Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005

On the Use of Clustering and local Singular Spectrum Analysis to Remove Ocular Artifacts from Electroencephalograms

A. R. Teixeira, A. M. Tomé
DETUA/IEETA
Universidade de Aveiro
3810-193 Aveiro, Portugal
email: ana@ieeta.pt

E. W. Lang, P. Gruber
Institute of Biophysics
University of Regensburg
93040 Regensburg, Germany
email: elmar.lang@biologie.uni-regensburg.de

A. Martins da Silva
Hospital St. António and ICBAS/IBMC
University of Porto
4099-001 Porto, Portugal
email: ams@icbas.up.pt

Abstract— **We present a method based on singular spectrum analysis [4], [5] to remove ocular artifacts (EOG) from an Electroencephalogram (EEG). After embedding the EEG signals in a feature space of time-delayed coordinates, feature vectors are clustered and the principal components (PCs) are computed locally within each cluster. Then we assume that the EOG artifact is associated with the PCs belonging to the largest eigenvalues. We incorporate a Minimum Description Length (MDL) criterion to estimate the number of eigenvectors needed to represent the EOG artifact faithfully. The EOG signal thus extracted is then subtracted from the original EEG signal to obtain the corrected EEG signal we are interested in.**

I. INTRODUCTION

Ocular activity creates artifacts in Electroencephalogram (EEG) signals, especially those recorded from frontal channels. When eyes move (blinking or other movements) the electric field around the eyes changes, producing an electrical signal known as Electrooculogram (EOG). As the signal propagates over the scalp, it also appears in the recorded Electroencephalogram (EEG). Figure 1 shows a segment of

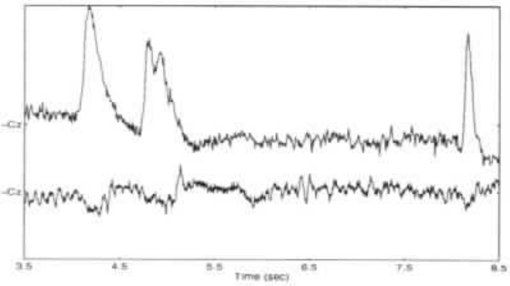


Fig. 1. A segment of 2 EEG signals recorded in Fp1-Cz and O1-Cz, respectively

two EEG channels. The large amplitude EOG artifact is more visible in frontal channels (Fp1-Cz) than in derivations of the occipital region (O1-Cz). In some studies, like single-trial event-related potentials (ERP), data from frontal electrodes located near the eyes are often not used in the analysis. Another strategy is to reject all segments of signals (epochs) that possibly contain artifacts. The task then is to manually or semi-automatically exclude all epochs which contain signals larger than a given arbitrary threshold. This is a tedious work and results in a substantial data loss. A variety of automatic procedures have been proposed to correct or remove ocular artifacts from EEG recordings. Some techniques are based on regression analysis [14], principal component analysis and more recently independent component analysis (ICA) [9], [8] or adaptive filtering techniques [7]. With ICA the performance was better than with PCA [9], where some remnants of the Electrooculogram (EOG) were still visible in the corrected data. However the identification of the components related with the ocular artifacts needs to be done manually in order to reconstruct the data without artifacts. Furthermore, those techniques were applied to the whole set of recorded EEG signals, which all included the EOG electrodes. In this work we will present a method based on singular spectrum analysis [4], [5] to remove ocular artifacts from EEG recordings. We will try to extract the EOG signal using the information from a single channel of the EEG recording only.

