COHERENCY AND SHARPNESS MEASURES BY USING ICA ALGORITHMS

An Investigation for Alzheimer's Disease Discrimination

Jordi Solé-Casals[1], François Vialatte[2], Zhe Chen[3] and Andrzej Cichocki[2]
[1]*Signal Processing Group, University of Vic, Sagrada Família 7, 08500 Vic, Spain*
[2]*RIKEN Brain Science Institute, LABSP, 2-1 Hirosawa, Saitama, 351-0106 Wako-Shi, Japan*
[3]*Neuroscience Statistics Research Lab., Dept. of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, U.S.A.*

Keywords: EEG, Alzheimer disease, ICA, BSS, Feature extraction.

Abstract: In this paper, we present a comprehensive study of different Independent Component Analysis (ICA) algorithms for the calculation of coherency and sharpness of electroencephalogram (EEG) signals, in order to investigate the possibility of early detection of Alzheimer's disease (AD). We found that ICA algorithms can help in the artifact rejection and noise reduction, improving the discriminative property of features in high frequency bands (specially in high alpha and beta ranges). In addition to different ICA algorithms, the optimum number of selected components is investigated, in order to help decision processes for future works.

1 INTRODUCTION

Alzheimer's disease (AD) is the most prevalent form of neuropathology leading to dementia; it affects approximately 25 million people worldwide and is expected to have a fast recrudescence in the near future (Ferri et al., 2006). Numerous clinical methods that are now available to detect this disease include brain imaging (Alexander, 2002), (Deweer et al., 1995), genetic studies (Tanzi and Bertram, 2001), and other physiological markers (Andreasen et al., 2001). However, these methods cannot be employed for the mass screening of a large population. A combination of psychological tests, such as Mini-mental score evaluation (MMSE), with electrophysiological analysis (e.g. electroencephalogram or EEG), would be a more efficient and inexpensive screening approach for detecting elderly subjects affected by AD.

The purpose of this study is to make quantitative comparisons of different independent component analysis (ICA) algorithms and to investigate their potential efficiency for preprocessing (such as noise reduction and feature extraction) the EEG data. The objective is to improve the discrimination between AD patients and age-matched control subjects.

2 EXPERIMENTAL DATA

In the course of a clinical study, mutlichannel EEG measurements (Deltamed EEG machine) were recorded from 37 elderly patients affected by Alzheimer's disease with a clinical follow-up treatment (labeled AD set) as well as from 39 age-matched controls (labeled Control set). The electrodes were located on 19 sites according to the 10-20 international system. Reference electrodes were placed between Fz and Cz, and between Cz and Pz. The sampling frequency was 256 Hz, with bandpass filter 0.17-100 Hz. When possible, three periods of 5 seconds were selected in a "rest eyes-closed" condition for each patient—only 4 subjects from the Alzheimer group did not allow us to extract three 5-second sessions and therefore were discarded in the current study. Hence, two groups of three 5-second signals were obtained for each of the 39 Controls, and 33 AD. The three independent sessions were chosen so as to minimize the presence of artifacts.

Solé-Casals J., Vialatte F., Chen Z. and Cichocki A. (2009).
COHERENCY AND SHARPNESS MEASURES BY USING ICA ALGORITHMS - An Investigation for Alzheimer's Disease Discrimination.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 468-475
DOI: 10.5220/0001430904680475

3 INDEPENDENT COMPONENT ANALYSIS AND BLIND SOURCE SEPARATION

3.1 Data Model

The experimental data are assumed to be generated by a probabilistic generative model that is described by two equations as follows:

$$\mathbf{x_t} = \mu + \mathbf{B}\mathbf{z_t} + \varepsilon_t, \quad (1)$$

$$\mathbf{z_t} = \mathbf{A}\mathbf{s_t}, \quad (2)$$

where t denotes the time index. Equation (1) is essentially a factor analysis (FA) model, where $\mathbf{z} \in \mathbb{R}^n$ are the hidden variables called "factor"; the $m \times n$ matrix $\mathbf{B}$ is called the "loading matrix"; $\mathbf{x}_t \in \mathbb{R}^m$ denote the observed multi-channel signals measured in the electrodes; $\mu \in \mathbb{R}^m$ denotes the constant mean vector that is often assumed to zero; and $\varepsilon_t \in \mathbb{R}^m$ denotes the additive uncorrelated noise that corrupts the measurements. The equation (2) describes a linear mixture model that is related to the blind source separation (BSS) problem of our interest that will be discussed later, where $\mathbf{s} \in \mathbb{R}^N$ denote the independent source signals originated from the brain; $\mathbf{A}$ denotes a linear mixing matrix that models the mixing process and the stationary propagation or scattering effect within a short timescale; and the mixed signals consist of the hidden factor $\mathbf{z}$ obtained in (1). Here, we assume a square mixing/demixing setting, in which $m > n = N$.

