleven healthy musicians (mean age: 23.2; SD: 3.7; 5 females), who listened to a 512 second-long piece of modern tango. The fMRI measurements were made in 3T scanner at sampling frequency of 0.5 Hz. Obtained fMRI scans went through common preprocessing routine, which is described in [6]. Overall, 231 fMRI scans corresponding to stimulus between 21 to 480 seconds were used for analysis. Six high-level features were obtained from stimulus audio, capturing timbral rhythmic and tonal information in music. The feature set, consisting of fullness, brightness, timbral complexity, pulse clarity, key clarity, and activity, were generated as a linear combination of 25 long-term and short-term features via PCA. The multi-stage process of their extraction and perceptual validation is described in [6].

### B. Group ICA extension

In this section, extension of GICA is introduced. For description of the conventional GICA algorithm with two-stage PCA reduction refer to [7]. The entire model can be divided into three-stage dimensionality reduction, followed by source separation by ICA. First, PCA separates signal and noise subspaces in each dataset. Next, MCCA is applied to select correlated subspace across all subjects’ responses. Particularly, it extracts an orthogonal set of canonical components from each subject data such that the canonical components are correlated across different datasets only on corresponding indices. Subsequently, the correlated subspace is extracted by selecting the canonical components that are correlated above the predefined threshold. The third stage dimensionality reduction is applied on concatenated data using PCA. The main reason is that number of signals in the concatenated data are commonly assumed to be larger than number of sources. In other words, we have the overdetermined data model, where the number of sources in the mixture is less than number of signals (or samples, statistically speaking). Therefore, prior subjecting to source separation, dimension reduction needs to be applied in order to produce determined mixtures with equalized sources and samples. Finally, BSS is achieved by ICA, extracting statistically independent sources in the concatenated data. The approach consists of the following steps:

Dimensionality reduction:

1. Reduce dimensionality of each dataset using PCA
2. Apply MCCA to select correlated subspace from each dataset
3. Concatenate reduced data from all subjects

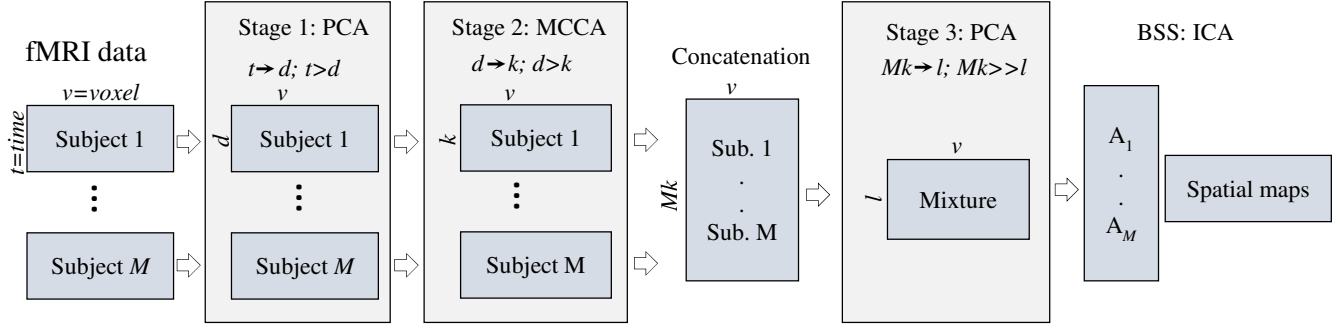


Fig. 1. Schematic view of the extended GICA model. Dimensions of data matrices after each stage are shown.

4. Reduce dimensionality of concatenated data using PCA

Blind source separation:

5. Apply ICA and get sources and unmixing matrix
6. Reconstruct time courses and spatial maps

Fig.1 depicts more detailed view of the entire processing chain. In the figure dimensionalities of data matrices after each step are shown for more clarity.

