

# Automatic removal of high-amplitude artefacts from single-channel electroencephalograms

A.R. Teixeira<sup>a</sup>, A.M. Tomé<sup>a,\*</sup>, E.W. Lang<sup>b</sup>, P. Gruber<sup>b</sup>, A. Martins da Silva<sup>c</sup>

<sup>a</sup> Departamento Electrónica, Telecomunicações e Informática/IEETA, Universidade Aveiro, 3810-193 Aveiro, Portugal

<sup>b</sup> Institute of Biophysics, University of Regensburg, D-93040 Regensburg, Germany

<sup>c</sup> HGSA and ICBAS/IBMC, University of Porto, 4099-001 Porto, Portugal

## ARTICLE INFO

### Article history:

Received 30 December 2005

Received in revised form 19 June 2006

Accepted 20 June 2006

### Keywords:

Singular spectrum analysis (SSA)

Embedding

Principal component analysis

Electrooculogram (EOG)

Electroencephalogram (EEG)

## ABSTRACT

In this work, we present a method to extract high-amplitude artefacts from single channel electroencephalogram (EEG) signals. The method is called local singular spectrum analysis (local SSA). It is based on a principal component analysis (PCA) applied to clusters of the multidimensional signals obtained after embedding the signals in their time-delayed coordinates. The decomposition of the multidimensional signals in each cluster is achieved by relating the largest eigenvalues with the large amplitude artefact component of the embedded signal. Then by reverting the clustering and embedding processes, the high-amplitude artefact can be extracted. Subtracting it from the original signal a corrected EEG signal results. The algorithm is applied to segments of real EEG recordings containing paroxysmal epileptiform activity contaminated by large EOG artefacts. We will show that the method can be applied also in parallel to correct all channels that present high-amplitude artefacts like ocular movement interferences or high-amplitude low frequency baseline drifts. The extracted artefacts as well as the corrected EEG will be presented.

© 2006 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Electroencephalographic recordings can be contaminated by eye blinking, head movements, muscle activity, heart beat and line noise. These artefacts pose a problem to the interpretation of recorded EEG signals, because in many cases they constitute the most prominent signals in terms of amplitudes. In particular, ocular activity represents the major source of artefacts in electroencephalogram (EEG) signals, especially when recorded from frontal channels. Such ocular artefacts are labeled electrooculograms (EOG). In fact, especially when measured at frontal locations of the scalp close to the eyes, the EOG signal amplitude can be several times larger than the brain-generated scalp potentials. Eye movements and blinks are very

frequent and inevitably will occur during EEG recordings while objects perform various tasks. To reduce the presence of such disturbing eye movement activity in EEG recordings, the subject is often asked to suppress eye blinking or to fixate the eyes onto a given target. However, this goal is never fully accomplished either because of the nature of the task to be examined or because the subject is not willing or able to cooperate.

Consequently, in studies related with single-trial event related potentials (ERP), data from frontal channels are often discarded at all. In others it is common to discard segments of the recorded EEG which are contaminated with ocular activities. This artefact rejection is commonly done either by visual inspection of the recordings or by setting some automatic detection criterion like eliminating the segments which

\* Corresponding author. Tel.: +351 234 370512.

E-mail addresses: [ana@ieeta.pt](mailto:ana@ieeta.pt) (A.M. Tomé), [elmar.lang@biologie.uni-regensburg.de](mailto:elmar.lang@biologie.uni-regensburg.de) (E.W. Lang), [ams@icbas.up.pt](mailto:ams@icbas.up.pt) (A. Martins da Silva).  
0169-2607/\$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.  
doi:[10.1016/j.cmpb.2006.06.003](https://doi.org/10.1016/j.cmpb.2006.06.003)

achieve an amplitude higher than a predefined threshold. The detection can be accomplished either on-line or off-line (after the recording session). Naturally, the performance depends on the criterion used. A fixed threshold, for example, will generally not accomplish the variability that characterizes most of the signals during recording sessions. The advantage of on-line detection lies in the possibility of additional recordings (at least in some evoked potential response tests) to compensate for the loss of data due to artefacts. Even though, discarding segments of EEG can lead to a significant loss of information which in some cases compromises the significance of the study either because it has not enough data or the artefact-free trials represent a biased measure of the recording session. A detailed review of reduction strategies in evoked potential studies can be found in Ref. [1]. In continuous recording sessions, like the ones resulting from long monitoring sessions studying epilepsy, huge artefacts are also present and constitute a serious problem for the visual inspection of the recordings. Because of their large amplitudes, artefacts resulting from eye movements are often masking the onset of focal seizures.

## 2. Background

The availability of digital EEG recordings allows the investigation of procedures trying to remove artefact components from the recorded brain signals. The primary goal will be to remove such superimposed artefacts without distorting the underlying brain signals. A variety of automatic procedures have been proposed in the literature to correct or remove ocular artefacts from EEG recordings. Some techniques are based on regression analysis, adaptive filtering techniques [2], principal component analysis, and more recently independent component analysis (ICA) [3–5] or other blind source separation techniques [6]. The traditional method is regression analysis which basically consists in the subtraction of the scaled EOG channel (or horizontal and vertical EOG recording channels) from the EEG signal. The most recent works use independent component analysis: Jung et al. [5] used the INFOMAX-algorithm, Urrestarazu et al. [7] and Zhou and Gotman [8] applied the joint approximate diagonalization of eigen-matrices algorithm (JADE), in Ref. [6], an approximate joint diagonalization of time-delayed correlation matrices (SOBI) was used while in Ref. [3], the fast fixed point algorithm (FASTICA) was applied. In all the works but [7], the EOG channels were included in the processed data set of signals though Vigário [3] had argued that the computation of the independent components can be achieved without the inclusion of EOG signals.

One important issue in ICA methods applied to EEG recordings is the identification of components related with ocular artefacts. Hence, in order to reconstruct the data without artefacts, it is needed to eliminate the components related with the artefacts. Most of the works did not give any emphasis to the task of identifying artefact-related components which seems to be achieved mostly in a visual/manual manner. Despite the variety of methods applied, it is not possible to conclude about their performance once they use distinct databases, different measures and goals. In Ref. [4], the au-

thors conclude that ICA performed better than PCA, where some remnants of the electrooculogram were still visible in the corrected data. In Ref. [9] the performance of regression methods, principal component analysis and an independent component analysis method were compared using real and simulated data and the authors conclude that ICA distorts the power of the EEG signal in the range of 5–20 Hz.

In this work, we will present a method based on singular spectrum analysis [10,11] to remove artefacts from EEG recordings. SSA as well as local SSA, the modification we are proposing in this work, considers univariate signals, hence in case of EEGs needs as input single channel recordings only, contrary to the methods discussed above. This is a definite advantage as artefacts appear different in different channels, some even may not contain artefacts at all in certain segments. Hence, artefacts can be processed more specifically in each channel if needed. Another advantage concerns the identification of artefact related components in projection methods which generally can become very tedious in methods like ICA. With local SSA there is a natural assignment of high-amplitude artefacts to signal components associated with the largest eigenvalues of the decomposition. Further local SSA does not need a proper reference signal, a separately recorded EOG signal, for example, as regression and adaptive filtering methods do. Also note that these reference signals never provide a pure reference to the artefact, as EOG signals, for example, always contain EEG contaminations also. Last but not least, the proposed method is simple to implement and robust. However, the method discussed cannot take advantage of available topographic information as methods can do which process all channels together.

The outline of the paper is as following: in next section, the SSA methodology is briefly explained, then the local SSA method is detailed. The implementation of the algorithm depends on the assignment of some parameters; Section 4 will discuss robust strategies to choose their values. In Section 5 simulations with artificial signals help to clarify the influence of these parameters on the performance of the algorithm. The method is then applied to single channel EEG recordings. It splits the recorded signal into two components: artefacts and undistorted EEG signals. In fact, the method identifies high-amplitude artefacts as the “signal” and the underlying, undistorted EEG signal as the “noise” component. Finally, the method is also applied in parallel to a set of channels containing paroxysmal epileptiform activity with high-amplitude artefacts to extract the latter simultaneously from all channels and reveal the undistorted EEG signals.

## 3. Singular spectrum analysis and denoising

In many signal processing applications sensor signals are contaminated with noise. The latter is generally assumed to be additive and non-correlated with the sensor signals. The general purpose of SSA analysis is the decomposition of a time series into additive components which can be interpreted as “trends”, “oscillatory” and “noise” components. The SSA strategy is widely used in climatic, meteorologic and geophysics data analysis [10,11].