3.2 Procedure

At the first stage, we apply principal factor analysis or principal component analysis (PCA) to perform dimensionality reduction. This is done by whitening the hidden factor $\mathbf{z}_t$ given $\mathbf{x}_t$, assuming that noise ε_t is uncorrelated.

Specifically, given observed samples $\{\mathbf{x}_t\}_{t=1}^T$, we can calculate the sample covariance matrix (assuming zero mean) and conduct its eigenvalue decomposition (EVD) as follows

$$\hat{\mathbf{C}}_x = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^T = \mathbf{U}\Lambda\mathbf{U}^T, \quad (3)$$

where $\mathbf{U}$ is the $m \times m$ orthogonal matrix that consists of eigenvectors as its column vectors, Λ is a diagonal matrix that consists of the diagonal entries as eigenvalues. Let $\mathbf{U}_n$ denote an $m \times n$ matrix that consists of the first n dominant eigenvectors, then we can estimate the noise covariance by

$$\hat{\Sigma} = \hat{\mathbf{C}}_x - \mathbf{U}_n \Lambda_n \mathbf{U}_n^T, \quad (4)$$

and the loading matrix by

$$\hat{\mathbf{B}} = \mathbf{U}_n \Lambda_n^{1/2}. \quad (5)$$

Finally, the *whitened* factor variable $\mathbf{z}$ is produced by a linear transformation

$$\mathbf{z}_t = \mathbf{Q}\mathbf{x}_t, \quad (6)$$

where $\mathbf{Q} = (\hat{\mathbf{B}}^T \hat{\Sigma}^{-1} \hat{\mathbf{B}})^{-1} \hat{\mathbf{B}}^T \hat{\Sigma}^{-1}$.

At the second stage, an ICA algorithm is implemented to perform BSS. Specifically, given $\mathbf{z}_t$, we intend to find an optimal demixing matrix $\mathbf{W}$, operated on the whitened signal by $\mathbf{y}_t = \mathbf{W}\mathbf{z}_t$, such that the components in $\mathbf{y}_t$ are mutually uncorrelated or independent. The estimated output signal $\mathbf{y}_t$ are assumed to be the source signals of interest up certain scaling and permutation ambiguity.

Upon PCA and ICA stages, we can apply a deflation procedure to identify the individual original source in the sensor space by backward projection. Specifically, given the output signal $\mathbf{y}_t = [y_1(t), y_2(t), \ldots, y_n(t)]^T$, we can also reconstruct the *incomplete* hidden factor by projecting the ith component of $\mathbf{y}_t$, denoted by $y_i(t)$, backward onto the subspace

$$\hat{\mathbf{z}}_t = \mathbf{W}^{-1}[0,\ldots,0,y_i(t),0,\ldots,0]^T \equiv \left[\mathbf{W}^{-1}\right]_i y_i(t), \quad (7)$$

where $\left[\mathbf{W}^{-1}\right]_i$ denotes the ith column vector of the matrix $\mathbf{W}^{-1}$. Furthermore, we can reconstruct the specific source of interest in the observed data space (i.e., the scalp signals contributed merely to the ith source)

$$\begin{aligned} \hat{\mathbf{x}}_t &= \mathbf{Q}^\dagger \hat{\mathbf{z}}_t = \mathbf{Q}^\dagger \mathbf{W}^{-1}[0,\ldots,0,y_i(t),0,\ldots,0]^T \\ &= (\mathbf{W}\mathbf{Q})^\dagger [0,\ldots,0,y_i(t),0,\ldots,0]^T, \end{aligned} \quad (8)$$

where $\mathbf{Q}^\dagger$ denotes the pseudoinverse of matrix $\mathbf{Q}$. Hence, by projecting $\hat{\mathbf{x}}_t$ to the original channels' positions, we essentially *identify* the source(s) of interest.

