# Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram

Wim De Clercq*, Anneleen Vergult*, Bart Vanrumste*, *Member, IEEE,* Wim Van Paesschen**, and Sabine Van Huffel*, *Senior Member, IEEE,*

**Department of Electrical Engineering ESAT-SCD(SISTA), Katholieke Universiteit Leuven, Leuven, Belgium.*

***Department of Neurology, University Hospital Gasthuisberg, Leuven, Belgium.*

*Corresponding author (wim.declercq@esat.kuleuven.ac.be)*

ABSTRACT

The electroencephalogram (EEG) is often contaminated by muscle artifacts. In this paper a new method for muscle artifact removal in EEG is presented, based on Canonical Correlation Analysis (CCA) as a Blind Source Separation technique (BSS). This method is demonstrated on a synthetic data set and on a real ictal EEG recording contaminated with muscle artifacts. The proposed method removed successfully the muscle artifact without altering the recorded underlying brain activity. The method outperformed a low pass filter with different cutoff frequencies and an Independent Component Analysis (ICA) based technique for muscle artifact removal. Moreover, the fast computational time and the ordering of the CCA components make the proposed method applicable in clinical practice.

***Index Terms**:* EEG, Muscle artifact removal, Canonical Correlation Analysis, Blind Source Separation

# Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram

## I. Introduction

The ElectroEncephaloGram (EEG) is frequently contaminated by electrophysiological potentials associated with muscle contraction due to biting, chewing and frowning. These muscle artifacts, also known as ElectroMyoGram (EMG) artifacts, obscure the EEG and complicate the interpretation of the EEG or even make the interpretation unfeasible. In the domain of Event Related Potentials (ERP), epochs heavily contaminated with muscle artifacts are typically rejected. When a small number of epochs is available, for instance in complex memory tasks, each rejected epoch will reduce the interpretability of the ERP [1]. In automated epileptic seizure detection, muscle artifacts considerably reduce the selectivity of the method [2]. In the field of seizure prediction, muscle artifacts obscure the interpretation of the nonlinear measures when looking for a pre-ictal state in scalp EEG recordings [3]. Hence, there is a clear need to remove these artifacts from the EEG.

Low-pass filters are commonly used to remove muscle artifact. However, as the frequency spectrum of the muscle artifacts overlaps with that of interesting brain signals [4], frequency filters not only suppress muscle artifacts but also valuable information, such as ictal beta activity [5]. Most EEG artifact removal techniques only consider ocular artifacts, while, to our knowledge, little effort has been made on the removal of muscle artifacts. Regression methods, investigated for eye blink and eye movement

artifact correction [6], [7], are impractical when muscle artifacts are considered, because no reference channel is available. A more recently explored approach is Independent Component Analysis (ICA) which separates the EEG into statistical independent components. This method was already successfully applied to ocular artifact removal [8], [9]. However, cross-talk can be observed when the separation of brain and muscle activity is considered [9]–[11]. Moreover, when applying ICA, the identification of the components containing artifacts in general, and muscle activity in particular, is not obvious, and requires additional user attention. ICA based techniques that try to overcome this problem, like constrained ICA (cICA) [12], cannot be applied for muscle artifact removal. This method extracts only that independent component that is most similar to a specific reference signal [13]. However, for muscle artifacts the construction of a good reference signal is impractical.

In this paper a new method for muscle artifact correction in EEG is presented, based on the statistical Canonical Correlation Analysis (CCA) [14] applied as a Blind Source Separation (BSS) technique [15], further referred to as BSS-CCA. The BSS-CCA method, as ICA and PCA, thus belongs to the group of data-driven techniques solving the BSS problem. The blind source separation can be defined as the determination of the original or underlying sources which generate multisensor signals, while knowing very little, if anything, of how these sources project on the sensors and making little assumptions on these source signals. The BSS-CCA method requires the sources to be uncorrelated and to be maximally correlated with a given function, which is in our case a time shifted version of the original signal. By consequence, in our case, the last requirement opposes the sources to be maximally autocorrelated, inducing an order on the sources based on the autocorrelation index. Moreover, the BSS-CCA method is computationally much faster than ICA [15].