SSA relies on the embedding of a sensor signal in the high-dimensional space of its time-delayed coordinates thereby creating what is called a signal trajectory matrix. The multidimensional signal obtained after embedding is projected onto the most significant directions computed using singular value decomposition (SVD) or principal component analysis (PCA) techniques. An embedding strategy was also used in conjunction with FastICA to decompose recordings from a single EEG channel into distinct components [12]. Embedding multidimensional signals in a feature space spanned by delayed coordinates followed by a diagonalization of time-delayed correlation matrices is also a technique to blindly extract filtered versions of underlying source signals [13].

In this work, after embedding the signal in delayed coordinates, we introduce into SSA a clustering step which groups together similar columns of the trajectory matrix. After having applied PCA to each cluster, the multidimensional data are projected locally into a subspace spanned by  $k$  eigenvectors associated with the  $k$  most significant eigenvalues. The choice of the number  $k$  of components is based on an application of the minimum description length (MDL) criterion [14]. In next sections, we will describe the main steps of the basic SSA analysis and then introduce the additional steps of our modified version of singular spectrum analysis, called local SSA (note that this term is also used by Yiou et al. [20] in a different context). Finally, the MDL principle to choose the number of significant eigenvalues, thus estimating the dimension of the “signal subspace”, is presented.

### 3.1. Embedding and SSA analysis

Embedding can be regarded as a mapping that transforms a one-dimensional time series  $x = (x[0], x[1], \dots, x[N - 1])$  into a multidimensional sequence of lagged vectors. Let  $M \in \mathcal{N}$  denote a window length with  $M < N$ . The embedding procedure forms  $L = N - M + 1$  multivariate vectors,  $\mathbf{x}_l, l = 1, \dots, L$ , which constitute the columns of the trajectory matrix

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

Note that the trajectory matrix has identical entries along its diagonals (Toeplitz matrix). The multidimensional signal vectors can be centered in the embedding space by computing

$$\mathbf{X}_c = \mathbf{X} \left( I - \frac{1}{L} \mathbf{j}_L \mathbf{j}_L^T \right) \quad (2)$$

where  $\mathbf{j}_L = (1, \dots, 1)^T$  is an  $L \times 1$  vector of unit values, and  $I$  represents an  $L \times L$  identity matrix. With the centered trajectory matrix  $\mathbf{X}_c$  an eigendecomposition of the  $M \times M$  covariance matrix  $\mathbf{S} = \langle \mathbf{X}_c \mathbf{X}_c^T \rangle$  is computed. In SSA analysis the strategy to choose the eigenvectors of  $\mathbf{S}$  in order to project and reconstruct the multidimensional signal is called grouping [10]. The choice of eigenvectors depends on the goal of the analysis [11]. For example, denoising can be achieved by projecting the  $M$ -dimensional signal onto the  $k < M$  eigenvectors corresponding

to the  $k$  largest eigenvalues. Then a “noise-free” signal can be obtained after reconstruction. Therefore, considering the matrix  $\mathbf{U}$  with  $k$  eigenvectors corresponding to its columns, the denoised multidimensional signal is obtained via

$$\hat{\mathbf{X}} = \mathbf{U} \mathbf{U}^T \mathbf{X}_c + \mathbf{X} \frac{1}{L} \mathbf{j}_L \mathbf{j}_L^T \quad (3)$$

Notice that it cannot in general be expected that  $\hat{\mathbf{X}}$  possesses identical elements along each descending diagonal with identical values like in case of  $\mathbf{X}$  (Eq. (1)). However, this can be accomplished by replacing the entries in each diagonal by their average along the diagonal to form the matrix  $\hat{\mathbf{X}}_d$ . This procedure assures that the Frobenius norm of the difference between the original matrix,  $\hat{\mathbf{X}}$ , and the transformed matrix,  $\hat{\mathbf{X}}_d$ , has minimum value among all possible solutions to get a matrix with all diagonals equal. Finally, the denoised one-dimensional signal is obtained by reverting the embedding procedure, i.e. by taking a sample from each diagonal of the matrix  $\hat{\mathbf{X}}_d$ .

### 3.2. Local SSA

In SSA, any time series like an EEG recording is considered a superposition of underlying informative signal components. For instance, if the signal is assumed to be contaminated with a non-correlated additive gaussian noise, the following model is considered:  $x[n] = y[n] + r[n]$ , where  $r[n]$  represents additive Gaussian white noise. The embedding step turns this univariate sequence into a multivariate signal vector. With respect to a projective subspace denoising, this reduction of the noise level of the recorded signal becomes a non-linear operation thereby. The clustering step, introduced in local SSA after the embedding step, serves to approximate this non-linear processing by a locally linear processing (the PCA or SVD) by choosing directions of maximum variance in sub-groups of column vectors of the trajectory matrix.

The necessary modifications of plain SSA are explained in detail in Ref. [15]. Basically it introduces a clustering step into the SSA technique and uses an MDL criterion to choose the signal-related, uncorrelated components in each cluster. For convenience we summarize the necessary steps of the whole procedure:

- After embedding, the multivariate signals  $\mathbf{x}_l, l = 1, \dots, L$ , representing the columns of the trajectory matrix  $\mathbf{X}$ , are grouped together into  $k$  clusters using any appropriate clustering algorithm like k-means [16]. After clustering, the set of indices  $\{c_l\}$  indexing the columns of  $\mathbf{X}$  is subdivided into  $q < L$  disjoint subsets  $c_1, c_2, \dots, c_q$ . Thus, any sub-trajectory matrix  $\mathbf{X}^{(c_l)}$  is formed with those columns of the trajectory matrix  $\mathbf{X}$  which belong to the subset  $c_l$  of indices.
- A covariance matrix  $\mathbf{C}^{(c_l)} = \langle \mathbf{X}^{(c_l)} (\mathbf{X}^{(c_l)})^T \rangle$  is computed in each cluster and its eigenvectors and eigenvalues are determined. To achieve denoising, the data is projected onto the eigenvectors which correspond to the  $k$  largest eigenvalues. Applying an MDL criterion (see next section), the number  $k$  of significant directions is estimated and can be different in each cluster. After denoising a sub-trajectory matrix  $\hat{\mathbf{X}}^{(c_l)}$  is reconstructed within each cluster. The further processing is very similar to the one described by Eq. (3).

- The clustering is reverted, i.e. each column of the extracted sub-trajectory matrix  $\hat{\mathbf{X}}^{c_i}$  will be assigned to a column of  $\hat{\mathbf{X}}$  according to the contents of subset  $c_i$ .
- The reconstructed one-dimensional signal  $\hat{x}[n]$  is obtained by reverting the embedding, i.e. by averaging over the entries of the corresponding descending diagonals of the reconstructed trajectory matrix  $\hat{\mathbf{X}}$ .

#### 4. The parameters of the local SSA algorithm

The implementation of the algorithm as described in the last section requires the assignment of the following parameters: the embedding dimension ( $M$ ) and the number ( $q$ ) of clusters to split the columns of the trajectory matrix. A third parameter ( $k$ ), representing the signal subspace in each cluster, can be assigned automatically using an MDL criterion. In the following we discuss heuristics to estimate the embedding dimension  $M$  and the number of clusters  $q$ . Furthermore, we discuss an MDL criterion to estimate the signal subspace dimension  $k$ . It turns out that the parameters of the model are easy to estimate and well justified.

##### 4.1. Embedding dimension and number of clusters

In SSA applications the choice of an embedding dimension  $M$  has to be considered. If no further knowledge is available,  $M$  should be chosen approximately half of the segment length ( $N$ ) [10]. To extract periodic signal components,  $M$  should be close to their periodicity [10]. A more general strategy is followed in Ref. [12] where a lower bound is suggested according to the frequency resolution contained in every column of the trajectory matrix, i.e.  $M > f_s/f_r$  where  $f_s$  represents the sampling frequency and  $f_r$  is the minimum frequency which is to be extracted.

In local SSA the number  $q$  of clusters has to be assigned also. Obviously, the number of samples  $N$  constitutes a natural upper bound. A more practical heuristic chooses  $q$  according to the number of vectors resulting in each cluster. In particular, the cardinality of each cluster cannot be lower than the embedding dimension  $M$ .