In addition, if we are only interested in denoising or getting rid of a specific component, we can set that specific output signal (say y_i) to zero while keeping other components intact, and apply the same above-described back projection procedure to recover the original scene. In our experiments, by ranking the output components, we always select the one that has the least absolute kurtosis value (i.e., the one close to Gaussian by assuming zero kurtosis statistic for Gaussian signal, positive kurtosis statistic for super-Gaussian signal, and negative kurtosis for sub-Gaussian signal).

3.3 Selection of CANDIDATE ALGORITHMS

For comparison, we have selected eight representative ICA algorithms.[1]

The selection criteria for these algorithms are based on several factors: (i) computationally efficiency; (ii) robustness; (iii) fewer degree of freedom (such as the choices of learning rate parameter, nonlinearity, or number of iterations); (iv) preference to batch method.

Specifically, the following eight ICA/BSS algorithms are among some of the most popular BSS methods in the literature. A brief description and configuration setup of each method is given below:

1. AMUSE (Algorithm for Multiple Unknown Signals Extraction)(Tong et al., 1991): A second-order batch BSS algorithm based on a two-stage eigenvalue decomposition.
2. SOBI (Second-Order Blind Identification) (Belouchrani et al., 1997): A second-order batch BSS algorithm based on joint diagonalization of time-delayed signal covariance matrices. In our experiments, the number of time-delay covariance matrices is set to be 30.
3. JADE (Joint Approximate Diagonalization of Eigen-matrices) (Cardoso and Souloumiac, 1993): A high-order statistics (HOS)-based ICA algorithm based on joint diagonalization of second- and fourth-order cross-cummulants. As a batch method, JADE algorithm requires no parameter tuning; however, it is computationally expensive and memory-storage demanding (with an order of $O(n^4)$).
4. Pearson-ICA (Karvanen et al., 2000): An iterative ICA algorithm based on Pearson system; the maximum number of iterations is set to be default value, 1000.
5. FastICA (Hyvarinen and Oja, 1997): A fixed-point ICA method for sequential source extraction, the fixed point is sought by maximizing the "negentropy" of each mixture. We used default parameter setup with `tanh` nonlinearity and maximum number of iterations as 1000.
6. Thin-ICA : (Cruces-Alvarez et al., 2004): A batch ICA algorithm for simultaneous blind signal extraction based on thin QR and SVD factorizations.
7. CCA-BSS (Canonical Correlation Analysis-based BSS) (Borga and Knutsson, 2001): A second-order BSS algorithm based on canonical correlation of temporal observations.
8. TFD-BSS (Time-Frequency Distribution Joint Diagonalization-based BSS) (Févotte and Doncarli, 2004): This method takes account of information in time and frequency and the source separation criterion is conducted in time-frequency domain based on joint diagonalization of the spatial time-frequency distribution.

[1]The selection of the ICA algorithms here is by no means exhaustive and it mainly reflects our preference criteria. For instance, some iterative ICA algorithms like Infomax or natural gradient were not chosen here because they typically have slow convergence speed.

The detailed description of algorithms are beyond the scope here; for relevant references, see (Cichocki and Amari, 2002). All of algorithms are implemented in MATLAB, some of them are available for download from the original contributors or in the ICALAB package (Cichocki et al., WWW).

For each algorithm, we have varied the number of independent components extracted (namely, n), from 3 to 10, and searched for the optimum within this range. It was found that almost all algorithms converge within our experimental setup, except for the FastICA algorithm which sometimes failed to converge; in which cases FastICA was excluded from the comparisons.

4 PERFORMANCE EVALUATION

4.1 Coherency

Coherency is an informative measure that characterizes how the phases of two time series (in our cases, two electrodes' recordings) are coupled to each other, hence it was often used for measuring interactions of two signals. Coherency or coherence is also closely related to the terms "phase locking" and "phase synchrony" that were proposed in the literature. Interestingly, Coherence is often viewed as an important measure for distinguishing AD and MCI in EEG analysis of clinical practice (Koenig et al., 2005), (Babiloni et al., 2006).

In consistent with the terminology in Nolte et al. (2004), given two electrodes' recordings $x_i(t)$ and $x_j(t)$, let $X_i(\omega)$ and $X_j(\omega)$ denote their corresponding Fourier transforms; then the *coherency* between $x_i(t)$ and $x_j(t)$ is defined as the *normalized* cross-spectrum

$$C_{ij}(\omega) = \frac{S_{ij}(\omega)}{\sqrt{S_{ii}(\omega)S_{jj}(\omega)}}, \qquad (9)$$

where $S_{ij}(\omega) = \langle X_i(\omega)X_j^*(\omega)\rangle$ denotes the cross-spectrum, and $\langle\cdot\rangle$ represents the expectation average.