BSS-CCA is already used for exploratory data analysis in functional Magnetic Resonance Imaging (fMRI)

[15], to extract temporal structures or spatial patterns, which are difficult to specify a priori, such as unexpected activations, drifts and motion related artifacts. The method is also applied to EEG data [16]. This pilot study demonstrated that BSS-CCA can be used for separating ocular artifacts, introduced in background EEG by an artificial mixing process. However, as for our knowledge, BSS-CCA is not yet applied on EEG for muscle artifact correction.

The remainder of the paper is organized as follows. In section II, BSS by CCA is outlined, and its application as artifact corrector of EEG data is described. The synthetic data and the real life EEG recording used to test the performance of the method are described in section III. For both the synthetic data and the EEG recording, the performance was compared with the performance of a low-pass-filter and an ICA based technique. The obtained results are shown in section IV. Furthermore, in sections V and VI our findings are discussed and conclusions are drawn. Finally, a semi-automatic method to remove muscle artifacts is presented in the Appendix.

## II. Blind source separation by canonical correlation analysis for artifact removal

### *A. Blind source separation by canonical correlation analysis*

In the blind source separation problem, the observed time course $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_K(t)]^T$, with $t = 1, ..., N$, with $N$ the number of samples and $K$ the number of sensors, is the result of an unknown linear mixture of a set of unknown source signals $\mathbf{s}(t) = [s_1(t), s_2(t), ..., s_K(t)]^T$, :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{1}$$

where $\mathbf{A}$ is the unknown mixing matrix. The goal is to estimate the mixing matrix and recover the original source signals $\mathbf{s}(t)$. This is carried out by introducing the de-mixing matrix $\mathbf{W}$ such that

$$\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t), \tag{2}$$

approximates the unknown source signals in $\mathbf{s}(t)$, by a scaling factor.

Unless there are extra constraints imposed, it is in general impossible to solve this problem. Canonical correlation analysis solves the problem by forcing the sources to be mutually uncorrelated and maximally correlated with a predefined function [17], [18].

Let $\mathbf{x}(t)$ be the observed data matrix with $K$ mixtures and $N$ samples, then we define the predefined function $\mathbf{y}(t)$ as a temporally delayed version of the original data matrix, to enforce the sources to be maximally autocorrelated [17], [18]:

$$\mathbf{y}(t) = \mathbf{x}(t-1). \tag{3}$$

For the remainder of the chapter, BSS-CCA refers to this autocorrelation version.

When the mean of each row from the data matrices $\mathbf{x}(t)$ and $\mathbf{y}(t)$ is removed, canonical correlation analysis obtains two sets of basis vectors, one for $\mathbf{x}$ and the other for $\mathbf{y}$, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Consider the linear combinations of the components in $\mathbf{x}$ and $\mathbf{y}$:

$$\begin{aligned} x &= w_{x_1}x_1 + ... + w_{x_K}x_K = {\mathbf{w_x}}^T\mathbf{x}, \\ y &= w_{y_1}y_1 + ... + w_{y_K}y_K = {\mathbf{w_y}}^T\mathbf{y}, \end{aligned} \tag{4}$$

CCA finds the vectors $\mathbf{w_x} = [w_{x_1}, ..., w_{x_K}]^T$ and $\mathbf{w_y} = [w_{y_1}, ..., w_{y_K}]^T$ that maximize the correlation $\rho$ between $x$ and $y$ by solving the following maximization problem:

$$\begin{aligned} \max_{\mathbf{w_x},\mathbf{w_y}} \rho(x,y) &= \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} \\ &= \frac{E[({\mathbf{w_x}}^T\mathbf{x})({\mathbf{w_y}}^T\mathbf{y})]}{\sqrt{E[({\mathbf{w_x}}^T\mathbf{x})({\mathbf{w_x}}^T\mathbf{x})]E[({\mathbf{w_y}}^T\mathbf{y})({\mathbf{w_y}}^T\mathbf{y})]}} \\ &= \frac{{\mathbf{w_x}}^T\mathbf{C_{xy}}\mathbf{w_y}}{\sqrt{({\mathbf{w_x}}^T\mathbf{C_{xx}}\mathbf{w_x})({\mathbf{w_y}}^T\mathbf{C_{yy}}\mathbf{w_y})}}, \end{aligned} \tag{5}$$