Singular Spectrum Analysis (SSA) is a technique applied to time series analysis. The method has been widely used in climatic, meteorologic and geophysics time series [1], [4], [5]. The general purpose of the SSA analysis is the decomposition of the time series of interest into several additive components that typically can be interpreted as "trend" components (slowly varying parts of the series), various "oscillatory" components and "noise" components [12], [5]. A related technique is Multichannel Singular Spectrum Analysis (MSSA) which is an extension of SSA to more than one time series [5]. The SSA or multichannel singular spectrum analysis (MSSA) methods mainly comprise the following steps: embedding, singular value decomposition (SVD) or principal component analysis (PCA), grouping and diagonal averaging. In this work we introduce a clustering technique before the principal component analysis so that the principal directions are computed locally in each cluster. The relevant principal components are chosen with a minimum description length criterion (MDL). Then data belonging to a cluster are reconstructed independently. Like in basic SSA free parameters are involved which must be adapted to the problem under study: the dimension of the embedding, the number of clusters and a grouping strategy. In this work we apply the algorithm to extract artifacts in Electroencephalograms (EEG) related with ocular movements (EOG). Hence we want to separate the signals into two groups: one having the components associated to EOG and the remaining components. As the EOG artifact represents a large-amplitude signal, it should be associated with the largest eigenvalues. The choice of the number of eigenvalues can be achieved using different strategies, however, experimentally the MDL criterion yields the most consistent results. The extracted EOG signal is subtracted from the original EEG signal to obtain the corrected EEG signal we are interested in.

The next section presents the basic steps of the SSA or MSSA technique incorporating the clustering step and the principles of the MDL criterion. Then some numerical simulations are presented which corroborate the consistency of the MDL criterion in the determination of the signal subspace dimension when compared with other strategies. The removal of the EOG from the EEG signal is also presented testing local SSA with a varying number of clusters.

II. Singular Spectrum Analysis and Clustering

The next paragraphs present basic principles of the methods used in the experimental set-up. The formulation follows the steps considered in SSA or MSSA methods: embedding, SVD, grouping and diagonal averaging. But we introduce a clustering step after the embedding phase of those methods, hence apply them only locally. We also modify the grouping strategy, the theoretical foundations of which are given in the next section.

A. Embedding

Embedding can be regarded as a mapping that transforms a one-dimensional time series $x_i = (x_i[0], x_i[1], \dots, x_i[N-1])$ into a multidimensional sequence of lagged vectors. Let M be an integer (window length) with $M < N$. The embedding procedure forms $K = N - M + 1$ lagged vectors that constitute the columns of the trajectory matrix

$$\mathbf{X_i} = \begin{bmatrix} x_i[M-1] & x_i[M] & \dots & x_i[N-1] \\ x_i[M-2] & x_i[M-1] & \dots & x_i[N-2] \\ x_i[M-3] & x_i[M-2] & \dots & x_i[N-3] \\ \vdots & \vdots & \ddots & \vdots \\ x_i[0] & x_i[1] & \dots & x_i[N-M] \end{bmatrix} \quad (1)$$

Note that the matrix has the same entries along the diagonals. Multichannel Singular Spectrum Analysis (MSSA) is an extension of SSA to multidimensional time series, i.e. $\mathbf{x}[n] = [x_1[n], x_2[n], \dots, x_L[n]]^T$. In that case the global trajectory matrix is a concatenation of component trajectory matrices formed for each $x_i[n], i = 1, \dots, L$

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_L]^T \quad (2)$$

with dimension $LM \times (N - M + 1)$, where L is the number of sub-matrices with the structure described by (1).

B. Clustering

In [6] a denoising technique is proposed which is based on a clustering of the data in the space of time-delayed coordinates, i.e. the grouping of columns of the trajectory matrix $\mathbf{X}$. The idea is to look for q disjoint sub-trajectory matrices $\mathbf{X}^{(c_i)}$ using a clustering algorithm (like K-means [2]) to group the columns of $\mathbf{X}$. The number N_{c_i} of columns in each sub-trajectory matrix obeys

$$\sum_{i=1}^{q} N_{c_i} = N - M + 1 \quad (3)$$

Each cluster formed by the columns of a sub-trajectory matrix is represented by a mean vector which is given by

$$\mathbf{m}_{c_i} = \frac{1}{N_{c_i}} \mathbf{X}^{(c_i)} \mathbf{j}_{N_{c_i}}, \; i = 1, \dots, q \quad (4)$$

where $\mathbf{j}_{N_{c_i}}$ represents a vector with N_{c_i} entries equal to 1. Note that after the clustering the set $\{k = 1, 2, \dots, N-M+1\}$ of indices of the columns of $\mathbf{X}$ is subdivided in q disjoint subsets $c_1, c_2, \dots c_q$. Thus sub-trajectory matrix $\mathbf{X}^{(c_i)}$ is formed with those columns of the matrix $\mathbf{X}$ which belong to the subset of indices c_i.