More formal description of the model is as follows. Let us denote dataset from each subject by  $\mathbf{X}_m \in \mathbb{R}^{t \times v}$ , where  $m = [1, 2, \dots, M]$  denotes number of subjects,  $t$  is number of fMRI scans and  $v$  is number of voxels. First, dimensions of each dataset is reduced using PCA. If we denote dimension reduction matrix (i.e. selected  $d$  eigenvectors) as  $\mathbf{V}_m \in \mathbb{R}^{t \times d}$ , then the reduced data for each subject (i.e. projection of data vectors to a new space) can be expressed as:

$$\mathbf{Z}_m = \mathbf{V}_m^{-1} \mathbf{X}_m \quad (1)$$

Note that pseudoinverse will be assumed in the cases when a matrix to be inverted is not square. Now  $\mathbf{Z}_m$  becomes an input for MCCA. Solving MCCA problem can be considered as finding orthogonal matrix  $\mathbf{B}_m$  and canonical components are obtained by projecting the input data to  $\mathbf{B}_m$ :

$$\mathbf{U}_m = \mathbf{B}_m^{-1} \mathbf{Z}_m = \mathbf{B}_m^{-1} \mathbf{V}_m^{-1} \mathbf{X}_m = \mathbf{D}_m \mathbf{X}_m \quad (2)$$

where,  $\mathbf{D}_m = \mathbf{B}_m^{-1} \mathbf{V}_m^{-1}$  and  $\mathbf{U}_m$  contains components, such that the canonical correlations, or correlation between corresponding components among different subjects are in decreasing order. The dimensions are reduced from  $d$  to  $k$  by selecting canonical components showing  $k$  highest mean correlations. Subsequently, datasets from all subjects are concatenated. The concatenated dataset  $\tilde{\mathbf{U}} \in \mathbb{R}^{Mk \times v}$  consists of  $M * k$  samples and is an overdetermined mixture. Then, dimensionality of the data is reduced using PCA to make the mixture determined:

$$\tilde{\mathbf{u}} = \mathbf{V}^{-1} \mathbf{u} \quad (3)$$

where,  $\mathbf{V} \in \mathbb{R}^{Mk \times l}$  is a dimension reduction matrix. Finally,  $\tilde{\mathbf{u}}$  is subjected to ICA decomposition to find sources of activations as independent components (IC) and mixing matrix, or temporal courses of IC-s. The model ICA follows

is  $\tilde{\mathbf{u}} = \mathbf{A}\mathbf{s}$ . In practice, ICA algorithm estimates the unmixing matrix  $\mathbf{W} = \mathbf{A}^{-1} \in \mathbb{R}^{l \times l}$  and IC-s such that:

$$\mathbf{y} = \mathbf{W}\tilde{\mathbf{u}} \quad (4)$$

According to ICA model applied here, aggregate unmixing matrix containing partitions unique to each subject, while the independent components are common. Reconstruction of temporal courses in the reduced data space is done by simply inverting the estimated unmixing matrix. However, we are interested in time courses in the original scan space. It can be achieved by inserting (3) in (4), partitioning aggregate data for each subject, and then inserting (2):

$$\mathbf{y}_m = \mathbf{W}_m \mathbf{V}_m^{-1} \mathbf{D}_m \mathbf{X}_m = \mathbf{W}_m \mathbf{X}_m \quad (5)$$

Here,  $\mathbf{y}_m$  denotes reconstruction of subject-specific spatial maps. Then, temporal courses can be estimated by:

$$\mathbf{T}_m = \mathbf{W}_m^{-1} = \mathbf{D}_m^{-1} \mathbf{V}_m \mathbf{W}_m \quad (6)$$

For ICA calculation FastICA algorithm was used [17].

### C. Analysis of extracted components by ICA

Within the temporal concatenation approach, each extracted common activation pattern may feature different temporal dynamics among subjects. The aim is to find the common maps temporally correlated with the stimulus. To this end, we first correlate subject-specific temporal courses of each IC with the time courses of the six stimulus features. The significance thresholds for correlations are estimated using Monte-Carlo simulation presented in [6] and the level of significance selected throughout this study is  $p < 0.01$ . If majority of the subject-specific temporal courses of a given IC are significantly correlated with the temporal course of any of the stimulus features, then the IC is considered to be stimulus-related and is selected for further analysis. At this point we reconstruct subject-specific spatial maps of the selected IC-s in order to observe individual differences in activation patterns. Ideally, we would have one or more similar activation patterns from all the subjects per acoustic feature. However, in practice similar spatial patterns from different subjects are not synchronized in time and show heterogeneous correlations with acoustic features, which makes them difficult to interpret.