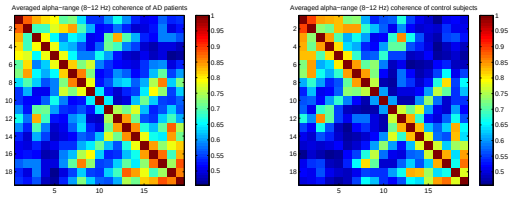

Figure 1: The mean alpha-range (8-12 Hz) coherence with raw EEG recordings (Left) averaged over all 33 AD patients, and (Right) averaged over all 39 control subjects.

And *coherence* is defined as the absolute value (or the magnitude) of the coherency, namely, $\mathrm{Coh}_{ij}(\omega) = |C_{ij}(\omega)|$.

In measuring the total coherence across the 19 electrodes, we are interested in comparing the "averaged coherence" of the original raw EEG recordings and that of the ICA-processed signals (by discarding one Gaussian-like component with the least absolute kurtosis statistic). The values of mutual coherence between electrodes are stored in a 19-by-19 symmetric matrix, with diagonal values being unity.

Specifically, we pay special attention to the alpha-range (8-12 Hz) coherence. The reasons for choosing the alpha-range are twofold: (i) The alpha wave is less noisy and therefore more reliable in the EEG recordings (because the subjects were all in rest conditions); (ii) The alpha-coherence is believed to a useful measure in characterizing AD subjects (Babiloni et al., 2006). Figure 2 illustrates the averaged alpha-range coherence (displayed in a 19-by-19 matrix) among all 33 AD patients using the raw EEG recordings. As seen in the figure, typically, neighboring channels have relatively high coherence values.

4.2 Spatial Sharpness Measure for ICA Sources

After ICA source extraction, components are obtained which are hoped to be representative of independent brain activities. However, all mathematical methods rely on mathematical assumptions and criteria which are not necessarily realistic in a real-world setting. Therefore we are also interested in investigating the biological plausibility of the ICA components.

One way to assess the plausibility of ICA components is to observe their spatial distribution—since brain activity arises from a specific area and then spreads over a larger region in the brain. During a given short time interval, it is likely that brain activity shall be located in delimited areas. After applying ICA and back-projecting the components, each three situations can be observed for each source during short time windows:

1. The source is spatially delineated in a localized and peaky area.
2. The contributing electrodes are spread all over the scalp without peaks.
3. The source is a combination of more than one peak in several separated locations.

The first situation is the only one that could be plausibly attributed to brain activity – note that sources representative of known EEG artifacts are also often spatially delineated, with a very sharp location (Delorme et al., 2001) – so that ICA algorithms extracting spatially sharp sources will provide good information both for brain signal analysis, and for artifact rejection. The second situation is the worst scenario: such a source is very unlikely to be attributed to brain activity, and would not be easily construed. The third situation is also unlikely to be representative of brain activity, and probably accounts for several independent activities which were not accurately separated by the ICA algorithm.

Naturally, we would need a measure that characterizes the "peakiness" of a distribution in a 2D representation of the scalp. Ideally, this measure shall keep the 2D structural information (note that 2D peak and 2D scrambled peak are not the same distributions), therefore the standard kurtosis measure for 1D signal is not suitable. Furthermore, this measure should reject the cases where the 2D distribution has multiple peaks.

To measure the spatial sharpness (or sparseness) of the extracted independent components, we conduct the following two-step procedure.

1. Gaussian smoothing
2. Calculating the kurtosis statistic of the Gaussian smoothed matrix

In the first step, after removing the source that has the least absolute kurtosis value (representing close-to-Gaussian noise), we extract the 2D topological information of each source. To this end we apply a bi-dimensional Gaussian smoothing procedure (Gonzalez and Woods, 1992). In our case, with a small number of electrodes, we represent the spatial information of each source with a matrix of spatial source distribution $\mathbf{D}$, an $\ell_1 \times \ell_2$ matrix (in our case, it is a 5×5 matrix reflecting the electrode layout, unused border positions are set to zero to obtain a square matrix). For each source, we convolve the matrix $\mathbf{D}$ with a Gaussian kernel $\mathbf{G}$. $\mathbf{G}$ is a 2D isotropic Gaussian distribution discretized over a square matrix whose dimension is d, with $d = \max(\ell_1, \ell_2)$, and whose standard deviation $\sigma = (d-1)/2$ such that 2σ encompasses the