C. Decomposition in the Clusters

A SVD is performed for each sub-trajectory matrix, $\mathbf{X}^{(c_i)}$, followed by the extraction of components corresponding to the k principal directions of the SVD.

a) Singular Value Decomposition: This step comprises the SVD of each sub-trajectory matrix $\mathbf{X}^{(c_i)}$, $i = 1, \dots, q$. Note that the index c_i will be omitted further on to improve readability. First the cluster mean is subtracted from all columns of the sub-trajectory matrix

$$\mathbf{Y} = \mathbf{X}^{(c_i)} - \mathbf{m}_{c_i} (\mathbf{j}_{N_{c_i}})^T \quad (5)$$

The covariance matrix $\mathbf{C}$ of sub-trajectory matrix $\mathbf{X}^{c_i}$ reads $\mathbf{C} = \langle \mathbf{Y}\mathbf{Y}^T \rangle$. Taking the eigenvalues in decreasing order of magnitude $(\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_{LM} \geqslant 0)$ and the corresponding eigenvectors $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{LM})$, the projections then can be defined as

$$\mathbf{v}_l = {\lambda_l}^{-1/2} \mathbf{Y}^T \mathbf{u}_l, \; (l = 1, \dots, d) \quad (6)$$

where $d = \max\{l | \lambda_l > 0\}$. The SVD of the zero mean data can be written as

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + ... + \mathbf{Y}_d \quad (7)$$

where $\mathbf{Y}_l = (\sqrt{\lambda_l}\mathbf{u}_l)\mathbf{v}_l^T = \mathbf{u}_l\mathbf{u}_l^T\mathbf{Y}$ are the elementary matrices.

b) Grouping: This step is dependent on the goal of the time series analysis and consists in forming sums of elementary matrices [5]. In this work we consider a denoising task where a large-amplitude EOG signal has to be extracted. So only two groups are considered. One of the groups contains the largest eigenvalues which can be related with the large amplitude EOG "noise". So we want to identify how many components of the second member of eqn. (7) need to be considered to extract the EOG artifact. The principles of that choice will be discussed in the next section as it is related with the k principal directions of the space of dimension LM. Then the extracted sub-trajectory matrix is reconstructed according to

$$\mathbf{X}_r^{(c_i)} = \mathbf{Y}_1 + \mathbf{Y}_2 + \ldots + \mathbf{Y}_k + \mathbf{m}_{c_i}(\mathbf{j}_{N_{c_i}})^T \quad (8)$$

Note that this decomposition of the clusters described above must be done in each cluster using the corresponding covariance matrix. Then eqn. (5), (6) and (7) are computed in the different sub-trajectory matrices.

D. Reverting the clustering

After reconstructing all sub-trajectory matrices, $\mathbf{X}_r^{c_i}$, the clustering process must be reversed to obtain the global reconstructed trajectory matrix $\mathbf{X}_r$. Then each column of the extracted sub-trajectory matrix, $\mathbf{X}_r^{c_i}$, will be assigned to a column of $\mathbf{X}_r$ according to the contents of subset c_i.

E. Diagonal averaging

The last step is to transform the reconstructed trajectory matrix into a one-dimensional or multidimensional signal with N samples. The global reconstructed trajectory matrix achieved in this last step, however, does not possess the structure presented in the embedding step. Hence for a one-dimensional signal, all the elements in each diagonal must be substituted by its mean such that the resulting matrix will be a trajectory matrix of the desired sequence. The averaging procedure guarantees that the Frobenius norm of the difference between the original matrix and the transformed matrix has minimum value among all possible solutions to get a matrix with all diagonals equal. In MSSA the diagonal averaging is done in the L separated sub-matrices of the global reconstructed trajectory matrix.