TABLE 1. Evaluation parameters for MCCA+GICA vs conventional GICA (last entry in the table). The output dimensionalities of the first and the third stages (PCA) are fixed to 80 and 40 respectively, while the output of the second stage (i.e. MCCA) is varied from 10 to 60

MCCA dimensions	Converged runs	Convergence rate (mean/SD)	Cluster quality index (Mean/SD)	Selected independent components	Common maps (feature/sub)	Active voxels in common map (left/right hemisphere)
MCCA-10	66	82/11	0.64/0.21	37	-	-
MCCA-20	72	74/13	0.78/0.19	27	-	-
MCCA-40	96	64/17	0.9/0.09	24	-	-
MCCA-60	100	59/14	0.89/0.1	31	1 (Bright./6 sub)	2317/2283
No MCCA	99	65/14	0.9/0.11	29	-	-

#### D. Evaluation of different parametrisations

Two general criteria were used for evaluation and comparison of the results from different trials. The first criterion is stability and reliability of ICA decomposition. Stochastic nature of ICA-based decomposition renders direct comparison of results unreliable. To address the issue the software package ICASSO [18] was utilized. This tool has been designed for analysis of the stability and robustness of ICA decomposition. Essentially ICASSO runs ICA repeatedly  $N$  times ( $N=100$  in this study), each time with randomly initialized unmixing matrix, clusters the extracted independent components, and provides multiple parameters for observing and visualizing the clustering and separation results.

One interesting parameter in our list is ICASSO cluster quality index:

$$I_q = \frac{1}{|C_m|^2} \sum_{i,j \in C_m} \sigma_{ij} - \frac{1}{|C_m||C_{-m}|} \sum_{i \in C_m} \sum_{j \in C_{-m}} \sigma_{ij}$$

where  $C_m$  denotes the set of estimated independent components in the cluster  $m$ ,  $|C_m|$  is the size of the cluster,  $C_{-m}$  is the set of indices outside the cluster  $m$ , and  $\sigma_{ij}$  is an absolute value of mutual correlations between estimated independent components extracted in different runs. Cluster quality index characterizes compactness and separation of clusters and is a good measure for estimating stability of the extracted component as well as detecting possible overfitting.

We propose three parameters for assessing the stability and reliability of ICA decomposition: 1. number of converged runs (out of  $N=100$ ). If ICA decomposition does not converge, the estimation of mixing matrix is not reliable. 2. mean/standard deviation of convergence rate across  $N$  runs. Convergence rate is the number of steps required for convergence. 3. Cluster quality index.

Since a stable ICA decomposition does not guarantee good quality of produced independent components, it is necessary to quantify the quality of the output in order to make comparison between different ICA decompositions possible. Last three of the six parameters assess the quality of produced components and consist of: 4. Number of the selected components, 5. Number of the common maps, and 6. Number of the active voxels in a common map(s).

### III. RESULTS

For simplicity we denote conventional group ICA as simply GICA, and the extension introduced in this study - as MCCA+GICA. Furthermore, to differentiate between the two stage PCA reductions, the term individual PCA will be applied to the reduction of each subject's data, while group PCA will refer to dimension reduction of concatenated data.

We first investigated different parametrizations of GICA in order to select appropriate number of sources. To this end, the number of dimensions was consecutively fixed for one of the two PCA stages while varying the other, and examined the GICA output. For individual PCA we wanted to retain most of the variance data and reduced the dimension from 231 to 80, explaining about 88% of variance in average over all subjects. As the tests involving different numbers of dimensions showed, individual PCA did not have as major influence on ICA decomposition as group PCA. It is expected because individual PCA does not define number of sources to be extracted by ICA. However, it increases SNR by separating signal and noise subspaces, and therefore, intuitively will have more influence on quality of independent components. For group PCA, conversely, significant influence on ICA decomposition was observed. Intuitively, it is probably related to the fact that it directly determines the number of sources in the mixture and suboptimal estimation of sources has devastating effect on ICA decomposition. Indeed, in our experiments ICA decomposition always failed above certain number of sources, regardless of the settings of individual PCA, or MCCA. For GICA, none of the tested number of dimensions for each PCA stage led to finding common spatial maps. Hence, it was difficult to evaluate quality of the produced independent components except counting the number of selected ones. Nevertheless, the first three evaluation parameters were helpful for dramatically narrowing down the range of all possible number of sources to a few. As a result, we reduced data to 80 and 40 dimensions after individual and group PCA respectively. Next, the developed MCCA+GICA extension was applied and different sizes of correlated subspace were tested. Throughout the tests, individual and concatenated PCA reduction outputs were fixed to 80 and 40 respectively, since this parametrization produced the best results and we wanted any other parameters except MCCA to be similar among GICA and MCCA+GICA. Average canonical correlations corresponding to the selected