##### 4.2. MDL criterion

The determination of the number of significant directions is based on the application of a maximum likelihood estimation of the parameter vector of the covariance matrix  $C^{(c_i)}$  of each cluster. This parameter vector is given by  $\theta_k = (\lambda_1 > \lambda_2 > \dots > \lambda_k, \sigma^2, u_1, u_2, \dots, u_k)$  representing  $k(M+1)+1$  parameters in total. The parameters  $\lambda_i$ ,  $i = 1, \dots, k$  represent the  $k \leq M$  largest eigenvalues of the covariance matrix,  $u_i$  the corresponding eigenvectors and  $\sigma^2$  corresponds to the mean of the  $M-k$  discarded eigenvalues. Using the maximum likelihood estimate of  $\hat{\theta}$ , then  $k$  will be the value that minimizes the following expression

$$\text{MDL}(k) = -\ln f(\mathbf{X}^{c_i} | \hat{\theta}_k) + \frac{1}{2} K \ln N, \quad k = 1, \dots, M-1 \quad (4)$$

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

$$L(\hat{\theta}_k) = N(M-k) \ln \left[ \frac{\prod_{i=k+1}^M \lambda_i^{1/(M-k)}}{1/(M-k) \sum_{i=k+1}^M \lambda_i} \right], \quad k = 1, \dots, M-1 \quad (5)$$

The negative log-likelihood  $-L(\hat{\theta})$  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 Eq. (4) was added as a regularization term to penalize complexity. The value of  $K$  is related with the number  $k(M+1)+1$  of parameters represented by the parameter vector  $\theta_k$ . However, the actual number of degrees of freedom is reduced by  $k(k+1)/2$  because of the normalization and orthogonality constraints imposed onto the eigenvectors of the covariance matrix. Considering real valued signals, the value of  $K$  is computed via [14] ( $k = 1, \dots, M-1$ )

$$K = [k(M+1)+1] - \left[ \frac{1}{2} k(k+1) \right] = kM - \frac{1}{2} k(k-1) + 1 \quad (6)$$

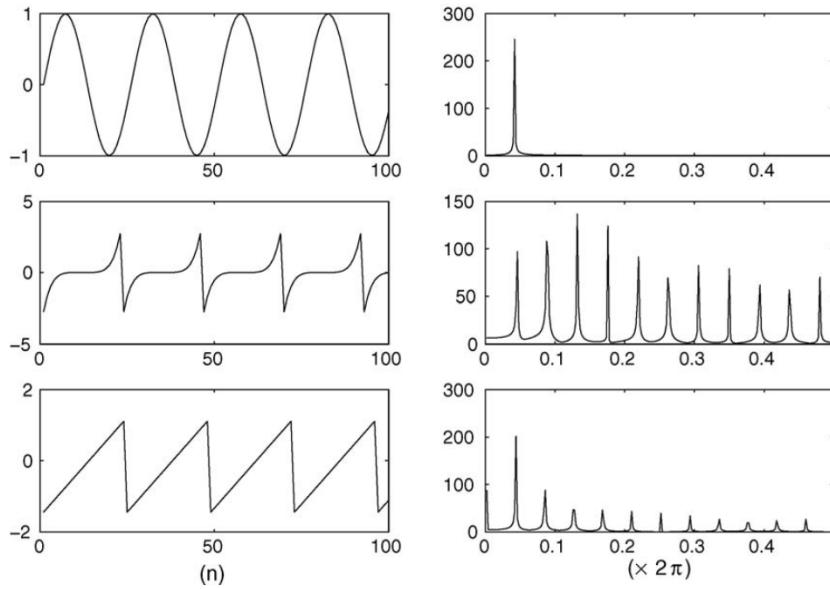
To elucidate the impact of the choice of parameters onto the performance of the algorithm, the next section first discusses some toy examples containing essential features of EEG data.

#### 5. Results and discussion

The results present and discuss artificial signals as well as EEG recordings. The aim of our experimental study with artificial signals is to demonstrate the performance of plain SSA and local SSA algorithms in dependence on the choice of the their parameters: embedding dimension  $M$ , number of cluster  $q$  and signal subspace dimension  $k$ . The latter is assigned by applying an MDL criterion. Then results obtained by applying local SSA to segments of a frontal EEG signal recording are presented. The influence of the number of clusters on the performance of the algorithm is illustrated as well. Finally, the analysis is extended to a multichannel EEG recording where the algorithm is applied in parallel to a subset of EEG channels. We can see that both the extracted signal and the corrected EEG signal provide useful information helping a visual inspection of problematic segments with high-amplitude artefacts.

##### 5.1. Artificial signals

In order to achieve conclusive insight into the choice of the free parameters of the algorithm some experiments were realized using artificial signals (see Fig. 1). The latter are represented by  $x[n] = y[n] + r[n]$ ,  $n = 0, \dots, 499$ , where  $y[n]$  is a zero-mean periodic signal and  $r[n]$  is a Gaussian white noise uncorrelated with the signal. The variance of the latter was chosen such that the ratio of energies (SNR) between  $y[n]$  and  $r[n]$  resulted to either 20 dB (high SNR) or 5 dB (low SNR) (see



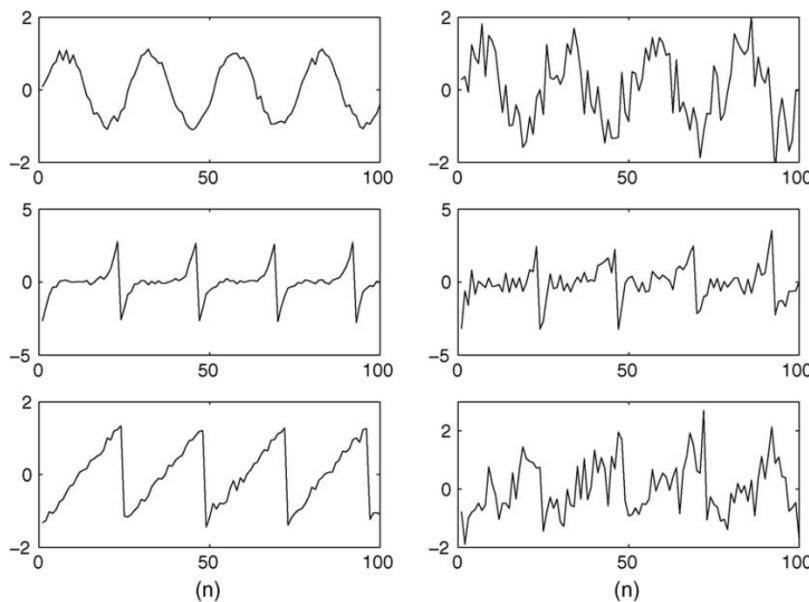
**Fig. 1 – Three artificial signals and their respective frequency contents. Top: sinusoid, middle: funny curve and bottom: sawtooth.**

Fig. 2). Note that using Gaussian noise is justified only by its simplicity. A more realistic approach would model colored noise with a frequency distribution similar to real EEGs. The periodic signals were selected with distinct frequency characteristics: (a) a narrow band signal (sinusoid), (b) a signal with energy only in a low frequency band (sawtooth) and (c) a signal whose contents spreads all over the frequency band (funny curve). These situations are sufficiently general to reflect essential features of real EEG recordings yet simple enough to ease a systematic investigation of the effect of the parameters of the method. The latter is described in Section 3.2 where the extracted signal represents a periodic wave,

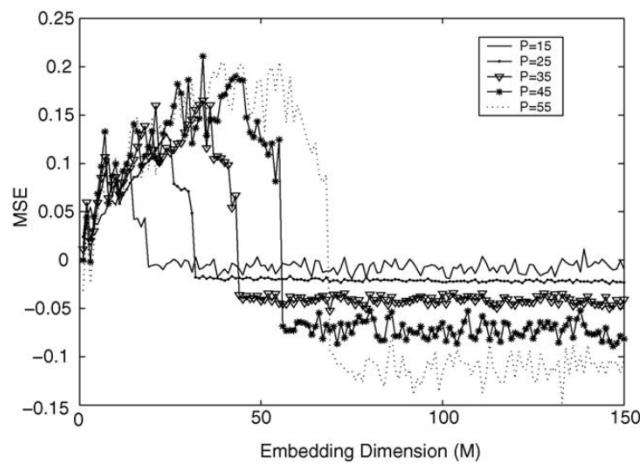
$\hat{x}[n] \simeq y[n]$ . The residual might approximate Gaussian noise, i.e.  $\hat{r}[n] = x[n] - \hat{x}[n]$ .

#### 5.1.1. Parameters of SSA

The heuristic rules proposed in the literature to choose the embedding dimension  $M$  point towards a minimum value that is related with the period of the signal to be extracted. The experiments confirm this heuristic under the following circumstances: (a) the extracted signal is wide-band and (b) the SNR is high, i.e. the extracted signal has an amplitude larger than the residual signal. Fig. 3 shows the dependence of the centered mean square error (MSE) on the embedding dimension



**Fig. 2 – The artificial signals with additive Gaussian noise. Left: SNR = 20 dB and right: SNR = 5 dB.**



**Fig. 3 – Relation between the mean square error (MSE) and the embedding dimension (M) for different signal periodicities (P). The actual values of the MSEs for a given period (P) were centered around zero to ease visual comparison.**

(M) considering different periods of  $y[n]$  and a SNR = 20 dB. However, decreasing the value of the signal-to-noise ratio to SNR = 5 dB, a similar result, i.e. a clearcut drop in the MSE above a critical embedding dimension could not be verified. And even choosing an embedding dimension higher than the respective period of the signal under consideration, (Fig. 5, left column) did not help in extracting the wide-band signals correctly.