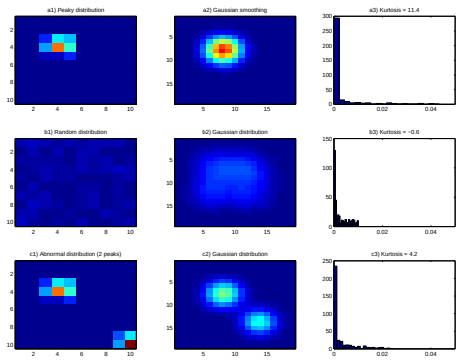

Figure 2: Illustrations of the kurtosis measure to Gaussian smoothed matrixes.

whole matrix **D**:

$$\mathbf{G}(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{10}$$

The smoothed matrix **O** is obtained by convolving the Gaussian kernel **G** and the spatial distribution of the source **D**:

$$\mathbf{O}(x,y) = \sum_{i=1}^{d}\sum_{j=1}^{d} \mathbf{G}(k,l)\mathbf{D}(x-i,y-j) \tag{11}$$

This matrix has interesting properties for our purpose: for "flattened" **D**, the smoothed matrix **O** remains flat. For peaky **D**, the smoothed matrix remains peaky; see Figure 2 for an illustration. For multiple peaks, the smoothed matrix's peaks are flattened. Therefore, matrix **O** roughly represents the spatial distribution information we are interested in extracting.

After the spatial information has been extracted by smoothing, the sparseness is computed in the second step using the conventional kurtosis measure. At this step the 2D spatial distribution does not matter anymore, therefore the kurtosis measure becomes well suited. Finally we obtain a single spatial sharpness measure for each source **s**, denoted by κ_s (sparseness of the smoothed matrix), using the kurtosis excess (Kenney and Keeping, 1962) for the matrix $\mathbf{O_s}$ (which represents the smoothed back-projected matrix from the source **s**):

$$\kappa_s = \frac{\mu_4(\mathbf{O_s})}{\mu_2(\mathbf{O_s})} - 3, \tag{12}$$

where $\mu_4(\mathbf{O_s})$ and $\mu_2(\mathbf{O_s})$ are the second and fourth order moments of the elements of $\mathbf{O_s}$.

The absolute value of the measure κ_s is close to zero for "flat distributed" elements in the matrix **D** (see second row illustration in Figure 2). On the other hand, the measure κ_s is highest when there is only one peak, and the measure value decreases when more peaks appear. After calculating this measure for each source (except for the rejected one), the averaged absolute value is used as an indicator of the spatial sharpness for all the sources

$$\kappa = \frac{1}{N}\sum_{s=1}^{N} |\kappa_s|. \tag{13}$$

This value will be also averaged within a moving temporal window across the complete duration of the data (explained later in experimental section).

5 EXPERIMENTAL RESULTS

5.1 Comparison on Coherence Change

We back projected the ICA components (by discarding only one component with the least absolute kurtosis statistic) to the original 19 electrodes. By comparing the original coherence matrix, we can calculate the relative positive increase of alpha-range coherence as well as the relative decrease of alpha-range coherence; they are summed over and then averaged over the total number of subjects and the counted number of electrodes (each with increased or decreased coherence value) based on 5 seconds of EEG recordings. In total, for each ICA algorithm with one specific number of n (i.e., the number of independent components), we calculate two statistics (mean$\pm$STD), one pair for the averaged coherence increase, and another pair for the averaged coherence decrease. The same procedure is applied to each independent session of EEG recordings. The mean statistics of averaged coherence increase and averaged coherence decrease are shown in Figure 3. As seen in the figure, for all ICA algorithms, the relative (averaged) coherence increase values drop down as the number of independent components increase, whereas the relative (averaged) coherence decrease values remain approximately constant regardless of the number of independent components. However, the overall positive coherence increase is much greater than the overall negative coherence decrease, thereby resulting a net increase of coherence for all channels. This phenomenon is anticipated because the employed PCA/ICA procedure essentially discard the noise components and keep the other components intact; on the other hand, the effect of noise reduction (and therefore coherence increase) is more pronounced when the reduced dimensionality is significant (i.e., with a small number of independent components). It is also noted that although Figure 3 only illustrates the result of the EEG recordings in one session, similar phenomena were also found in other two sessions.

