A Comparative Study on the Results of Algorithms Implementing Blind Source Separation of EEG Data

Surendra Pattanaik
Department of Electrical Engineering
Gandhi Institute for technological Advancement, Bhubaneswar, Odisha, India

Abstract- Independent component analysis (ICA), applied to EEG data, has proven capable of separating artifacts and brain sources. Of the variety of ICA and blind source separation algorithm now available, which are more efficient at processing EEG data? Here we defined efficiency to mean blind separation of the data to into near "dipolar" components having scalp maps consistent with synchronous activity in a single cortical region, here 20- ICA algorithms as well as PCA whitening , and PROMAX decomposition is applied to 71- channel data from 14 subjects and ranked the resulting decomposition by the number of near dipolar components they identified, by this measurement Infomax and Pearson ICA ranked highest though similar near dipolar components are returned by most of the ICA based algorithm.

Keywords –EEG, ICA, Blind Source Separation, PROMAX

I. INTRODUCTION

For separating artifacts from EEG data ICA is widely used. ICA can separate brain EEG activity from non brain artifacts [1, 2]. ICA able to separate signals from multichannel data, whose time course are maximally independent of each other. Since local (<100μm) connection is vastly denser in the cortex than long range connection, long range synchronization of cortical field activities should be much weaker than short range synchronization and the far field EEG signal recorded on the scalp should arise within compact cortical patches. In practice ICA decomposition may return 20 or more components whose scalp maps are compatible with generation in such a patch [4, 6].

It still remains unclear, however which ICA algorithm returns the most dipolar components when applied to EEG data. The three ICA algorithms most often applied to process EEG data are Infomax ICA[7],SOBI[8], and Fast ICA[9]. However, there are a large variety of other available ICA algorithms that may be possible for EEG decomposition. All ICA algorithms have the same overall goal[10] and generally produce near identical results when applied to near idealized(models) source mixture. However, since their approaches to independence differ and since EEG brain and non brain source signals are likely not perfectly independent, the different ICA algorithm may likely to return different results when applied to same EEG data.

II. MATERIAL AND METHODS

A. EEG data used for testing

Fourteen subjects (7 male, 7 female) participated in the study [5]. A series of 8 consonant letters were presented visually at screen center, 3-7 of which were black (to be memorized) the rest green (to be ignored). A central fixation symbol was presented for 5 sec at the beginning of each trial. A series of letters was then presented for 1.2 sec each with 200 ms gaps. Following this, a dash appeared on the screen for 2-4 s to signal a memory maintenance period during which the subject had to retain the sequence of memorized letters until a (red) probe letter was presented. The subject then had to press one of two buttons with their dominant hand (index finger or thumb) to indicate whether or not the probe letter was part of the memorized letter set. Auditory feedback 400 ms after the button press informed the subject, whether their answer was correct or not (Fig 1). The next trial began when the subject pressed another button. Each subject performed 100-150 task trials.

EEG data were collected from 71 channels (69 scalp and two periocular electrodes, all referred to right mastoid) at a sampling rate of 250 Hz with an analog pass band of 0.01 to 100 Hz (SA Instrumentation, San Diego). Input impedances were brought under 5 kΩ by careful scalp preparation. Data were analyzed by custom Matlab scripts built on the open source EEGLAB toolbox [6]. Continuous data were first high-

pass filtered above 0.5 Hz using an FIR filter. Epochs were selected 0.7 s before and after each letter presentation in the experiment (memorize, ignore, and probe). The mean channel values were removed from each epoch, and between 1 and 16 noisy data epochs were removed prior to ICA decomposition. Criteria for epoch removal were high-amplitude, high-frequency abnormalities such as those accompanying coughs, sneezes, jaw clenching, etc. The number of data samples in each dataset was about 250,000.

B. ICA algorithms

We used a total of 23 linear decomposition algorithms, 20 ICA algorithms plus principal component analysis (PCA), whitening/Sphering, and Promax [11]. We downloaded Matlab code for most of the algorithms from the Internet (see Table 1). The ICA algorithms all performed complete decomposition in which the number of returned components is equal to the number of channels:

$$S = WA \qquad (1)$$

Where *A* is the data matrix of size, number *of channels* by *number of time points*, *W* is an unmixing matrix of size, number *of ICA components* by *number of channels*, and *S* is the ICA component activation time courses of size *number of ICA components* by *the number of time points*.