#### 5.1.2. Parameters of local SSA

The simulations show that the local SSA algorithm works better if the embedding dimension is chosen higher than the minimal dimension given by the above mentioned heuristics when the SNR is low (5 dB). With a low SNR, i.e. when the amplitude of  $r[n]$  is close to the amplitude  $y[n]$  of the periodic signal, local SSA performs better than plain SSA with the exception of narrow band signals like the sinusoid. Table 1 shows the MSE error for all cases using embedding dimensions lower ( $M = 11$ ) and higher ( $M = 36$ ) than the period of the respective waves considered. The extracted periodic waves are shown in Figs. 4 and 5 and the best results were obtained with local SSA with an embedding dimension of  $M = 36$ .

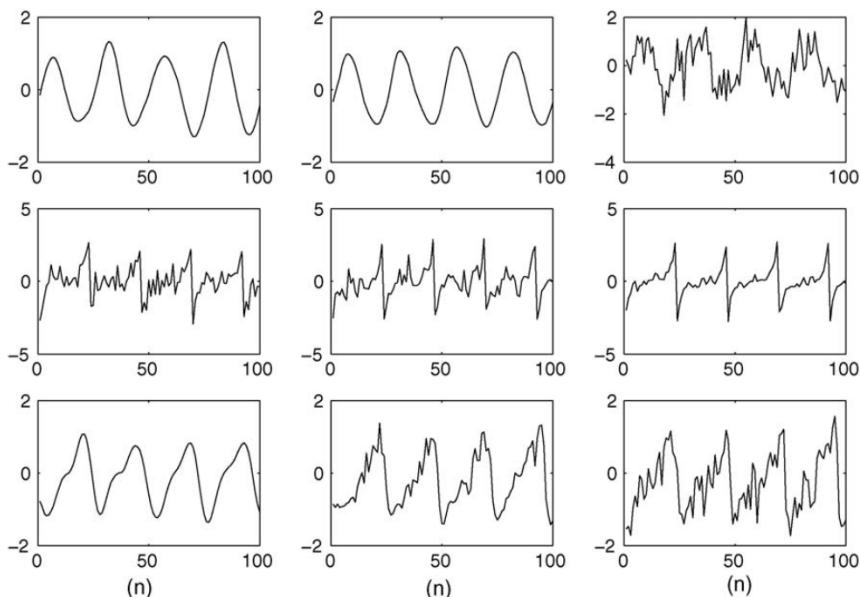
We verified that having a high embedding dimension favors the spread of the random noise signal into more directions and the MDL criterion yields a better estimate of the signal

**Table 1 – MSE between original and extracted periodic waves when SNR = 5 dB using different numbers of clusters (q)**

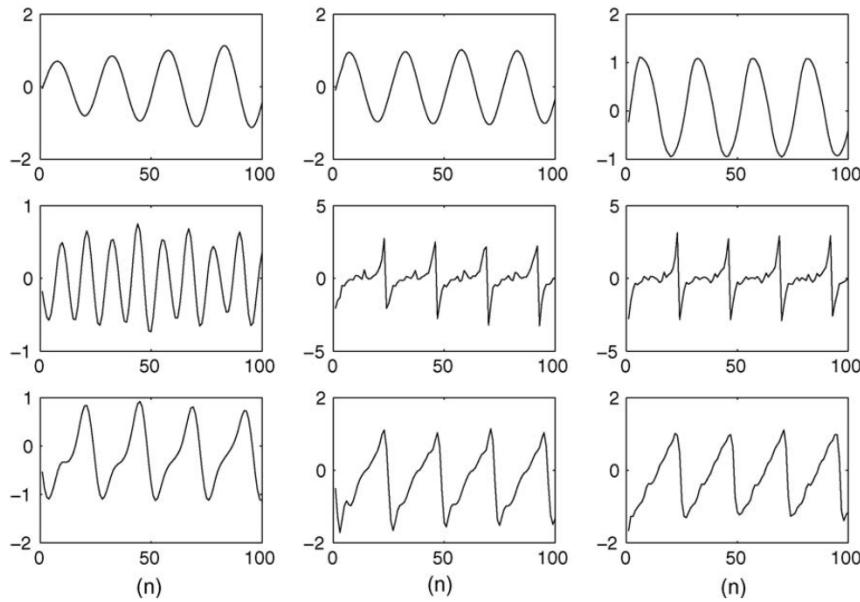
Sinusoid			Funny			Sawtooth		
$P = 26^a$			$P = 24^a$			$P = 24^a$		
$q = 1$	$q = 3$	$q = 5$	$q = 1$	$q = 3$	$q = 5$	$q = 1$	$q = 3$	$q = 5$
M = 11	0.041	0.031	0.317	0.310	0.091	0.045	0.177	0.076
M = 36	0.012	0.006	0.009	0.758	0.041	0.031	0.219	0.052

The MDL is used to choose the number of directions. Note that  $q = 1$  corresponds to plain SSA.

<sup>a</sup> Period.



**Fig. 4 – Signals  $\hat{x}[n]$  extracted with local SSA using the parameters: SNR = 5 dB, M = 11. Left:  $q = 1$ (SSA), middle:  $q = 3$  and right:  $q = 5$ .**

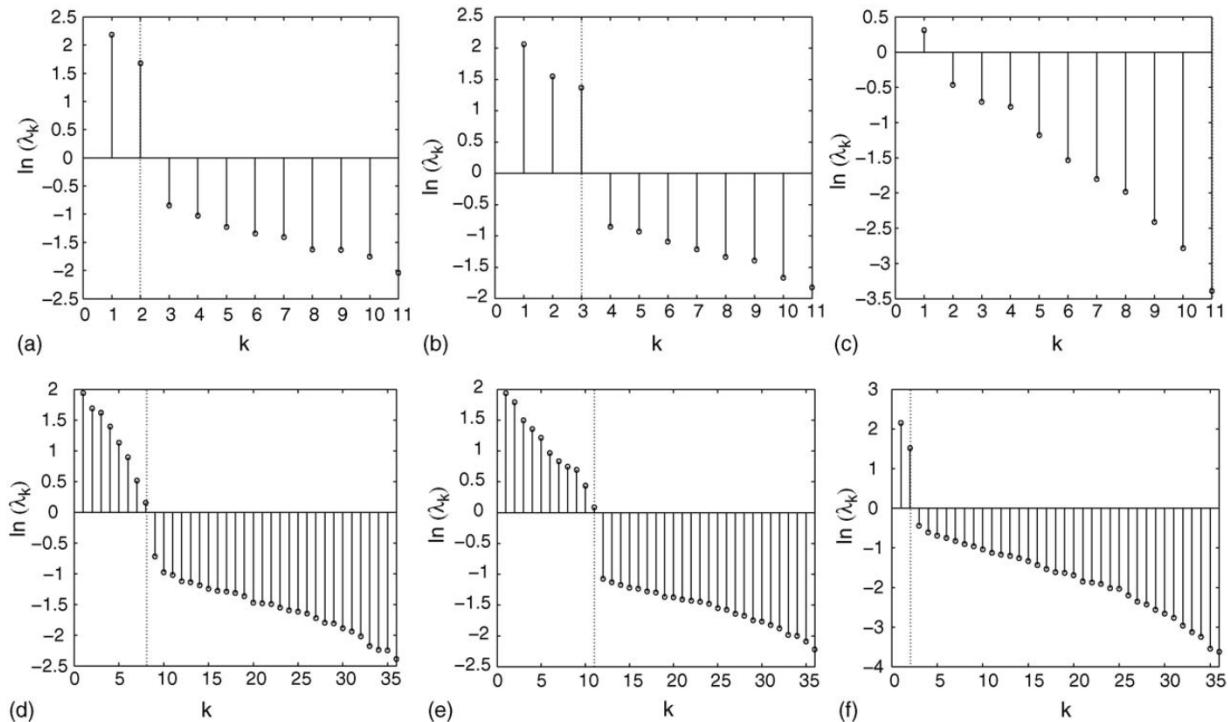


**Fig. 5 – Signals  $\hat{x}[n]$  extracted with local SSA using the parameters: SNR = 5 dB,  $M = 36$ . Left:  $q = 1$ (SSA), middle:  $q = 3$  and right:  $q = 5$ .**