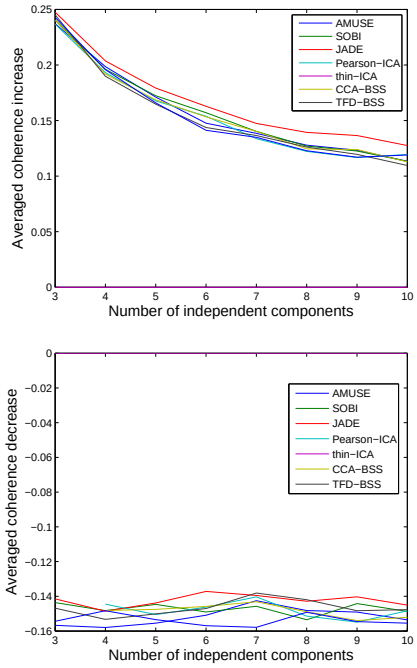

Figure 3: Top panel: performance comparison of averaged alpha-range coherence increase (mean statistics). Bottom panel: performance comparison of averaged alpha-range coherence decrease (mean statistics). The comparison was made between 7 ICA algorithms (excluding FastICA) on AD subjects.

Using averaged positive coherence increase value as an indicator of efficiency, by comparing different ICA/BSS algorithms, it was found that the JADE algorithm is the best, followed by two second-order statistic based ICA algorithms: SOBI and CCA-BSS. Interestingly, all of these three algorithms exploit information of temporal correlation, and two second-order ICA/BSS algorithms are very computationally efficient (compared to others except for AMUSE). By summing together the averaged relative coherence increase and the averaged relative coherence decrease, we obtain the result shown on Figure 4. In this case, JADE algorithm remains the best in achieving the highest net increased coherence, followed by SOBI and thin-ICA. As seen in the figure, in order to obtain net positive alpha-range coherence change, an optimum number of independent components is around 5 or 6. Interestingly, this number is consistent with other earlier investigations using the same EEG data set (Solé-Casals et al., 2008) (Cichocki et al., 2005) (Vialatte et al., 2005).

Likewise, for the control subjects, we can conduct the same analysis; the performance comparison of average net alpha-range coherence between seven ICA algorithms is shown in Figure 5. It seems that the

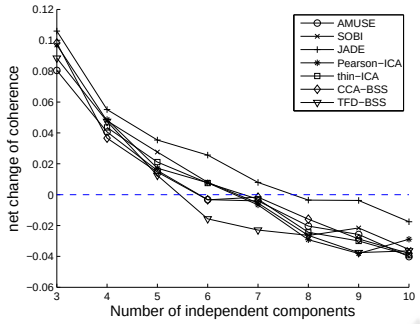

Figure 4: Performance comparison of averaged net alpha-range coherence change between 7 ICA algorithms (excluding fastICA) on AD subjects.

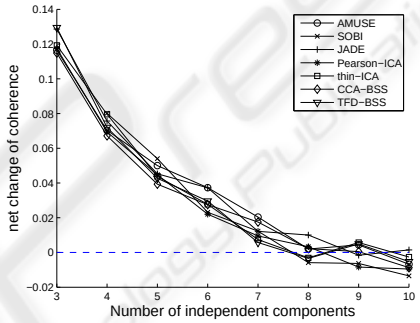

Figure 5: Performance comparison of averaged net alpha-range coherence change between 7 ICA algorithms (excluding fastICA) on control subjects.

ICA component threshold for attaining positive coherence is slightly increased to around 7 or 8. This also implies that the EEG recordings of the AD subjects are less coherent than those of the control subjects because there are more source components in the recorded signals.

5.2 Comparison on the Spatial Sharpness Measure

When the spatial sharpness measure is applied to ICA components, the time evolution of the component has to be taken into account: it is not unlikely that a brain activity may flow from one brain area to another area. We used overlapping shifted time-windows of w=500 msec over the total T=5 seconds, and measured the spatial power distribution of the back-propagated source on the scalp at each time step. At each step, a sharpness measure value κ is obtained for each source, the final result will be given by the average sharpness $\overline{\kappa}$:

$$\overline{\kappa} = \sum_{t=1}^{T-w} \frac{\kappa(t,w)}{T-w} \qquad (14)$$

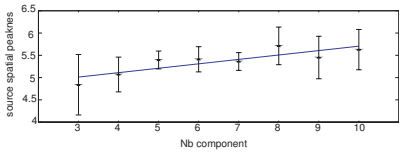

Figure 6: Relation between the number of components extracted and the average sharpness κ_m of the sources. Stars represents average for all algorithms, error bars represents the standard deviation for all algorithms.

where $\kappa(t,w)$ is the smoothed kurtosis for the source from time t to time $t+w$.