III. Selection of Signal Subspace

The grouping strategy discussed in this work is only meant to find the k directions corresponding to the largest eigenvalues of the covariance matrix which has the dimension $p = LM$. After ordering the estimated eigenvalues of $\mathbf{C}$ by decreasing order ($\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$) we can consider two strategies: one based on variance and the other using information theoretic principles. The first strategy is often applied in pattern recognition applications in order to find the number of directions to project the data while preserving a high percentage of variance (95% is a common value). There the dimension k is chosen such that the following expression achieves the required threshold (VAR).

$$VC(k) = \frac{\lambda_1 + \lambda_2 ... + \lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_p} \geqslant VAR \quad (9)$$

The alternative strategy assumes that the covariance matrix is modeled by the following equation

$$\mathbf{C} = \mathbf{\Psi} + \sigma^2\mathbf{I}$$

where $\mathbf{\Psi}$ is a semi-positive definite symmetric matrix of rank $k \leqslant p$ and represents the covariance of the signals, while $\sigma^2\mathbf{I}$ represents the covariance of the noise (assumed to be gaussian white noise). The largest k eigenvalues of $\mathbf{C}$ are then related to the signals, while the remaining eigenvalues are related with the noise (ideally all of value σ^2).

The most widely used criterion to determine the dimension of the signal subspace k are based on the application of information theoretic principles like the MDL criterion [15], [3]. That determination is based on the application of a maximum likelihood estimation of the parameter vector of $\mathbf{C}$ which is $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \ldots, \lambda_k, \sigma^2, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k)$. Where λ_i, $i = 1, \ldots, k$ are the k largest eigenvalues of $\mathbf{C}$, and $\mathbf{u}_i$ the corresponding eigenvectors and σ^2 corresponds to the mean of the discarded eigenvalues.

Using the maximum likelihood estimate of $\widehat{\boldsymbol{\theta}}$, then k will be the value that minimizes the following expression

$$MDL(k) = -\ln f(\mathbf{X}|\widehat{\boldsymbol{\theta}}) + \frac{1}{2}K \ln N,\ k = 0, \ldots, p-1 \quad (10)$$

where N is the number of observations available to estimate the covariance matrix $\mathbf{C}$ and $f(\mathbf{X}|\widehat{\boldsymbol{\theta}})$ denotes the conditional probability density parameterized by $\widehat{\boldsymbol{\theta}}$. The log likelihood function $L(\widehat{\boldsymbol{\theta}}) = \ln f(\mathbf{X}|\widehat{\boldsymbol{\theta}})$ represents the accuracy of the representation of the data with the parameter vector $\widehat{\boldsymbol{\theta}}$ and depends on the discarded eigenvalues

$$L(\widehat{\boldsymbol{\theta}}) = N(p-k) \ln \left[\frac{\prod_{i=k+1}^{p} \lambda_i^{1/(p-k)}}{\frac{1}{p-k} \sum_{i=k+1}^{p} \lambda_i} \right] \quad (11)$$

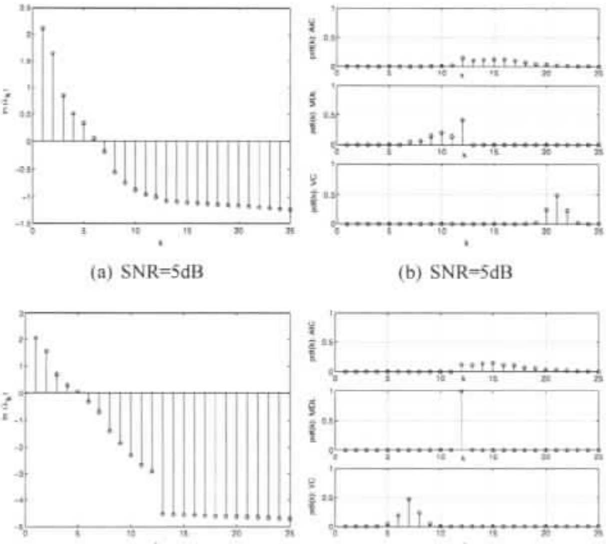
The negative log-likelihood $-L(\widehat{\boldsymbol{\theta}})$ is recognized to be a standard measure of the training error. However it has been reported that the simple maximization of this term tends to result in the phenomenon of over-fitting. Thus the second term in eqn. (10) was added as a regularization term to penalize complexity. The value of K is related with the number of parameters in $\boldsymbol{\theta}$ and the complexity of its estimation. Considering real valued signals, the value of K is computed according to

$$K = k + 1 + pk - k^2/2 - k/2 = pk - k^2/2 + k/2 + 1 \quad (12)$$

Several works [10], [11] report the behavior of the MDL criterion (and other criteria like Akaike's information criterion (AIC)) for the estimation of the dimension of the signal subspace for various applications (like model order selection in autoregressive models). In what concerns the MDL criterion, it has been concluded that the smaller eigenvalues must be clustered together and must exhibit a gap to the larger eigenvalues in order to result in a reliable estimation of the signal subspace dimension.