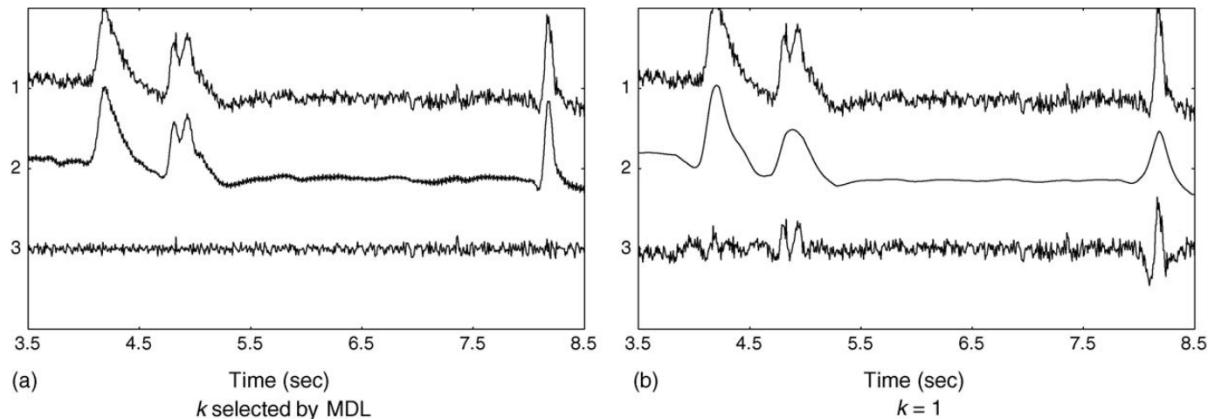
subspace. The poor results of Fig. 4 (right column) mostly result because here the MDL criterion overestimates the signal subspace dimension by selecting all the possible directions in some (or all) clusters.

Fig. 6 illustrates the phenomenon for the funny curve when  $M = 11$  (top) and  $M = 36$  (bottom). The following observations can be made:

- With  $M = 11$ , the logarithm of the eigenvalues decreases linearly, hence there is no grouping neither a clear gap, so all directions were selected. In fact, an experimental study with artificial data shows that the MDL criterion provides a more reliable estimation of the signal subspace if the number of samples to estimate the covariance matrix is large [15].



**Fig. 6 – (a-f) Funny curve: eigenvalue plot and MDL selection (dotted line) in each of the three clusters. Top  $M = 11$  where in (c) all directions are selected and bottom  $M = 36$ .**



**Fig. 7 – (a and b)** Extracted artefact and corrected EEG signal obtained with local SSA using six clusters. Top: original EEG, middle: extracted EOG and bottom: corrected EEG.

- The small eigenvalues to be associated with the noise signals must be clustered together and must be separated by a gap from the larger eigenvalues to obtain a reliable estimate of the signal subspace dimension. This confirms observations reported in other studies as well [17,14].
- For an embedding dimension larger than the period of the waves, like it is the case with  $M = 36$  in Fig. 6, the performance of the algorithm increases with the number of clusters. However, the cardinality of each cluster constitutes a natural upper bound to the number of clusters.

## 5.2. EEG analysis

The EEG signals were chosen from a database of epileptic patients recorded on long-term EEG monitoring sessions. The EEG signals were recorded using 19 electrodes placed according to the 10–20 system and mounted with a common ground reference at Fz. The signals are filtered and digitalized at a sampling rate of 128 Hz and stored as European data format (EDF), using an EEG Galileo recording system. Monopolar brain signals using the Cz electrode as reference were visualized using EEGLAB [18] and processed using SSA and local SSA methodologies.

**Table 2 – Subspace dimension  $k$  and number of samples  $N_{c_i}$  in each cluster**

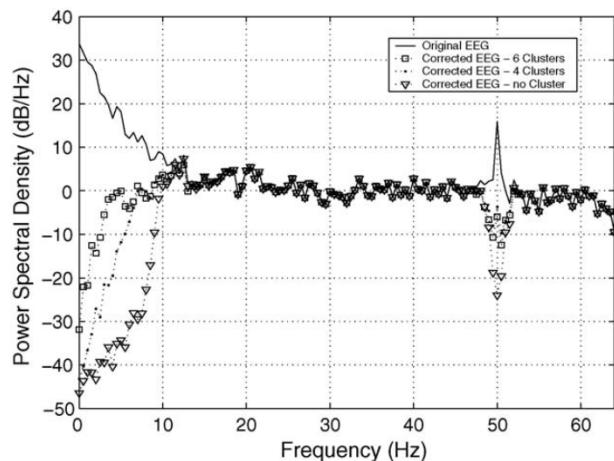
Label	Dimension ( $k$ )	$N_{c_i}$
(SSA)		
1	9	1624
Local SSA ( $q = 4$ )		
1	7	101
2	7	956
3	7	132
4	8	435
Local SSA ( $q = 6$ )		
1	6	128
2	10	112
3	5	754
4	7	208
5	8	324
6	10	98

### 5.2.1. SSA and local SSA in EEG analysis

The first EEG example shows the results of the analysis of 13 s of a frontal channel (Fp1–Cz) recording with high-amplitude eye movements. The one-dimensional signal was embedded using  $M = 41$  in all three experiments performed. We tested different numbers of clusters  $q$  in the clustering step of the local SSA algorithm and the signal subspace dimension was estimated using an MDL criterion. Table 2 exhibits the subspace dimension  $k$  assigned by MDL both for plain SSA and local SSA (with  $q = 4$  and  $q = 6$ ).

A visual inspection of the extracted EOG using the algorithm and the corrected EEG (the difference between original EEG and the extracted EOG) by a clinical expert revealed no significant difference between the different versions computed for different numbers of clusters. It was also observed that the 50 Hz interference is consistently removed together with the EOG signal in all cases.

Fig. 7 illustrates the difference between using either the MDL criterion to select the subspace dimension or using only one direction corresponding to the largest eigenvalue to reconstruct the EOG signal. In the second case the extracted



**Fig. 8 – Power spectral density(dB/Hz) vs. frequency (Hz) of original signal (full line), corrected EEG signal using local SSA—six clusters (□), four clusters (●) and SSA (Δ).**

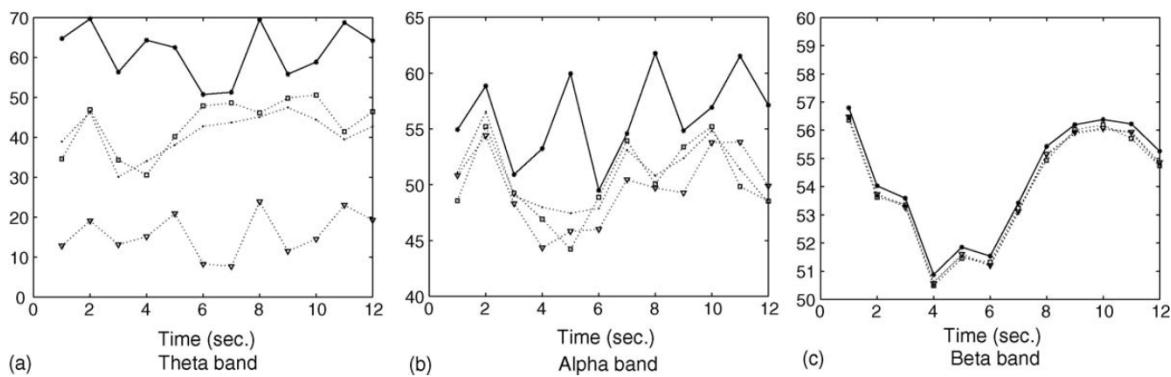


Fig. 9 – (a–c) Energy along the segment: original EEG (\*), corrected EEG signal using local SSA—six clusters (□), corrected EEG signal using local SSA—four clusters (●), corrected EEG using SSA ( $\Delta$ ).