In order to obtain a fair comparison, we select the AD patients (33 subjects) as the experimental data since it was found that these recordings have generally poorer signal-to-noise ratio (SNR) than the control subjects. For each patient, three sessions of 5 seconds are used. For each session, the median value of 33 sharpness measure values for each patient is retrieved. This median value yields the representative value of sharpness for a given session. Since we are interested in a generally well-suited algorithm, and the goal is to search for the most consistent results, the final indicator will be given by the average κ_N of each session:

$$\kappa_N = \frac{\kappa_{N1} + \kappa_{N2} + \kappa_{N3}}{3}, \quad N \in [3,10] \qquad (15)$$

where N denotes the number of selected components, and $\kappa_{N1}, \kappa_{N2}, \kappa_{N3}$ denote the value κ_N for sessions 1, 2, and 3, respectively. The value κ_N is computed for each algorithm, and for all possible numbers (3∼10) of components.

Figure 6 displays the overall result obtained from this indicator. The straight line is obtained by a linear least square regression, showing a linear increase (Pearson $R^2 = 0.74$, $p < 0.05$). Therefore, the more sources are extracted, the more spatially sharp solution is obtained. Standard deviation of κ_m over the ICA algorithms is best for 5 to 7 components: within this range all ICA algorithms produce similar results, which indicates that the underlined true number of sources is likely to be close to this range.

When taken independently, each algorithm does not show the same performance. Figure 7 represents the distributions of κ_m for all algorithms. The overall increasing reported in Figure 6 is visible, as well as the consistency over components 5 to 7.

According to this measure, the overall winner is ThinICA and the overall worst algorithm is TFBSS. When analyzing the most stable period (5 to 7 components),the best algorithms, in decreasing order of efficiency, would be CCABSS, ThinICA, AMUSE and JADE.

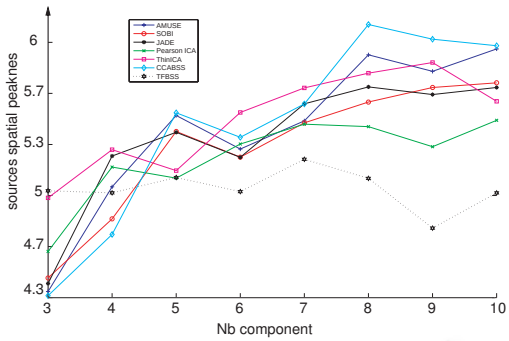

Figure 7: Spatial sharpness k_m of the sources, depending on the chosen algorithm. Sharpness is reported using the overall median kurtosis measure presented above. Except for ThinICA, for 3 components the results are poor; in general, the best algorithm depends on the selected number of components.

6 CONCLUSIONS

In this paper, we have proposed several measures or criteria to compare several popular ICA algorithms in an investigation of feature extraction of EEG signals for the purpose of discriminating Alzheimer's disease. As a powerful signal processing tool used in the preprocessing step, ICA was found useful in artifact rejection, improving SNR, and noise reduction, all of which are important for feature selection at the later stage. We also investigate the neurophysiological plausibility of the ICA outputs in terms of the sharpness measure.

It was found that, in general, ICA algorithms are particularly useful for feature extraction in high frequency bands, especially in high alpha and beta ranges; in contrast, in low-frequency bands, little gain has been obtained compared to the baselines. This fact is more or less anticipated, because EEG signals are usually contaminated by noise at high-frequency bands, but are more resistant to noise at low frequency bands. Moreover, the optimum number of selected components seem to depend on the selected algorithms, but the overall observations seem to indicate the number should be in the range from 4 to 7. In terms of the overall average performance, it seem that the JADE, SOBI, thinICA, and CCABSS algorithms give more consistent and better results.

ACKNOWLEDGEMENTS

First autor acknowledges support from the Ministerio de Educación y Ciencia of Spain under the grant TEC2007-61535/TCM, and from the Universitat de

Vic under the grant R0912.

REFERENCES

Alexander, G. E. (2002). Longitudinal pet evaluation of cerebral metabolic decline in dementia: A potential outcome measure in alzheimer's disease treatment studies. In *American Journal of Psychiatry, vol. 159, pp. 738-745*.