IV. Numerical Simulations

In this section we will present numerical simulations concerning the application of the proposed method. We start by presenting an experimental study comparing the consistency of the described criteria (Variance, MDL and AIC) to select the subspace dimension k. This study is conducted using artificially mixed signals. The second experiment deals with the denoising of one EEG signal incorporating the MDL criterion on the subspace dimension selection k and varying the number of clusters q (or sub-trajectory matrices) formed with the data embedded in time-delayed coordinates.



(a) SNR=5dB (b) SNR=5dB (c) SNR=20dB (d) SNR=20dB

Fig. 2. Parameters of the simulation- N=5000, A=5X3, M=5: (a) (c) $\ln(\lambda_k)$ versus k and (b) (d) pdf(k) of AIC, MDL and VC versus k (top-down)

A. On the use of the MDL criterion

Artificial signals ($\mathbf{s}$) are mixed with a random matrix ($\mathbf{A}$) and gaussian noise is added, i.e. $\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \epsilon[n]$. Using a segment of N samples of a L dimensional signal, $\mathbf{x}[n]$ is embedded in a space of delayed coordinates of dimension M. The input space then has dimension LM and 1000 simulations where realized varying the mixing matrix and adding random noise such that the SNR remains constant.

We use different criteria to determine the dimension of the signal subspace k: MDL, Akaike's information and percentage of variance (VC). The Akaike (AIC) criterion is very similar to the MDL criterion, its definition can be found in [11], [15]. The experiments were performed using an input space of dimension $LM = 25$ and estimating the parameters with

TABLE I

Mean Absolute Error for the estimate of noise variance using MDL to estimate subspace dimension

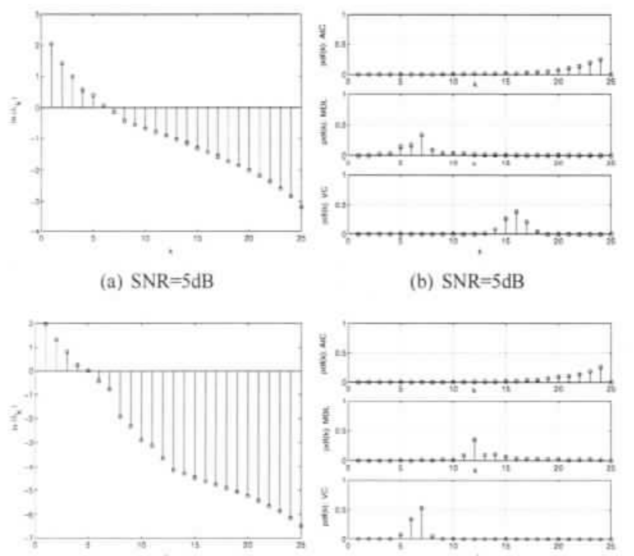
	SNR		
N	20 dB	10 dB	5 dB
50	0.0070	0.0520	0.0958
5000	0.0007	0.0095	0.0230

$N = 50$ (Fig. 3) and $N = 5000$(Fig. 2). We can see (fig. 2 b and d) that the estimated probability density function of the signal subspace k exhibits a pronounced peak when the MDL criterion is applied. For this criterion, when the SNR is high ($\geqslant 20dB$) the probability has a value close to 1 for $k = 12$. It has to be noticed that there is a clustering of the lower eigenvalues (see fig. 2-c) as well as a clear gap. For lower SNR ($5dB$) the clustering is not so evident, and there is no gap between higher and lower eigenvalues (see fig. 2-a). However, the largest value of the pdf occurs also for $k = 12$. Using the AIC, the estimated values for k yield values in the range $(12-16)$ with almost the same probability both for low and high SNRs. The percentage of variance (VC) results in completely different estimates around $k = 7$ when $SNR = 20dB$ and $k = 21$ if $SNR = 5dB$. This has to be expected, because a constant threshold is used, hence when the noise level increases it is as important as another signal.