with $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ the within-set covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathbf{C_{xy}}$ the between-sets covariance matrix. The maximization problem can be solved by setting the derivatives of Eq. (5) with

respect to $\mathbf{w_x}$ and $\mathbf{w_y}$ to zero, which after manipulation results in the following two eigenvalue problems [18], [19]:

$$\begin{cases} \mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\hat{\mathbf{w}}_{\mathbf{x}} &= \rho^2\hat{\mathbf{w}}_{\mathbf{x}}, \\ \mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\hat{\mathbf{w}}_{\mathbf{y}} &= \rho^2\hat{\mathbf{w}}_{\mathbf{y}}. \end{cases} \tag{6}$$

In the BSS-CCA version it is sufficient to solve Eq. 6 for $\hat{\mathbf{w}}_{\mathbf{x}}$ resulting in $K$ solutions $\{\rho_i, \hat{\mathbf{w}}_{\mathbf{x_i}}\}$, $i = 1...K$, sorted by autocorrelation $\rho_i$. The $K$ estimates of the sources $\mathbf{s}_i(t)$ are then given by $\mathbf{z}_i(t) = \hat{\mathbf{w}}_{\mathbf{x_i}}^T\mathbf{x}(t)$.

### *B. Applying BSS-CCA to EEG for artifact correction*

BSS-CCA is effective in performing source separation if the following assumptions are valid: (1) the mixing medium is linear and propagation delays are negligible, (2) the time courses of the sources are uncorrelated, (3) the time courses of the sources have different autocorrelation structures, and (4) the number of sources is equal as or less than the number of sensors. In the case of EEG signals, assumption (1) is satisfied, because volume conduction is thought to be linear and instantaneous [20]. The sources in case of EEG, are related to brain activity and artifacts, which are anatomically and physiologically separate processes, meaning that those processes are uncorrelated, hence assumption (2) is fulfilled. It will be demonstrated in this study that brain activity produces structured signals having a high autocorrelation, whereas muscle activity is less structured and encompass more properties related to temporally white noise. Therefore different sources seem to have different autocorrelation structures and hence assumption (3) seems to be satisfied. It is questionable whether assumption (4) is fulfilled because the effective number of uncorrelated sources contributing to the scalp EEG is not known.

Nevertheless, simulation studies have demonstrated that BSS-CCA can be used for separating ocular artifacts that are introduced in background EEG by an artificial mixing process [16], as already mentioned in the introduction.

Once BSS-CCA is applied to the EEG and the sources, or components, contributing to the EEG are

derived, then the manual or automatical identification of the components containing the artifacts is required. Next, the EEG is reconstructed without these components, by projecting the selected non-artifactual components back onto the scalp. The enhanced EEG $\mathbf{x}_{clean}(t)$ is then:

$$\mathbf{x}_{clean}(t) = \mathbf{A}_{clean}\mathbf{z}(t), \tag{7}$$

with $\mathbf{z}(t)$ the sources obtained by BSS-CCA, and $\mathbf{A}_{clean}$ the mixing matrix with the columns representing activations of the artifactual sources, set to zero.

## III. Materials and Methods

### A. Construction of the synthetic data

The aim of the simulation study was to evaluate the performance of the proposed method in removing muscle artifacts. Several simulations with different signal to noise ratios were performed.

*1) Brain activity:* Three EEG epochs of 10 s containing respectively mainly delta (<4Hz), alpha (8-13Hz) and beta (>13Hz) activity, were selected as underlying brain signals. In the selected EEGs no muscle artifact was present according to the visual inspection of an experienced neurophysiologist. The data was collected from 21 scalp electrodes placed according to the International 10-20 System [21] with additional electrodes $T1$ and $T2$ on the temporal region. The sampling frequency was 250 Hz and an average reference was used. The three EEG epochs containing delta, alpha and beta activity were stored in three 21-by-2500 dimensional matrices $\mathbf{B_1}$, $\mathbf{B_2}$ and $\mathbf{B_3}$, respectively.