Andreasen, N., Minthon, L., Davidsson, P., Vanmechelen, E., and et al. (2001). Evaluation of csf-tau and csf-a?2 as diagnostic markers for alzheimer disease in clinical practice. In *Am Med Assoc, vol. 58, pp. 373-379*.

Babiloni, C., Ferri, R., Binetti, G., Cassarino, A., Forno, G. D., Eercolani, M., Ferreri, F., Frisoni, G., and et al. (2006). Fronto-parietal coupling of brain rhythms in mild cognitive impairment: A multicentric eeg study. In *Brain Research Bulletin, pp. 63–67*.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. In *IEEE Trans. Signal Processing, vol. 45, pp. 434–444*.

Borga, M. and Knutsson, H. (2001). A canonical correlation approach to blind source separation. In *Technical Report LiU-IMT-EX-0062, Department of Biomedical Engineering*.

Cardoso, J. F. and Souloumiac, A. (1993). Blind beamforming for non-gaussian signals. In *IEE Proceedings - Part F, 140, 362–370*.

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing*. Wiley, New York.

Cichocki, A., Amari, S., Siwek, K., and et al., T. T. (WWW). Icalab toolboxes. http://www.bsp.brain.riken.jp/ICALAB.

Cichocki, A., Shishkin, S. L., Musha, T., Leonowicz, Z., Asada, T., and Kurachi, T. (2005). Eeg filtering based on blind source separation (bss) for early detection of alzheimer's disease. In *Clinical Neurophysiology, 116, pp. 729–737*.

Cruces-Alvarez, S. A., Cichocki, A., and Lathauwer, L. D. (2004). Thin qr and svd factorizations for simultaneous blind signal extraction. In *Proc. European Signal Processing Conference (EUSIPCO), Vienna, Austria, pp. 217–220*.

Delorme, A., Makeig, S., and Sejnowski, T. (2001). Automatic artifact rejection for eeg data using high-order statistics and independent component analysis. In *3rd ICASSP International Workshop, San Diego,*.

Deweer, B., Lehericy, S., Pillon, B., Baulac, M., and et al. (1995). Memory disorders in probable alzheimer's disease: the role of hippocampal atrophy as shown with mri. In *British Medical Journal, vol. 58, p. 590*.

Ferri, C. P., Prince, M., Brayne, C., and et al., H. B. (2006). Global prevalence of dementia: a delphi consensus study. In *The Lancet, vol. 366, pp. 2112-2117*.

Févotte, C. and Doncarli, C. (2004). Two contributions to blind source separation using time-frequency distributions. In *IEEE Signal Processing Letters, 11, pp. 386–389*.

Gonzalez, R. and Woods, R. (1992). *Digital Image Processing*. Addison-Wesley.

Hyvarinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. In *Neural Computation, 9(7) pp. 1483–1492*.

Karvanen, J., Eriksson, J., and Koivunen, V. (2000). Pearson system based method for blind separation. In *Workshop on Independent Component Analysis and Blind Signal Separation, ICA2000, Helsinki, pp. 585–590*.

Kenney, J. F. and Keeping, E. S. (1962). *Mathematics of Statistics. Part 1*. Van Nostrand, Princeton, NJ.

Koenig, T., Prichep, L., Dierks, T., Hubl, D., Wahlund, L., John, E., and Jelic., V. (2005). Decreased eeg synchronization in alzheimer's disease and mild cognitive impairment. In *Neurobiology of Aging, 26, pp. 165–171*.

Solé-Casals, J., Vialatte, F., and Cichocki, Z. C. A. (2008). Investigation of ica algirithms for feature extraction of eeg signals in discrimination of alzheimer disease. In *Proc. International Conference on Bio-Inspired Systems and Signal Processing, Biosignals, pp. 232–235*.

Tanzi, R. E. and Bertram, L. (2001). New frontiers in alzheimer's disease genetics. In *Neuron, vol. 32, pp. 181-184*.

Tong, L., Soon, V., Huang, Y. F., and Liu, R. (1991). Indeterminacy and identifiability of blind identification. In *IEEE Trans. CAS, vole. 38, pp. 499–509*.

Vialatte, F., Cichocki, A., Dreyfus, G., Musha, T., Rutkowski, T., and Gervais, R. (2005). Blind source separation and sparse bump modelling of time frequency representation of eeg signals: New tools for early detection of alzheimer's disease. In *Proc. IEEE Workshop on Machine Learning for Signal Processing, pp. 27–32*.