In the study the number of samples N used on the different estimates was also varied and we conclude that the consistency of the MDL estimate increases when N increases. For instance, for $N = 50$ and $SNR = 20dB$ (fig. 3) the maximum value of pdf also occurs when $k = 12$ but its amplitude amounts to 0.5 only. Using the MDL criterion, the maximum likelihood estimates of the noise variance (as the mean of discarded eigenvalues) are compared with the variances of the added noise. Table I shows these results. The entries of the table correspond to a mean absolute error and we can also see the dependence of the error of the estimate with the number of samples. In conclusion this study shows that the MDL criterion gives a more consistent estimate of k when compared with the other criteria under the same conditions.

B. Denoising the EEG signal

A frontal (Fp1-Cz) EEG channel sampled at $128Hz$ was used. A segment of the signal with $N = 1664$ samples ($\simeq 13s$) containing high-amplitude EOG artifacts was considered. The one-dimensional signal ($L = 1$) was embedded using $M = 40$ in all three experiments performed. We tried different numbers of sub-trajectory matrices (clusters) q in the clustering step of the described method. The results will be compared visually in the time domain, but in some cases the power spectral density computed by the Welch method [13] is also considered. EEG studies usually concentrate on the frequency contents of a set of bands: theta (3.5-7.5Hz), alpha (7.5-13Hz) and beta (13-25Hz). We also compared instantaneous measures of energy in some of those bands just to evaluate the differences between the corrected EEG and original EEG. The energy was estimated in segments of $2s$, windowed with a Hamming



(a) SNR=5dB (b) SNR=5dB

(c) SNR=20dB (d) SNR=20dB

Fig. 3. Parameters of the simulation: $N = 50$, $A = 5 \times 3$, $M = 5$: (a) (c) $\ln(\lambda_k)$ versus k and (b) (d) pdf(k) of AIC, MDL and VC versus k (top-down)

window, corresponding to a frequency resolution of $0.5Hz$ with an overlap of 50%.

Table II collects some parameters of the method when the

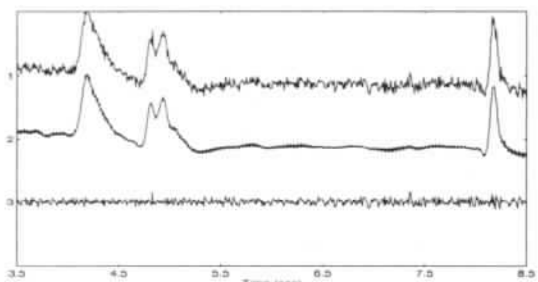


Fig. 4. Signals obtained using 6 clusters and reconstruction with the subspace dimension given by MDL: original EEG, extracted EOG and corrected EEG.

number of clusters changes. Note that for a dimension of the input space of $M = 40$ the number of clusters cannot be increased beyond 6. Otherwise there are clusters with numbers of samples close to M and the MDL criterion tends to overestimate the subspace dimension then. Also note that the vector samples for each cluster (when the algorithm is configured with clustering) are distributed inhomogeneously along the segment. The figure 5 shows the distribution of the cluster labels (using as time information for the column vector the time information of the first sample) where different amplitude levels belong to different clusters. Notice that the cluster #3 has segments of the signal where there is no EOG artifacts (for instance in the interval 5.5- 7.5s compare with fig.4) A visual inspection of the extracted EOG using the algorithm and the

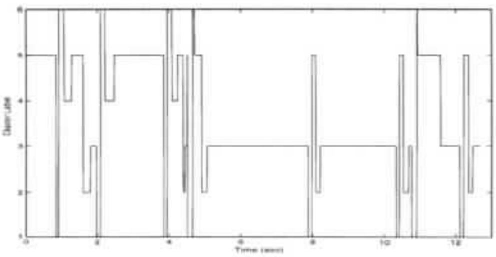


Fig. 5. Cluster labeling along the data segment with $C = 6$. The time assigned to a data vector is the time occurrence of the first sample in the vector

corrected EEG (the difference between original EEG and the extracted EOG) reveals no significant difference between the different versions computed for different numbers of clusters. Fig.4 shows a segment of those signals. It was also verified that the $50Hz$ interference is removed with EOG signal.