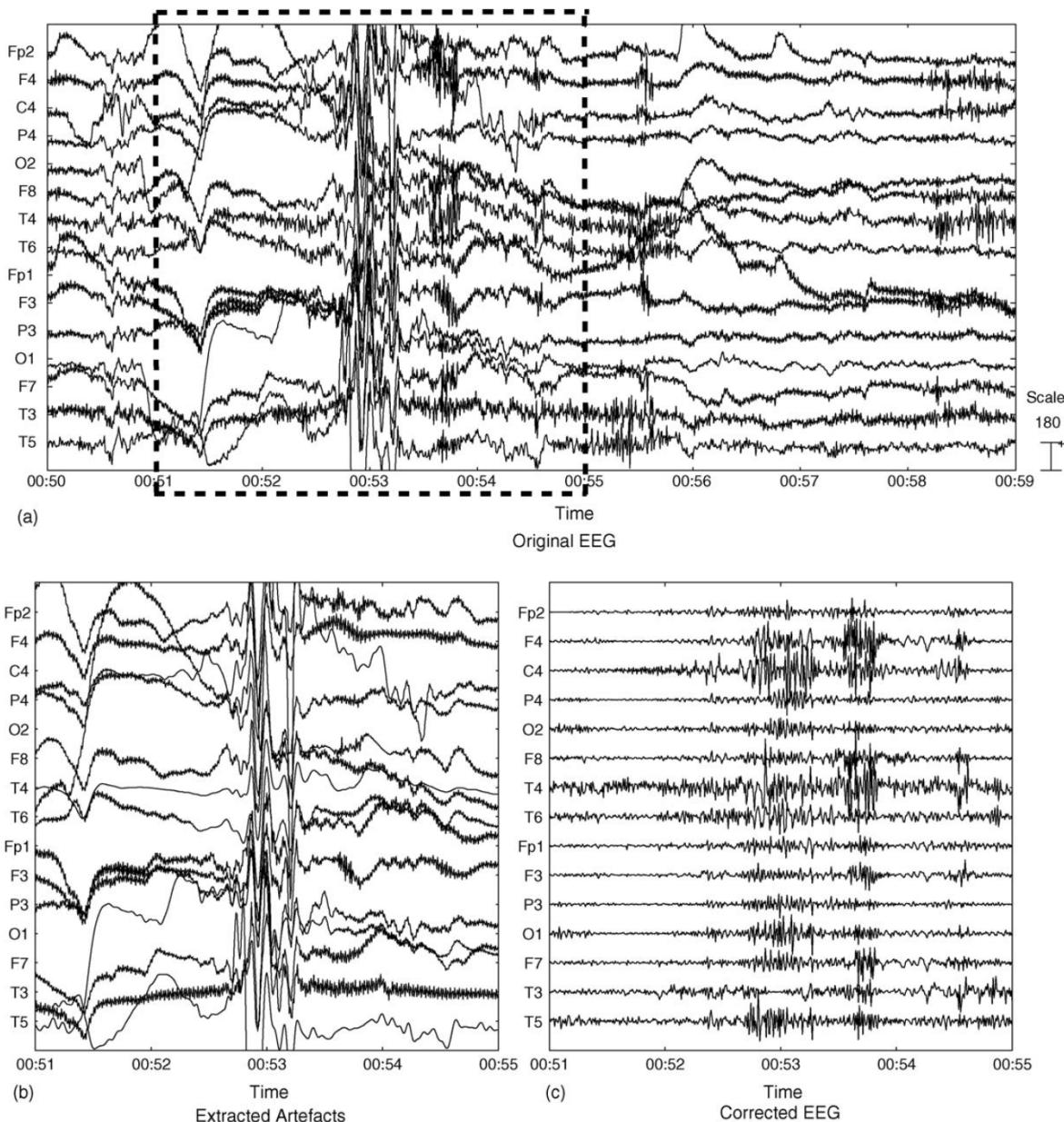


Fig. 10 – (a–c) First segment of EEG signal recordings using Cz as reference electrode. The extracted artefacts and the corrected signals are shown only for the frame indicated.

signal corresponds to the EOG artefact but the corrected EEG still contains some remnants of the original EOG signal. This demonstrates that with EOG artefacts the dominant PC is not sufficient to represent the large amplitude artefact as it was the case with the water artefact in case of 2D NMR spectra [13].

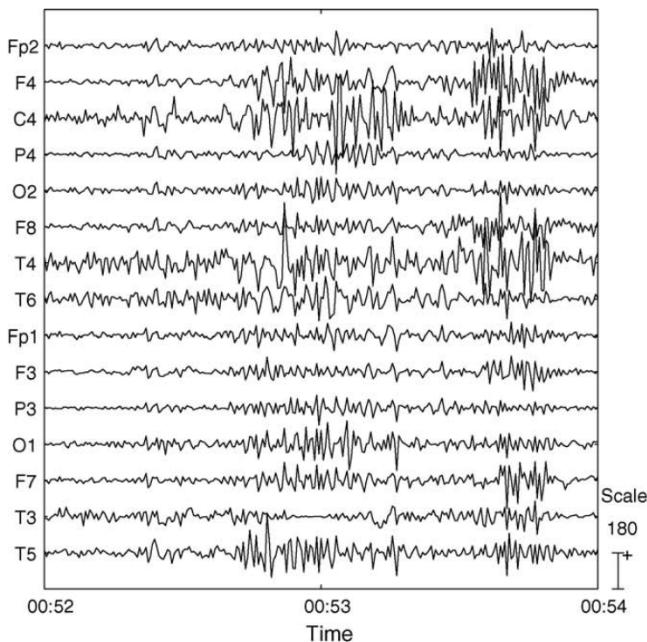
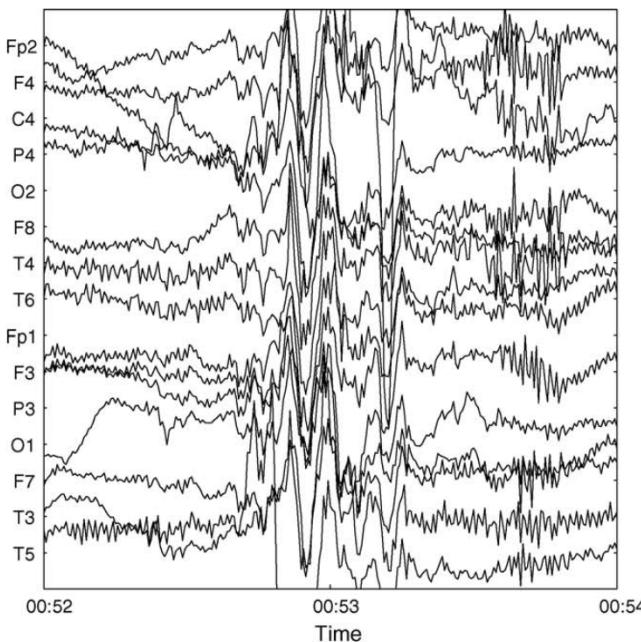
**5.2.1.1. Analysis in frequency domain.** In the frequency domain the power spectral density computed by the Welch method [19] was considered. EEG studies usually concentrate on the frequency content of the following set of frequency bands: theta waves (3.5–7.5 Hz), alpha waves (7.5–13 Hz) and beta waves (13–25 Hz). We also compared instantaneous measures of the energy in some of those bands just to evaluate the differences between the corrected EEG and their original counterparts. The energy was estimated in segments of 2 s, windowed with a Hamming window, corresponding to a frequency resolution of 0.5 Hz with an overlap between adjacent windows of 50%.

The power spectral density (psd) of all the signals confirms that the 50 Hz line noise is also present in the extracted EOG signal (see Fig. 8). We also verify that the logarithm of the power spectral density of the corrected EEG (residual signal) exhibits a drop in the low frequency band (corresponding to theta and alpha bands <10 Hz), whereas substantial spectral density builds up in the original signal due to the presence of the low frequency EOG artefact. The beta band is very similar to the original EEG except in close proximity to the 50 Hz line noise. Comparing the log psd of the corrected EEG, calculated with different numbers of clusters, it is seen that the variation of the latter influences solely the low frequency band and that the decrease in psd diminishes with increasing  $q$  (see Fig. 8). Thus, low frequency signals are less affected at higher  $q$ -values. Nonetheless the figures shown demonstrate that the high-amplitude, low frequency artefact is extracted efficiently. Note that with  $q = 6$  only below 3 Hz the psd is di-

minished strongly. But small amplitude transitory signals in this frequency range might not be affected as the local SSA method proposed relies on the fact that components related to the largest eigenvalues of the decomposition correspond to a high energy content. This is why, for example, the 50 Hz noise is removed. It represents a small amplitude signal which is present persistently in the recordings. Thus, enough energy is accumulated and this artefact becomes associated with a component corresponding to a large eigenvalue of the local SSA decomposition. These observations are corroborated by instantaneous energies estimated in the characteristic bands for the original signal versus the corrected EEG signal (see Fig. 9). The energy in the beta band is preserved in the corrected EEG while it is altered in the other bands. Again with an increasing number of clusters the difference in estimated energies between original and corrected signals is decreasing.