ICA learns the unmixing weight matrix that makes the component time courses as temporally independent from each other as possible. However, the approach of each ICA algorithm to estimating and/or approaching this independence is different. Extended Infomax [10], Infomax [7], Pearson ICA [12], and ERICA [13] belong to the class of natural gradient algorithms [14], differing only in the way they estimate the component probability distributions. SOBI [8] is a second-order method that takes advantage of temporal correlations in the source activities. SOBI, SONS, AMUSE, icaMS, FOBI, EVD, and EVD 24 all use time delay covariance matrices [15]. Other algorithms, such as so-called FastICA [9], maximize the negentropy of their component distributions or their fourth order cumulant (JADE) [16]. For all these algorithms, we selected the default time delays (4) implemented in the downloaded software implementations. Possibly better ICA decompositions might be obtained in some cases with other choices. We refer the reader to more complete documentation [9, 15] for the implementation of each algorithm.

ICA differs from PCA in that it identifies sources of distinct information in the data, thereby relaxing PCA's spatial orthogonality constraint. Because it has been used to decompose EEG and ERP data (e.g. [17]), we also applied Promax, another non-orthogonal linear decomposition method that maximizes some higher power (such as the fourth power) of the projection of the data on each component axis. Finally, we included Sphering or whitening, often used as preprocessing before ICA decomposition. Sphering decorrelates the signals between all pairs of channel-centered components [7].

B. Method for testing ICA algorithms

After computing all 23 decompositions for each of the 14 EEG datasets, we localized a best-fitting single equivalent dipole corresponding to each component using a single equivalent dipole in a spherical 4-shell head model (radius: 71, 72, 79, 85 mm; conductances: 0.33, 0.0042, 1, 0.33 μS) in the DIPFIT plugin (version 1.02) of the EEGLAB toolbox. (Version 4.515) [6]. Peri-ocular (eye) channels were excluded from dipole fitting. Note that modeling each component maps with a single dipole is somewhat idealistic, since in particular some ICA components represent apparently bilateral synchronous source activities (see, e.g., Fig 2b of [4]). However, here such components appeared to be rare here, as in other decompositions of more than 32 channels.

We also computed a measure of independence between component maps based on a log likelihood function. This measure presents limitations as it is based on the assumption of a source activity probability distribution proportional to 1/Cosh(x). It thus should not be taken as an absolute measure of the independence of the returned component activities.

$$LL = -\left(\ln\left(|\det(w)| \right) - \Sigma\Sigma \frac{\cosh(s)}{N} - M \cdot \ln(\pi) \right)$$

Where *W* represents the weight matrix, *S* represents the ICA activities, *N, the* number of time points and *M* the

number of components

TABLE I
MEAN COMPONENT DIPOLARITY

Algorithm (Matlab func.)	D%	LL	Origin
Extended Infomax (runica)	29.9	178	EEGLAB 4.515
Pearson	29.1	169	ICAcentral (6)
Infomax (runica)	28.2	160	EEGLAB 4.515
ERICA	26.9	184	ICALAB 1.5.2
SONS	25.4	183	ICALAB 1.5.2
SHIBBS	23.7	169	ICAcentral (5)
FastICA*	23.5	169	ICAcentral (2)
JADE (jader)	23.4	169	EEGLAB 4.515
TICA	23.4	169	ICALAB 1.5.2
JADE optimized (jade_op)	21.4	169	ICALAB 1.5.2
JADE w/ time delay (jade_td)	20.2	169	ICALAB 1.5.2
eeA	19.0	305	ICAcentral (8)
Infomax (icaML) †	18.8	212	ICA DTU Tbox
FOBI	18.6	169	ICALAB 1.5.2
SOBIRO (acsobiro)	17.9	167	EEGLAB 4.515
EVD 24	17.7	169	ICALAB 1.5.2
EVD	17.0	169	ICALAB 1.5.2
SOBI	16.1	583	EEGLAB 4.515
icaMS†	10.6	169	ICA DTU Tbox
AMUSE	8.5	169	ICALAB 1.5.2
PCA	3.1	583	EEGLAB 4.515
Promax	33.7	467	EEGLAB 4.515
Whitening/Sphering	57.6	164	EEGLAB 4.515

* A symmetric approach to optimizing the weights for this algorithm returned similar results.
† No whitening by default.

Table 1: ICA component dipolarity for each algorithm. The first column gives the algorithms tested (and in parenthesis the Matlab function, when ambiguous). The second column (near-dipole, ND%) indicates the percentage of returned components having an equivalent dipole projection to the scalp with less than 10% residual variance from the component scalp map. The third column, log likelihood (LL), gives a rough estimate of component time course independence (2) (here lower numbers suggest more independence). The fourth column (Origin) indicates the online source. EEGLAB (here, former version 4.515) is available at sccn.ucsd.edu/eeglab, the ICALAB toolbox at www.bsp.brain.riken.go.jp/ICALAB/. The ICA DTU Toolbox is available at mole.imm.dtu.dk/toolbox/ica. An ICAcentral database giving links to code for these and other algorithms is available at www.tsi.enst.fr/icacentral. Numbers (in parentheses) to the right of the ICAcentral source label indicate the respective entry number in the ICAcentral database.