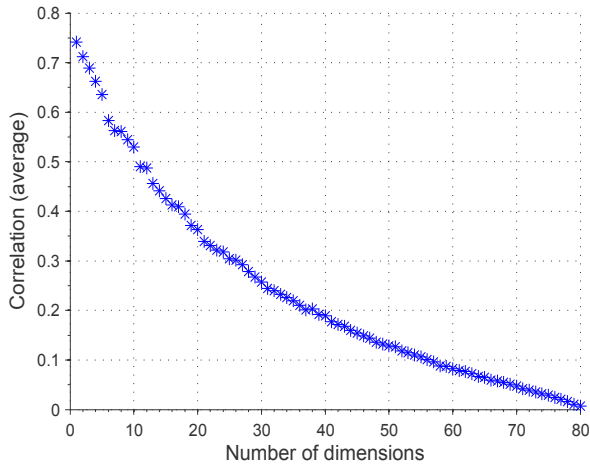


Fig. 2. Average correlations between canonical components

number of canonical components are shown in Fig.2. As depicted in the figure, the examined correlated subspace sizes (see Table 1) correspond to a wide range of thresholds for canonical correlations varying from fairly strict (above 0.5) to very liberal (above 0.1). In Table 1, six parameters summarizing MCCA+GICA output for different sizes of correlated subspace are shown. For comparison, GICA results for the same PCA outputs are also provided (labelled as 'No MCCA'). In terms of ICA stability, it is evident that increasing the size of correlated subspace improves ICA convergence (see Table 1). However, the rate of improvement quickly decreases, and for 40 and 60 components (MCCA-40 and MCCA-60 in the table) first three evaluation parameters are very similar. And yet, the parameters related to quality show more contrasting picture. For MCCA-40, none of the stimulus features were selected and common maps were not extracted, whereas for MCCA-60 two acoustic features including Brightness and Activity were selected, and common map related to Brightness was found with six contributing subjects. The common map in Fig.3 shows large clusters of bilateral auditory cortex activations, which is in line with previous findings [6, 9].

To summarize, the empirically selected parametrization of GICA after multiple tests was the following: individual PCA reduced dimensions from 231 to 80 and group PCA reduced aggregate data from 880 to 40. For MCCA+GICA: individual PCA reduced data from 231 to 80, MCCA reduced each dataset from 80 to 60, and group PCA – from 660 to 40. Based on our evaluation parameters (see Table 1), it can be seen that MCCA+GICA shows improvement over conventional GICA both in ICA convergence and quality of produced components.

#### IV. DISCUSSION

In this study we introduced MCCA-based extension of GICA, applied it on challenging fMRI data and compared the obtained results with the outcome from the conventional GICA. Interestingly, the latter failed to extract stimulus-related common spatial maps for any of the tested parametrizations.

The reason behind improved results by MCCA+GICA is that finding correlated subspace reduces complexity of

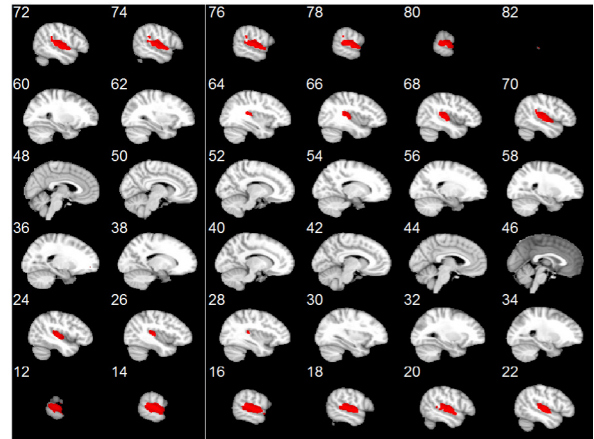


Fig. 3. Common spatial map, averaged across six contributing subjects

sources and therefore, has positive effect on inter-subject variability in terms of the temporal courses. As a result, for a given component, temporal courses from more than half of the subjects showed significant correlations with acoustic features. Reduced complexity of sources also explains faster convergence of ICA decomposition observed for the extended GICA model. In favour of this argument, Karhunen et al. [12] indicated that CCA alone already provides separation of sources at some degree using second-order statistics. Subsequently, in the next stage, partly separated sources form the new mixture of less complex sources in turn is decomposed further by ICA using higher-order statistics. Hence, benefits of introduced extension is expected and justified.

