Investigating the Reliability of ICA Sources Obtained After PCA Preprocessing

D. Cordes[1], R. R. Nandy[1]
[1]University of Washington, Seattle, Washington, United States

Introduction

Independent Component Analysis (ICA) is a promising technique for analyzing fMRI data [1]. Unfortunately, the size of fMRI datasets often renders this technique as being computationally intractable, and certain compromises must be made to perform the analysis. One such compromise is to project the full dataset onto a lower-dimensional subspace that, in some well-defined sense, captures the essence of the data. The most common method for achieving this aim is Principal Component Analysis (PCA). PCA is used for two reasons in ICA, for computational efficiency and to prevent ICA from splitting components into multiple groups. In our investigation we focus on the first aspect by using pseudo real and simulated data. It is demonstrated herein, that PCA reduction in ICA can lead to both better and worse source estimates (compared to unreduced data) depending on the relative amplitude of the activation.

Methods

Six normal volunteers in their twenties participated in the fMRI experiment. All scans were done on a GE 1.5 Tesla MRI scanner with the parameters: 20 slices, coronal acquisition, 64 x 64 resolution, 7mm slice thickness, 2mm gap, TR 2 s, FOV 24 cm, bandwidth, ± 62.5 kHz, TE 40ms, flip angle 82°, 165 time frames. Standard resting-state data and motor activation data (30 second on/off task, repeated 5 times) were collected. ICA was carried out using FastICA with and without PCA preprocessing [2]. Two type of simulations were performed. In the first simulation, an ICA motor component (obtained from an activation study) with signal source intensity S_{ip}, where i labels the particular component and p the pixel, was added to the resting-state data set R_{tp} (t labels time) according to $X_{tp} = R_{tp} + cA_{ti}S_{ip}$ where A_{ti} specifies the element of the mixing matrix of source S_{ip} and c is an arbitrary scaling constant that is adjusted to obtain different relative amplitudes of the simulated motor activation. Median values among all 20,000 pixel intensities used for the relative amplitudes were 0.5%, 1%, and 2%. The simulated data set $\{X_{tp}\}$ was then reduced by PCA to different dimensions (161, 150, 100, 80, 60, 50, 40, 30, 20) and analyzed by ICA. The dominant component with maximum overlap to S_{ip} was then found and compared using modified ROC methods. The advantage of using resting-state data is that the constructed data set has automatically the proper spatial dependence, noise characteristic and temporal autocorrelation structure. In another simulation we investigated the dependence of the ICA component extracted (after PCA preprocessing) as a function of the shape of the source distribution. We assumed that the source activation pattern can be parameterized by an exponential power family of the form $p_X(x) = C\,e^{-\frac{|x|^\alpha}{\beta}}$ where α is a free parameter describing the shape of the distribution. This particular parameterization can capture the supergaussian nature of the ICA motor component but lacks any skewness in the distribution [3]. For a fixed value of $\alpha \in [0.3,2.5]$ in steps of 0.1 we generated 20,000 independent identically distributed (iid) points from the distribution The iid assumption of the spatial distribution is not valid in real fMRI data. However, the spatial dependency is completely ignored in the current formulation of ICA for fMRI and therefore does not affect our simulations. The generated numbers represent a spatial intensity pattern (similar to S_{ip}) that was then multiplied with a hypothesized hemodynamic response function consisting of a convolved boxcar function with six seconds delay (similar to A_{ti}), and added to resting-state data as before. The amplitude of the hemodynamic response function was varied to achieve median relative amplitudes of 0.5%, 1%, and 2%, respectively.

Results and Conclusion

To obtain accurate ICA sources using PCA reduction in a preliminary step, the strength of the signal sources and their spatial distribution play a major role. Indeed, in both the pseudo-real and simulated data, PCA-preprocessed ICA can be beneficial if the signal is strong (≥ 1% relative amplitude) but can also fail to detect all the activations associated with the paradigm when the dimension was reduced too aggressively (< 50) or the relative amplitude was low (< 1%). In this case, when the signal is weak, PCA reduction may not capture enough of the observed variance to extract components corresponding to highly localized brain activation from very noisy data as such components explain only a small fraction of the total observed variance across the brain.

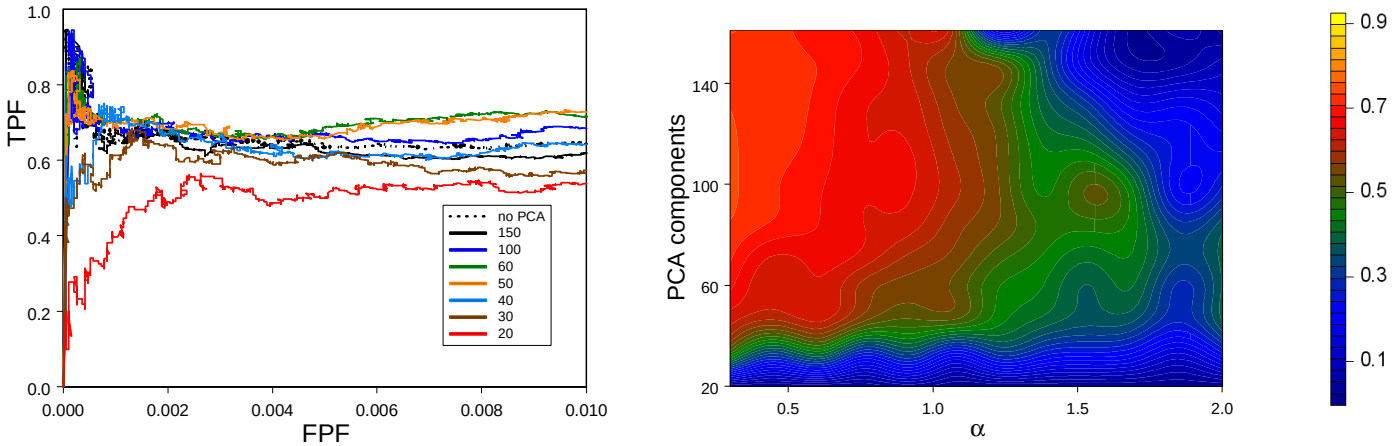

Fig.1. Modified ROC curves for an ICA source using simulated data processed by ICA with different degrees of PCA reduction for 1% activation (left: resting-state plus activation pattern from real motor data, right: TPF at FPF=0.005 for resting-state plus an activation pattern from an exponential power family parameterized by α.

References

[1] McKeown et al. Analysis of fMRI data by blind separation into Independent Spatial Components. 1998. Human Brain Mapping **6**:160-188.
[2] Hyvärinen et al. 1997. A fast fixed-point algorithm for Independent Component Analysis. Neural Computation **9**(7):1483-1492.
[3] Stone et al. 2002. Spatiotemporal ICA of Event-Related fMRI Data Using Skewed Probability Density Functions. Neuroimage **15**, 407-421.