III RESULTS

A. Dipolarity of ICA components

Table 1 indicates, for each ICA algorithm, the percentage of equivalent dipoles whose scalp projection had residual map variance from the returned component map below 10%. (vertical line in Fig. 2) The cumulative numbers of equivalent dipoles at each residual variance for each algorithm are represented in Figure 1.

Extended Infomax, implemented in EEGLAB runica, returned the largest number of such components among all ICA algorithms, although Pearson, simple (super-Gaussian) Infomax, and ERICA gave near-equivalent results. All these algorithms use natural gradient descent [14]. Two implementations of infomax ICA (Infomax and icaML) did not perform the same. The Infomax version (runica) in the EEGLAB toolbox was been specifically developed to process large amount of EEG data. icaML does not use whitening by default, whereas EEGLAB Infomax does. Also, Pearson ICA uses the same optimization procedure as FastICA, yet returned more dipolar components. These differences point to the importance of details of the method used to separate ICA components.

SOBI, another widely used algorithm for decomposing EEG data, and the other time-dependent algorithms did not return as many dipolar components as the natural gradient-based algorithms. As mentioned previously, some other choices of time delays for these algorithms might change their results. Interestingly component time courses returned by SOBI were far from maximally independent by our log likelihood (LL) measure, which returned a value equal to the one obtained for PCA. However, SOBIRO, a variant of SOBI using a robust orthogonalization method (Section of 4.3.1 [15]) returned a LL value similar to other ICA algorithms.

As expected, Principal Component Analysis returned only a small number of dipolar components (those accounting mostly for large eye artifacts). Promax returned a quite high number of near-dipolar components, but visual inspection revealed that their maps looked quite similar, all seeming to account in part for eye movement artifacts. The mean log likelihood measure for Promax decomposition (467) was, as expected, also considerably higher than those for the ICA decompositions.

Like the ICA algorithms, whitening or sphering also returned many dipolar components and a low mean log likelihood estimate. Whitening performs second-order decorrelation whereas ICA algorithms explicitly or implicitly optimize independence using higher-order statistics. Most ICA algorithms (see Table 1) use data pre-whitening as a starting point for ICA decomposition. Whitened components have stereotyped scalp maps consisting of a diffuse projection peaking at each respective data channel. This might only represent physiologically plausible projections of cortical EEG sources if the cortex were smooth and unfolded, whereas the largest portion of the human cortex is in its many folds or sulci. The low independence log likelihood value (164) for whitening was only about 4% higher than the LL values for Infomax and Extended Infomax). Therefore, Infomax and other ICA decomposition approaches that begin with pre-whitening actually reduce their number of dipolar component maps during training, as these maps more flexibly adapt to the spatial projections of the brain and non-brain sources of information in the data.

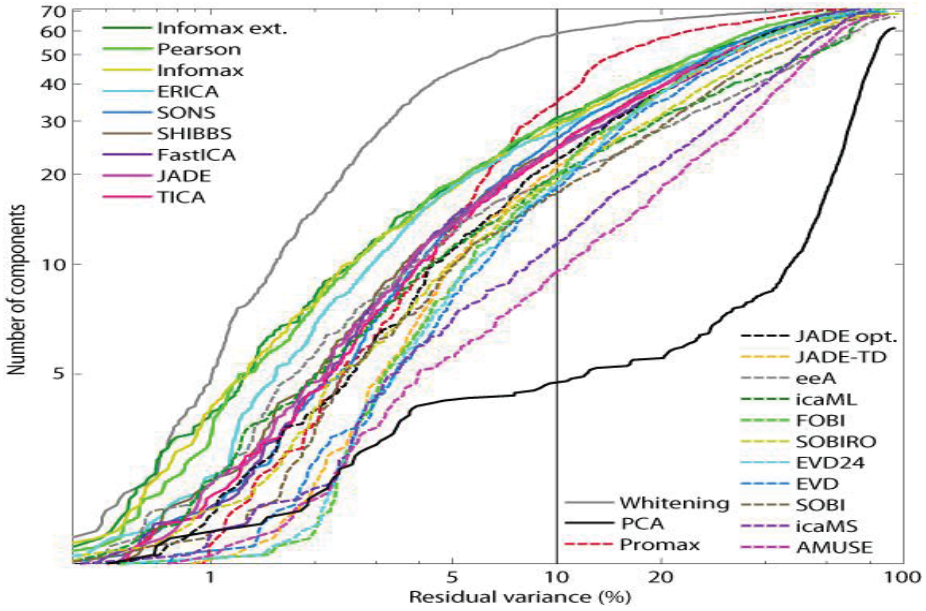

Fig. 1. Number of bipolar component returned as a function of residual variance for each algorithm. Since there were 71 channels in each data set, the ordinate indicates the cumulative number of components at each residual variance. The algorithms are ranked according to their intersection with the vertical line at 10% of residual variance (see Table 1).