*2) Muscle activity:* To obtain only muscle activity, necessary for the simulation study, it is not sufficient to select muscle artifacts in the EEG as these events will not only contain muscle activity but also activity from the brain. To remove the EEG activity we used ICA (SOBI) [22] to decompose these events. Note that for a large number of events, which we visually inspected, no clear separation was established.

For those events where a clear separation between EMG and brain activity was obtained, we selected a component that accounted for muscle artifact together with the corresponding field distribution. This procedure was repeated for three different 10 s EEG epochs recordings from three different subjects. The selected components and their corresponding field distributions (topographical maps) are shown in Fig. 1 (a-c). The individual components were reconstructed into a conventional EEG format. The resulting average referenced signal $\mathbf{M}$ containing all three independent muscle activities is shown in Fig. 1 (d).

*3) Data and measures to test the performance:* In the simulation study the average referenced muscle artifact signal $\mathbf{M}$ was superimposed on the average referenced signal $\mathbf{B}$ containing only brain activity:

$$\mathbf{X}(\lambda) = \mathbf{B} + \lambda \cdot \mathbf{M}, \tag{8}$$

with $\lambda \in \mathbb{R}$, representing the contribution of muscle activity. The Root Mean Squared (RMS) value of the brain signal is then equal to

$$RMS(\mathbf{B}) = \sqrt{\frac{1}{K \cdot N} \sum_{k=1}^{K} \sum_{n=1}^{N} \mathbf{B}_{kn}^2}, \tag{9}$$

with $N$ equal to the number of time samples and $K$ equal to the number of EEG channels. The RMS value of the muscle artifact signal is equal to

$$RMS(\lambda \cdot \mathbf{M}) = \sqrt{\frac{1}{K \cdot N} \sum_{k=1}^{K} \sum_{n=1}^{N} (\lambda \cdot \mathbf{M}_{kn})^2}. \tag{10}$$

An important measure is the Signal to Noise Ratio (SNR) which is defined as follows,

$$SNR = \frac{RMS(\mathbf{B})}{RMS(\lambda \cdot \mathbf{M})}. \tag{11}$$

Changing the $\lambda$ parameter alters the signal to noise ratio of our simulated signal. Several simulations with different signal to noise ratios were performed. The performance is expressed in terms of the Relative Root Mean Squared Error (RRMSE):

$$RRMSE = \frac{RMS(\mathbf{B} - \hat{\mathbf{B}})}{RMS(\mathbf{B})}, \tag{12}$$

where $\hat{\mathbf{B}}$ is the signal after muscle artifact removal.

The simulations were carried out for the three different brain signals $\mathbf{B_1}$, $\mathbf{B_2}$ and $\mathbf{B_3}$, described in section III-A.1

### *B. Methods for comparison*

For comparison, the proposed technique for muscle artifact removal is compared with other commonly used techniques for this purpose:

- Low pass filter: a low pass butterworth filter of order 8 was used with four different cutoff frequencies equal to 10Hz, 15Hz, 20Hz and 30Hz.
- ICA: as it is not in the scope of this paper to compare all different ICA algorithms, we limit ourself to the Joint Approximate Diagonalization of Eigen-matrices (JADE) algorithm [23]. This ICA algorithm was previously used in a study for eye and muscle artifact removal in ictal scalp EEG recordings [10]. The muscle artifact was removed as follows. The ICA components were calculated by using the JADE algorithm. The selection of the components accounting for the muscle artifact was based on visual inspection and the signal was reconstructed, excluding the components related to the artifact. It is worth mentioning that, for the muscle artifact removal based on ICA (JADE) to be unbiased, a different ICA algorithm (namely SOBI) as the JADE algorithm was used in the construction of the simulation study in section III-A.2.

### *C. Ictal EEG recording*

Because ictal EEG is often severely contaminated with muscle artifact, complicating the determination and localization of the ictal onset, the paper is focussed on the artifact correction of ictal EEG. Fig. 6 (a). shows a 10 s epoch of a scalp EEG recording from a long-term Epilepsy Monitoring Unit. The acquisition settings were the same as for the brain activity in section III-A.1. This EEG contains ictal activity from a patient with Mesial Temporal Lobe Epilepsy (MTLE). The epileptic activity is mainly

present on the $T2$, $F8$ and $T4$ electrodes. The seizure EEG is contaminated with muscle artifact and eye blinks. Again we applied CCA, ICA and frequency filtering to this epoch. The results were then visually inspected. Error measures could not be used as we did not have a-priori information available of the brain (epileptic) sources.