### 5.2.2. Multichannel EEG analysis

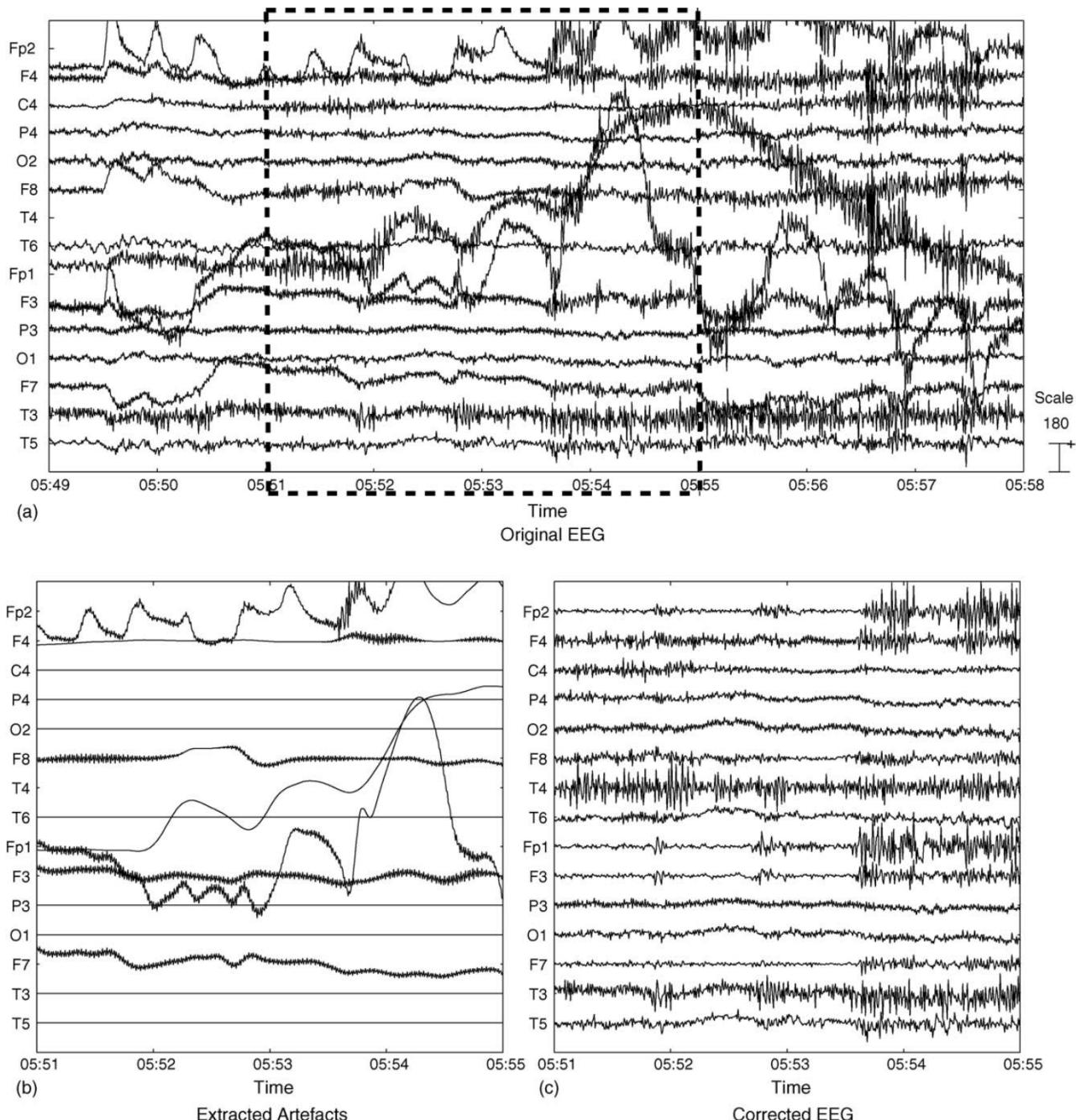
The signals of the set of channels recorded along the monitoring session suffer from distinct forms of distortion. In particular, the high-amplitude interference arising from ocular movements are more visible in frontal channels, while electrode artefacts show up in various channels spread over the scalp. We will present results using three data segments (with  $N = 1280$  samples) recorded from a patient which suffered from a partial complex seizure from the right temporal focus followed by a secondary generalization. The three segments correspond to EEG signals preceding the epileptic seizure onset and are corrupted by high-amplitude artefacts: the first segment starts 28 min before seizure onset, the second 24 min before and the last segment starts at seizure onset as verified by simultaneous video recordings of concomitant body movements. The analysis is performed in parallel in more than one channel using an embedding dimension of  $M = 41$ . The number of clusters is automatically assigned in each channel us-



**Fig. 11 – Detailed view of the first segment around 53 s shown on an expanded scale. Left: original EEG and right: corrected EEG.**

ing the given heuristics which aim to prevent over-fitting due to the MDL criterion but simultaneously uses the maximum number of clusters consistent with these heuristics. The simulations start with a maximal number of clusters  $q_{\max}$ , checking afterwards if all clusters end up with a cardinality higher than  $M$ , in which case the signal subspace dimension in each cluster is chosen as  $k \leq (M/2)$ . If both criteria are not met, then the number  $q$  of clusters is decreased and the process is repeated until a reliable decomposition in each cluster is achieved. Each channel is processed separately and Table 3 presents the number of clusters in each processed signal for the three processed segments.

- Segment 1: All channels are processed one after the other by the algorithm (see Fig. 10). The corrected EEG  $\hat{x}[n]$  (Fig. 10(b)) clearly exhibits the high-amplitude components of the original signals in an undistorted way. In most of the channels, an instantaneous frequency analysis (spectrogram) of  $\hat{x}[n]$  reveals that the frequency contents is mainly in the low frequency range (<10 Hz) and also around 50 Hz. The corrected EEG (see Fig. 10(c)) mainly possesses the high frequency (>10 Hz) contents of the original signal. However, in T4 and T6 bursts of theta (3–7 Hz) waves and sharp slow waves can be seen to occur around 53 s. This region is zoomed out for convenience in Fig. 11. The



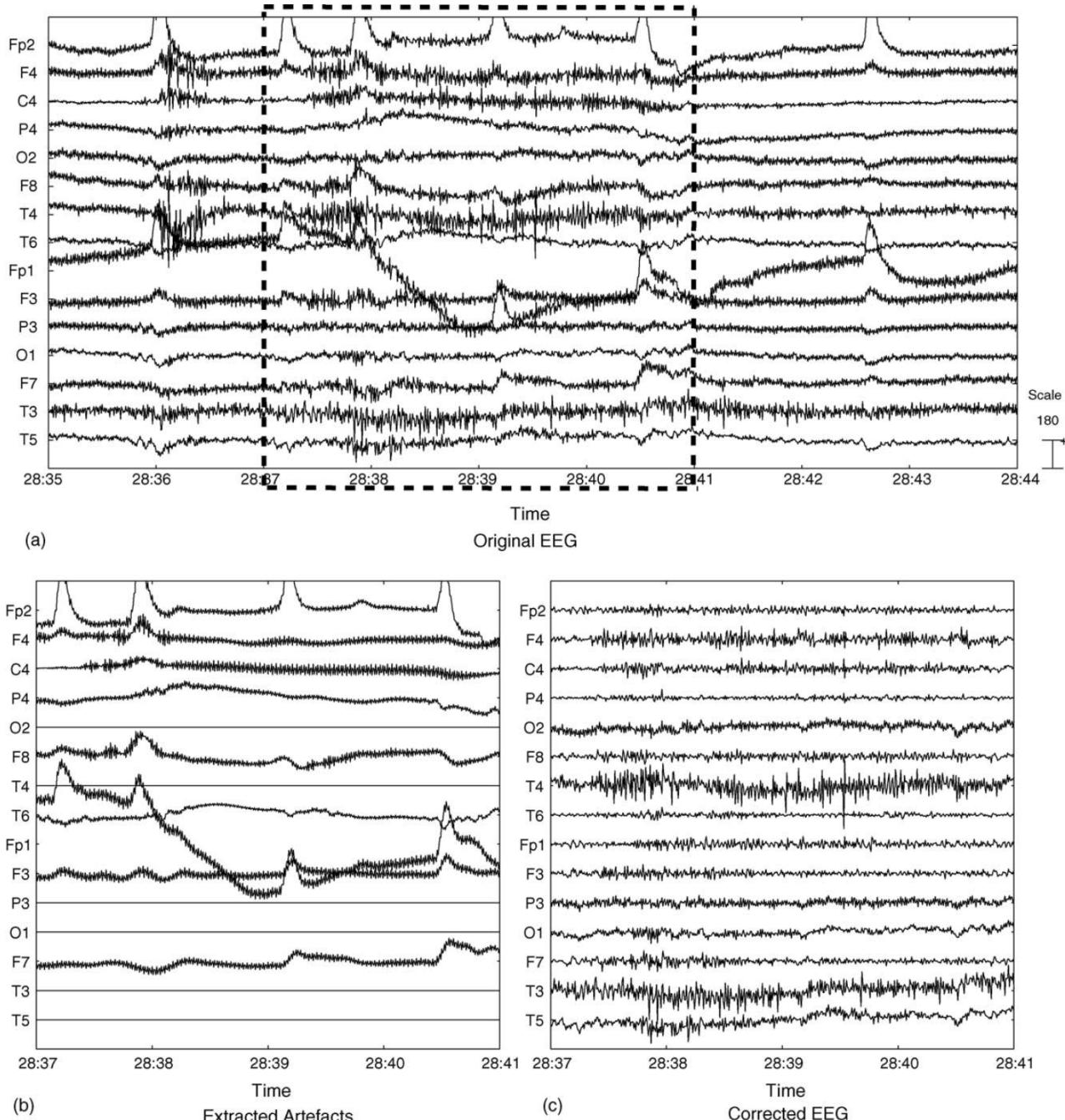
**Fig. 12 – (a–c) Second segment of EEG signals recordings using Cz as reference electrode. The extracted artefacts and the corrected signals are shown only for the frame indicated.**