B. Comparison of ICA algorithms

We compared ICA algorithms using the Amari measure [18] of distance between weight matrices. Although Amari distance has typically been used in simulations to measure the deviation of a given ICA decomposition from ground truth, it may also be used to compare any set of weight matrices from ICA decompositions.

We computed the pairwise distance between all pairs of decompositions for all datasets, took the means across data sets, projected the resulting 23-by-23 matrix into its principal 3-dimensional subspace using PCA, and then rotated this subspace using Infomax (tunica) for visualization.

Figure 2 represents the axes of this projection. The ICA algorithms were clustered into two groups (circled manually in Figure 2A), the first group consisting of the instantaneous ICA algorithms which all returned relatively large numbers of near-dipolar components with low residual variance. The second group consists of algorithms that depend on time-domain relationships. These returned a lower number of near-dipolar components (see Table 1). In this second group, two subgroups (left and right) can be discerned.

The instantaneous ICA algorithms that returned the fewest near-dipolar components, icaML and eye, stand apart in Fig. 2; eye and icaML also had an unusual log likelihood (Table 1). Unlike other ICA algorithms, icaML did not use Sphering. The results of PCA (Fig. 2) Are clearly isolated from those of the ICA algorithms. Results of Promax and whitening (Sphering) also appear relatively isolated.

Instantaneous ICA decompositions tended to return more near-dipolar components than blind decompositions using temporal differences. To test whether the most bipolar components returned by these two classes differed, we repeated the visualization using only the scalp maps of the 10 most bipolar components for each algorithm (Fig. 2B). Clearly, results of PCA, Whitening/Sphering, and Promax decompositions stand apart from the results of ICA and other blind source decompositions. A distinction between instantaneous ICA (clustered on the right) and blind source decompositions using time delays (clustered on the left) remains, with one noticeable exception – the results of the time-sensitive optimized SOBIRO algorithm closer to those of the instantaneous ICA decompositions. Overall, the most nearly dipolar components returned by all the ICA algorithms were similar, but differed from those of the other linear decompositions.