The power spectral density of all the signals confirms that the

TABLE II

SUBSPACE DIMENSION k AND NUMBER OF SAMPLES IN EACH CLUSTER N_{c_i}

	Labels	Dimension (k)	Size (N_{c_i})
no Cluster	1	9	1624
4 Clusters	1	7	101
	2	7	956
	3	7	132
	4	8	435
6 Clusters	1	6	128
	2	10	112
	3	5	754
	4	7	208
	5	8	324
	6	10	98

$50Hz$ line noise is also present in the extracted EOG signal (see fig. 6). We also verify that the power spectral density of the corrected signal is very low in the lower frequency band (corresponding to theta and alpha bands $< 10Hz$) whereas it is much increased in the original signal by the presence of the EOG artifact. But the beta band is very similar to the original EEG except in the proximity of 50Hz line noise. Comparing the power spectral density of the corrected EEG we can verify that the number of clusters influence mainly the lower frequency band and we can see that the best results are are achieved when $q = 6$ (see fig 7). We also tried to evaluate the frequency contents of theta and beta bands along the data segment. The lower band was chosen just because it is reported in the literature [8] that PCA methods remove the lower frequency contents. The figure 8 shows both results. The beta band is the dominant band in a frontal electrode and does not appear altered in the power density estimated for the whole segment. In the lower bands using clustering ($q = 6$), the energy contents in the extracted EEG is more close to the original EEG and it is more close when there is no ocular movement as is the case in the interval $5.5s - 7.5s$.

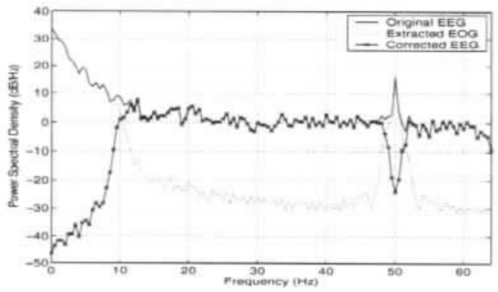


Fig. 6. Power Spectral Density(dB/Hz) versus Frequency (Hz) of Original EEG, Extracted EOG and Corrected EEG

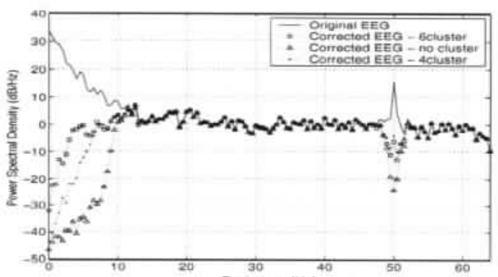
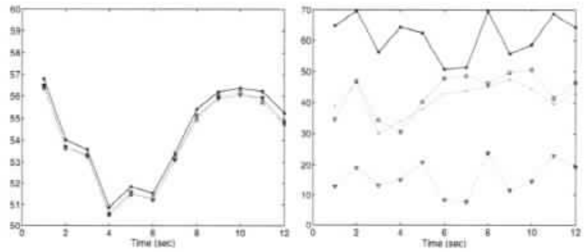


Fig. 7. Power Spectral Density(dB/Hz) versus Frequency (Hz) of original signal(⋆), Corrected EEG Signal using 6 Clusters (□), 4 Clusters(•) and no Clustering(△)



(a) Beta band(13-25Hz) (b) Theta band (3.5 -7.5 Hz)

Fig. 8. Energy along the segment: Original EEG (⋆); Corrected EEG with 6 clusters (□ -); Corrected EEG with 4 clusters and (•); Corrected EEG: no Cluster (∇)