In addition, we proposed means for quantifying output of ICA-based method, in order to: *a)* estimate reasonable number of sources for ICA-based methods, should an automatic model order selection method fail. *b)* compare results from multiple different ICA parametrizations. A somewhat intuitive explanation of why evaluation parameters are useful for estimating number of sources stems again from the fact that suboptimal estimation of number of sources has strong influence on ICA decomposition. Over-estimated number of sources leads to complete failure of ICA decomposition, which will be easily reflected in first three evaluation parameters. Under-estimated number of sources, on the other hand, is primarily exhibited in reduced quality of produced independent components, and will probably be detected by last three parameters. It should be noted that the proposed set of evaluation parameters is still under the development. For example, parameters such as convergence rate and converged rounds can be improved by normalizing with respect to the number of sources.

Temporal concatenation approach assumes common spatial maps and individual time courses among subjects. The reconstructed subject-specific spatial activation patterns were very similar as expected, but within the standard model only part of the corresponding temporal courses were significantly correlated with the stimulus features. Large variability in temporal dynamics among subjects' responses elicited from a complex stimulation is presumably the major contributing factor to the failure of the conventional temporal

concatenation approach. In light of this, and with the supporting evidence in [9], perhaps spatial concatenation approach with the constraint of common mixing matrix for all subjects would be more appropriate for the specific cases such as the one considered here, although the opposite has been suggested previously [8]. We are planning to test the extension of spatial concatenation strategy with MCCA.

## V. ACKNOWLEDGMENT

This work was financially supported by TEKES (Finland) under grant 40334/10, Department of Mathematical Information Technology at the University of Jyväskylä, Academy of Finland (project 272250), the Fundamental Research Funds for the Central Universities [DUT14RC(3)037], XingHai Scholar in Dalian University of Technology in China, and National Natural Science Foundation of China (Grant No. 81471742). The first and the second author contributed equally to this work. The authors would also like to thank to Elvira Brattico and Asoke Nandi for their contribution.

## VI. REFERENCES

- [1] V. D. Calhoun, J. Liu and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *Neuroimage*, vol. 45, pp. S163-S172, 2009.
- [2] K. N. Carvalho, G. D. Pearlson, R. S. Astur and V. D. Calhoun, "Simulated Driving and Brain Imaging," *CNS Spectr.*, vol. 11, pp. 52-62, 2006.
- [3] E. Formisano, F. De Martino, M. Bonte and R. Goebel, "'Who' is saying 'what'? Brain-based decoding of human voice and speech," *Science*, vol. 322, pp. 970-973, Nov 7, 2008.
- [4] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, pp. 1634-1640, Mar 12, 2004.
- [5] K. N. Kay, T. Naselaris, R. J. Prenger and J. L. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, pp. 352-355, 2008.
- [6] V. Alluri, P. Toivainen, I. P. Jaaskelainen, E. Glerean, M. Sams and E. Brattico, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *Neuroimage*, vol. 59, pp. 3677-3689, Feb 15, 2012.
- [7] V. Calhoun, T. Adali, G. Pearlson and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum. Brain Mapp.*, vol. 14, pp. 140-151, 2001.
- [8] V. J. Schmithorst and S. K. Holland, "Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data," *Journal of Magnetic Resonance Imaging*, vol. 19, pp. 365-368, 2004.
- [9] F. Cong, T. Puoliväli, V. Alluri, T. Sipola, I. Burunat, P. Toivainen, A. K. Nandi, E. Brattico and T. Ristaniemi, "Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis," *J. Neurosci. Methods*, vol. 223, pp. 74-84, 2014.
- [10] H. Hotelling, "Relations between two sets of variates," *Biometrika*, pp. 321-377, 1936.
- [11] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433-451, 1971.
- [12] J. Karhunen, T. Hao and J. Ylipaavalniemi, "A generalized canonical correlation analysis based method for blind source separation from related data sets," in *Neural Networks (IJCNN), the 2012 International Joint Conference On*, 2012, pp. 1-9.
- [13] Y. Li, T. Adali, W. Wang and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *Signal Processing, IEEE Transactions On*, vol. 57, pp. 3918-3929, 2009.
- [14] F. Cong, A. K. Nandi, Z. He, A. Cichocki and T. Ristaniemi, "Fast and effective model order selection method to determine the number of sources in a linear transformation model," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012, pp. 1870-1874.
- [15] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions On*, vol. 19, pp. 716-723, 1974.
- [16] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.
- [17] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, pp. 1483-1492, 1997.
- [18] J. Himberg, A. Hyvarinen and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage*, vol. 22, pp. 1214-1222, Jul, 2004.