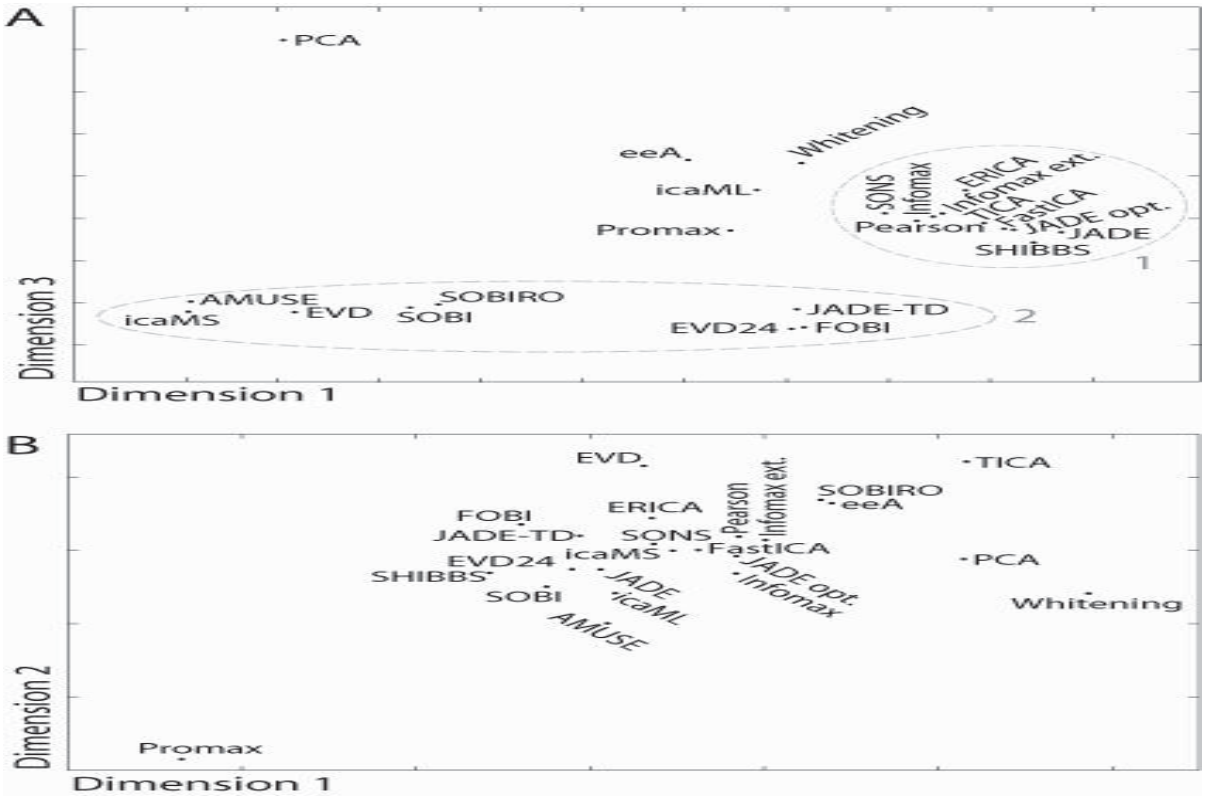

Fig. 2. A. Clustering of the results of all 23 decomposition using Amari distance between each pair of decomposition weight matrices. Infomax ICA (runica) applied to the 3-dimensional principal PCA subspace was used to project this multidimensional matrix onto axes of maximal interest for visualization. The two indicated decomposition clusters were selected visually. B. Visualizing relationships among the same decompositions considering only their 10 most dipolar components. Here PCA, sphere, and Promax clearly stand apart from the ICA decompositions.

IV.CONCLUSION

I have shown that some ICA and blind source algorithms return larger numbers of EEG components with nearly dipolar scalp maps than others. For decomposition of EEG data, timing-insensitive and timing-sensitive ICA or blind source separation algorithms return somewhat different decompositions that differ somewhat in the degree of dipolarity of their component maps. However, the most nearly dipolar components returned by the two decomposition classes do not appear to differ strongly. Results of blind source separation clearly differ, however, from component scalp maps isolated using the second-order variance-based methods PCA, whitening, and Promax.

REFERENCES

[1] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163-78., 2000.

[2] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clin Neurophysiol*, vol. 111, pp. 1745-58., 2000.

[3] S. Makeig, M. Westerfield, T. P. Jung, S. Enghoff, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, pp. 690-4, 2002.

[4] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends Cogn Sci*, vol. 8, pp. 204-10, 2004.

[5] J. Onton, A. Delorme, and S. Makeig, "Frontal midline EEG dynamics during working memory," *Neuroimage*, 2005.

[6] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods*, vol. 134, pp. 9-21, 2004.

[7] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput*, vol. 7, pp. 1129-59., 1995.

[8] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electronics Letters*, vol. 36, pp. 2050-2053, 2000.

[9] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw.*, vol. 13, pp. 411-30, 2000.

[10] T. W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A Unifying Information-theoretic Framework for Independent Component Analysis," *Comput. Math. Appl.*, vol. 31, pp. 1-21, 2000.

[11] A. E. Hendrickson and P. O. White, "A quick method for rotation to oblique simple structure," *British Journal of Statistical Psychology*, vol. 17, pp. 65-70, 1964.

[12] J. Karvanen and V. Koivunen, "Blind separation methods based on Pearson system and its extensions," *Signal Processing*, vol. 82, pp. 663-673, 2002.

[13] S. Cruces, L. Castedo, and A. Cichocki, "Robust blind source separation using cumulants," *Neurocomputing*, vol. 49, pp. 87-117, 2002.

[14] S. Amari, "Natural gradient learning for over- and under-complete bases In ICA," *Neural Comput*, vol. 11, pp. 1875-83, 1999.

[15] A. Cichocki and S. Amari, *Adaptive Blind Signal Processing*: Wiley, 2002.

[16] J. F. Cardoso, "Higer-order contrasts for independent component analysis," *Neural Comput*, vol. 11, pp. 157-192, 1999.

[17] J. Dien, W. Khoe, and G. R. Mangun, "Evaluation of PCA and ICA of simulated ERPs: Promax vs. infomax rotations," *Hum Brain Mapp*, 2006.

[18] S. Amari, A. Cichocki, and H. Yang, "A new algorithm for blind signal separation," in *Advances in Neural Information Processi Systems*, vol. 8, M. M. D. Touretzsky, and M. Hasselmo, Ed. Cambridge, MA: MIT press, 1996, pp. 757-763.