bursts of spikes is now also clearly visible in the frontal channels.

- Segment 2: This segment shows typical eye movement artefacts most visible in the frontal channels. Then only these frontal channels and channel T4, monitoring temporal cortex, were processed. In Fig. 12 we can see 4 s of this analysis: the extracted signal is only related with the EOG artefact and the 50 Hz line noise (Fig. 12(b)) and the corrected EEG has the lower amplitude components of the signal. In T4 a burst of spikes (after 5 min 51 s) can be seen while in other channels (F4 and F8) single spikes also occur during the same period. Comparing the cor-

rected T4 channel (Fig. 12(c)) with the corresponding channel before the seizure onset (Fig. 13(a or c)) we can verify that both exhibit a pronounced burst of spike waves. The paroxysmal activity in T4 before the seizure initiation indicates the possible origin of the epileptogenic focus.

It is now possible to compare the corrected signal recorded in the T4 channel (Fig. 12(c)) in this segment with the corresponding signal in the segment preceding the seizure onset (Fig. 13(a or c)). As both are substantially similar this can be taken as an indication of the epileptogenic focus.



**Fig. 13 – (a–c)** Third segment of EEG signals recordings using Cz as reference electrode. The extracted artefacts and the corrected signals are shown only for the frame indicated.

**Table 3 – Number of clusters in each processed channel (“-” indicates not processed)**

EEG channels	Number of clusters		
	First segment	Second segment	Third segment
Fp2-Cz	6	10	8
F4-Cz	3	2	6
C4-Cz	3	–	7
P4-Cz	3	–	8
O2-Cz	7	–	–
F8-Cz	6	6	3
T4-Cz	5	6	–
T6-Cz	3	–	10
Fp1-Cz	6	5	7
F3-Cz	6	7	7
P3-Cz	3	–	–
O1-Cz	3	–	–
F7-Cz	3	8	4
T3-Cz	6	–	–
T5-Cz	6	–	–

- Segment 3: This segment precedes the onset of a partial complex seizure followed by a generalization and shows paroxysmal epileptiform activity in the temporal right regions. The frontal channels show ocular artefacts. In addition, channels C4, P4 and T6 show electrode artefacts (a drift in baseline) of low frequency and high-amplitude. It can be verified in Fig. 13(b) that those artefacts as well as the 50 Hz could be removed. Furthermore, channels Fp1 and Fp2 of the corrected EEG (Fig. 13(c)) show spike waves that were masked by the high-amplitude artefacts in the original signal. The focus starts in T4 and T6 with spikes, shows bilateralization and later generalization. Note that channel T4, contrary to channel T6, has not been processed and shows bursts of spikes which show up along all the data segment.

## 6. Conclusions

In this work, we presented a modified SSA method, called local SSA, to remove, from EEG recordings, high-amplitude and low frequency artefacts which accumulate enough energy to be associated with large eigenvalues of the eigendecomposition. We process the raw data to which no digital filtering has been applied. The method needs the information contained within a single channel only, hence can be applied to each channel separately. Thus, only channels which contain such artefacts need to be processed. Our results confirm that local SSA shows good performance in removing artefacts like eye or head movements, baseline drifts and line noise. In summary, with the method proposed we can separate EEG signal recordings into two components: artefacts and undistorted EEG. It has to be pointed out that local SSA does not require any user intervention to select the components of the reconstruction as in conventional ICA methods, for example. Furthermore, the user can choose to process a subset of channels keeping others unprocessed which also allows a comparison of the outcomes of the algorithm with

non-processed channels. Although this is ongoing work, we present a method that is intended to help a visual inspection of the EEG recordings by an experienced clinician, hence might be useful in some critical segment analysis like the onset of ictal seizures.

## Acknowledgements

This work was supported by grants from the DAAD and the CRUP as well as the DFG which is gratefully acknowledged.

## REFERENCES

- [1] G. Gratton, Dealing with artifacts: the EOG contamination of the event-related brain potential, *Behav. Res. Methods, Instrum. Comput.* 30 (1) (1998) 44–53.
- [2] P. He, G. Wilson, C. Russel, Removal of ocular artifacts from electroencephalogram by adaptive filtering, *Med. Biol. Eng. Comput.* 42 (2004) 407–412.
- [3] R.N. Vigário, Extraction of ocular artefacts from EEG using independent component analysis, *Electroencephalogr. Clin. Neurophysiol.* 103 (1997) 395–404.
- [4] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T.J. Sejnowski, Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects, *Clin. Neurophysiol.* 111 (2000) 1745–1758.
- [5] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. McKeown, V. Iragui, T.J. Sejnowski, Removing electroencephalographic artifacts by blind source separation, *Psychophysiology* 37 (2000) 163–178.
- [6] C.A. Joyce, I.F. Gorodniyksy, M. Kutas, Automatic removal of eye movement and blink artifacts from EEG data using blind component separation, *Psychophysiology* 41 (2004) 313–325.
- [7] E. Urrestarazu, J. Iriarte, M. Alegre, M. Valencia, C. Viteri, J. Artieda, Independent component analysis removing artifacts in ictal recordings, *Epilepsia* 45 (9) (2004) 1071–1078.
- [8] W. Zhou, J. Gotman, Removing eye-movement artifacts from the EEG during the intracarotid amobarbital procedure, *Epilepsia* 46 (3) (2005) 409–414.
- [9] G.L. Wallstrom, R.E. Kass, A. Miller, J.F. Cohn, N.A. Fox, Automatic correction of ocular artifacts in the EEG: a comparison of regression and component-based methods, *Int. J. Psychophysiol.* 53 (2004) 105–119.
- [10] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & HALL/CRC, 2001.
- [11] M. Ghil, M. Allen, M.D. Dettinger, K. Ide, et al., Advanced spectral methods for climatic time series, *Rev. Geophys.* 40 (1) (2002) 3.1–3.41.
- [12] C.J. James, D. Lowe, Extracting multisource brain activity from a single electromagnetic channel, *Artif. Intell. Med.* 28 (2003) 89–104.
- [13] A.M. Tomé, A. Teixeira, E.W. Lang, K. Stadlthanner, A. Rocha, R. Almeida, DAMUSE—a new tool for denoising and blind source separation, *Digital Signal Process.* 15 (4) (2005) 400–421.
- [14] A.P. Liavas, P.A. Regalia, On the behavior of information theoretic criteria for model order selection, *IEEE Trans. Signal Process.* 49 (8) (2001) 1689–1695.
- [15] A.R. Teixeira, A.M. Tomé, E.W. Lang, P. Gruber, A.M.d. Silva, On the use of clustering and local singular spectrum analysis to remove ocular artifacts from

- electroencephalograms, in: IJCNN2005, IEEE, Montréal, Canada, 2005, pp. 2514–2519.
- [16] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, 1995.
- [17] Z. Leonowicz, J. Karvanen, T. Tanaka, J. Rezmer, Model order selection criteria: comparative study and applications, in: International Workshop Computational Problems of Electrical Engineering, 2004.
- [18] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics, *J. Neurosci. Methods* 134 (2004) 9–21.
- [19] A.V. Oppenheim, R.W. Schafer, Digital Signal Processing, Prentice-Hall, 1975.
- [20] P. Yiou, D. Sornette, M. Gill, Data-adaptive wavelets and multi-scale SSA, *Physica D* 142 (2000) 254–290.