V. Conclusions

In this work we presented an extension of SSA by incorporating a clustering step (i.e a grouping of the trajectory matrix in sub-matrices). We apply this local SSA method to extract EOG artifacts from EEG recordings. Because of the nature of the problem (a large amplitude artifact - EOG) we can also simplify the grouping strategy of the SSA method by introducing an MDL criterion. We also study the consistency of three subspace dimension selection criteria: a variance criterion, an MDL criterion and Akaike's information criterion. The MDL criterion yielded the most consistent results, hence it was chosen to be applied to the EOG artifact removal. The criterion proves to be effective when the EEG channel to be corrected contains large amplitude ocular artifacts. If, however, only the direction corresponding to the largest eigenvalue is used in the EOG extraction, a corrected EEG results which still shows remnants of ocular artifact. The method works best with EEG recordings (like frontal ones) where ocular artifacts show large amplitudes. Only then a small number of principal directions suffices to span the subspace of the EOG "noise" signal. However, we point out that the method uses the recording from a single EEG channel only, while the most widely used method (regression analysis) needs all recorded EOG channels. This preliminary experimental study needs to be extended to a larger set of signals with other characteristics in order to reach more definite conclusions concerning some free parameters (like the embedding length M and the number k of sub-trajectory matrices)

References

[1] M. R. Allen and A. W. Robertson. Distinguishing modulated oscillations from coloured noise in multivariate data sets. *Climate Dynamics*, 12:775–784, 1996.

[2] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.

[3] Eran Fishler and Hagit Messer. On the use of order statistics for improved detection of signals by the MDL criterion. *IEEE Transactions on Signal Processing*, 48(8):2242–2247, 2000.

[4] M. Ghil, M.R. Allen, M. D. Dettinger, K. Ide, and et al. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):3.1–3.41, 2002.

[5] Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & HALL/CRC, 2001.

[6] Peter Gruber, Kurt Stadlthanner, Ana M. Tomé, Ana R. Teixeira, Fabian J. Theis, Carlos G. Puntonet, and Elmar W. Lang. Denoising using local ICA and a generalized eigendecomposition with time-delayed signals. In *LNCS 3195, Proc. ICA' 2004*, pages 993–1000, Granada, 2004. Springer.

[7] P. He, G. Wilson, and C. Russel. Removal of ocular artifacts from electroencephalogram by adpative filtering. *Medical & Biological Engineering & Computing*, 42:407–412, 2004.

[8] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J. Mckeown, Vicente Iragui, and Terrence j. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000.

[9] Tzyy-Ping Jung, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, and Terrence J. Sejnowski. Removal of eye activity artifacts from visula event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111:1745–1758, 2000.

[10] Zbigniew Leonowicz, Juha Karvanen, Toshihisa Tanaka, and Jacek Rezmer. Model order selection criteria: comparative study and applications. In *International Workshop Computational Problems of Electrical Engineering*, 2004.

[11] Athanasios P. Liavas and Philip A. Regalia. On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, 49(8), 2001.

[12] V. Moskvina and K. M. Schmidt. Approximate projectors in singular spectrum analysis. *SIAM Journal Mat. Anal. Appl.*, 24(4):932–942, 2003.

[13] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing*. Prentice-Hall, 1975.

[14] Garrick L. Wallstrom, Robert E. Kass, Anita Miller, Jeffrey F. Cohn, and Nathan A. Fox. Automatic correction of ocular artifacts in the EEG: a comparison of regression and component-based methods. *International Journal of Psychophysiology*, 53:105–119, 2004.

[15] Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustic, Speech and Signal Processing,*, 33:387–392, 1985.

Proceedings

of the

International Joint Conference on Neural Networks (IJCNN) 2005

July 31 - August 4, 2005

Hilton Montréal Bonaventure Hotel

Montréal, Québec, Canada

Co-organized by:

THE INTERNATIONAL NEURAL NETWORK SOCIETY (INNS)

IEEE COMPUTATIONAL INTELLIGENCE SOCIETY

IJCNN 2005 Conference Proceedings

Additional copies may be ordered from:
IEEE Order Dept.
445 Hoes Lane / PO Box 1331
Piscataway, NJ 08855-1331 USA

Phone: (Toll Free) +1-800-678-4333
Email: customer-service@ieee.org
Web: shop.ieee.org

Bound Edition:
IEEE Catalog No. 05CH37662
ISBN: 0-7803-9048-2
ISSN: 1098-7576

CD Edition:
IEEE Catalog No. 05CH37662C
ISBN: 0-7803-9049-0