On model complexity estimation in non-square ICA of fMRI data

R. Mutihac[1]

[1]Dept. of Electricity & Biophysics, University of Bucharest, Bucharest, Romania

Objective

Independent component analysis (ICA) is a data-driven multivariate exploratory analysis based on the covariance paradigm, which is able to estimate the functional activity without reference to the experimental protocol and potentially may reveal new features in the data. Non-square ICA model alleviates data overfitting and allows statistical assessment of the estimated independent components (ICs) though inevitably deviates from the best linear fit. The purpose of our study was to estimate the dimensionality of the data model subspace in compliance with a robust structural measure due to Friedman [1] in projection pursuit.

Methods

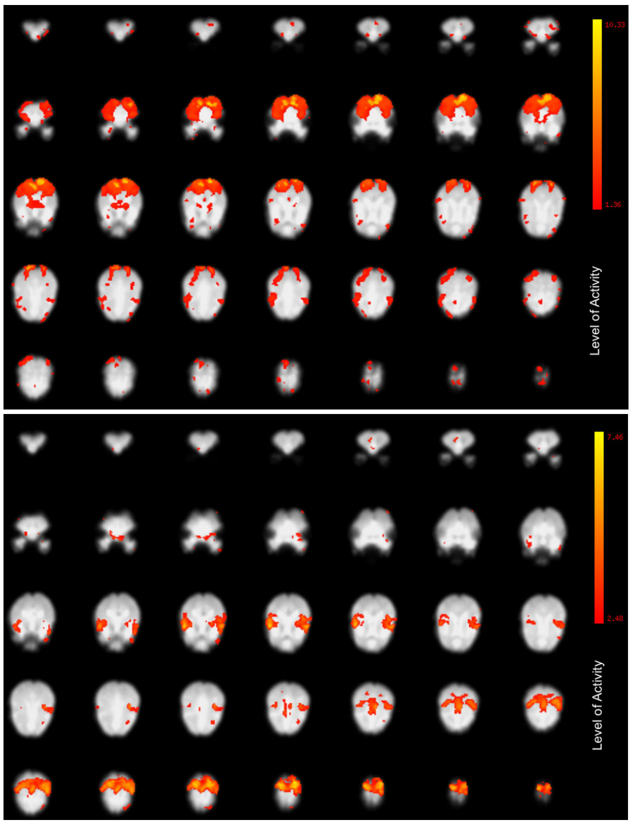
The model subspace in non-square ICA of fMRI data has no relationship to the number of timepoints; rather we assumed that there are more time points than sources of brain activity; a reasonable assumption in most recent fMRI studies. ICA is not endowed with internal means to estimate the optimal subspace that encompasses most of the non-Gaussianity in data, though PCA is commonly run prior to ICA unmixing to reduce dimensionality while retaining maximum of variance in data. The structural index used here is the distance from normality of a scalar random variable (distribution of the voxel values in a component) expressed as an integral-squared metric which is symmetric, rather than Kullbach-Leibler divergence which is not. Friedman [1] suggested a moment approximation for the integral, while we followed an iterative empirical evaluation of the inverse of the cumulative density function for each estimated IC. The monotony of the structural index estimation ensures less sensitivity to noise; the larger its value, the more structural information of data the respective component contains. In its plot versus the ICs, we selected the abscissa of the steepest flexing point as an estimation of the data model dimensionality and re-run non-square ICA; the difference between data and the estimated ICs was treated as an error term for statistical significance assessment of decomposition.

Results

Full-brain fMRI data originated from one subject performing a visuo-motor task in one session of *12* runs of *4.5* blocks each. Data were fully pre-processed in SPM2 and plugged in (spatial) FastICA [2]. Applying our procedure, data were confined to *15* components among which *2* of them displayed activity predominantly in the visual cortex (Fig. 1-*up*) and in the motor cortex (Fig. 1-*mid*), respectively. Their associated activities were correlated with the model time course of the experimental paradigm generated by Waver [3] with $r > 0.7$ (Fig. 1-*bottom*). The dimensionality that resulted was compared with common measures of information content like Akaike's information criterion (AIC), minimum description length (MDL), and probabilistic ICA [3].

Conclusions

Non-square ICA separates components in terms of structural information which are more likely to admit interpretations that are neurophysiological meaningful since they presumably correspond to distinct physical or physiological processes. Comparisons with real life fMRI data indicated the Friedman's index as a trade-off between MDL, statistically consistent but sensitive to spatial smoothness, and the more conservative AIC, and in good agreement with probabilistic ICA.



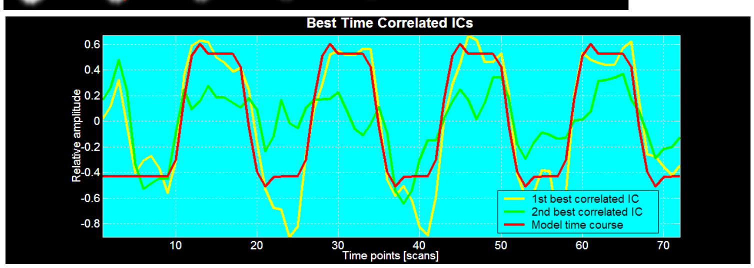


Fig. 1. *Best time correlated IC (up), second best time correlated IC (mid), and their associated time courses of activation (bottom).*

References

[1] J.H. Friedman, Exploratory projection pursuit, *J. Am. Stat. Assoc.*82(397):249-266, 1987.

[2] A. Hyvärinen and E. Oja, A fast fixed-point algorithm for ICA, *Neural Comput.*9:1483-1492, 1997.

[3] C.F. Beckmann and S.M. Smith, Probabilistic ICA for fMRI, *IEEE Trans. on Medical Imaging*, 23(2):137-152.