## IV. Results

### *A. Results on the synthetic data*

Fig. 2 (a) shows a simulated signal $\mathbf{X}$ with a signal to noise ratio equal to 0.65. For this simulated signal, the brain activity signal $\mathbf{B_2}$, which is shown in Fig. 2 (b), is used. The (normalized) components obtained by applying the BSS-CCA algorithm on $\mathbf{X}$ are shown in Fig. 3 (a). The autocorrelation of the components is shown in Fig. 3 (b). The components accounting for the muscle artifact are present in the three lowest autocorrelated components. The corresponding field distributions are shown in Fig. 3 (c-e). Both the derived time courses and field distributions are very similar to the ones of the original muscle sources shown in Fig. 1 (a-c). Fig. 4 shows the RRMSE as a function of the number of lowest autocorrelated CCA components that are excluded in the EEG reconstruction. A minimum is reached when the last three components are removed. Excluding those components in the reconstruction of the EEG results in the cleaned EEG shown in Fig. 2 (c). The RRMSE of the BSS-CCA processed data was equal to 0.11.

Fig. 2 (d-g) shows the results of applying the low pass filter on the simulated signal $\mathbf{X}$ with cutoff frequencies of 10Hz, 15Hz, 20Hz and 30Hz respectively. Because the frequency spectrum of the muscle artifacts overlaps with that of the brain signal, the low pass filters were not capable of removing the whole artifact without altering the underlying brain activity. The 10 Hz low pass filter removed most of the artifact activity, but also altered the fast activity present in the $\mathbf{B_2}$ EEG recording. Moreover, the

slow activity related to the muscle artifact between t$= 0 - 2$ s was not removed. The higher the cutoff frequency, the less complete the filter removed the muscle artifact. In Fig. 2 (h) the result of removing the artifactual JADE components is shown. As some of the muscle activity was present in ICA components containing brain activity, the removal of the muscle artifact based on the three muscle artifact components, which were selected by visual inspection, was not complete. The RRMSE of the 10HZ , 15Hz, 20Hz and 30Hz low pass filtered data was equal to 0.35, 0.33, 0.36 and 0.51 respectively. The RRMSE of the JADE processed data was equal to 0.25.

The RRMSE as a function of signal to noise ratio on the $\mathbf{B_2}$ brain signal for the different artifact removal techniques (BSS-CCA, ICA(JADE) and the low pass filter with cutoff frequency at 10Hz, 15Hz, 20Hz and 30Hz) is shown in Fig. 5. For high signal to noise ratios the 10 Hz filter performed worst because the brain signal $\mathbf{B_2}$ has mainly alpha (8-13Hz) activity, and consequently, the 10 Hz filter altered the brain signal the most. For lower SNR the performance of the 10 Hz filter compared to the other low pass filters improved as it removed most of the muscle artifact. Compared to the low pass filters, the performance of ICA was better for all signal to noise ratios. The BSS-CCA method outperformed all methods for all SNRs.

The results of the simulation study on the $\mathbf{B_1}$ and $\mathbf{B_3}$ brain signals containing delta and beta activity are similar to that of $\mathbf{B_2}$.

*B. Results on the ictal EEG recording*

This section shows the results obtained by applying the method to the epoch of multichannel EEG containing ictal EEG and muscle artifacts shown in Fig. 6 (a). The (normalized) components as obtained by BSS-CCA are depicted in Fig. 7 (a); their corresponding autocorrelation is shown in Fig. 7 (b). We

observed that the muscle activity is well separated from the ictal component (component 3), and is present in the lowest autocorrelated CCA components. Fig. 6 (b). shows the reconstructed EEG after excluding components (9-21).

For comparison, the low pass filter with cutoff frequency equal to 10 Hz and the JADE algorithm were also applied. Fig. 7 (c) shows the ICA components as obtained by JADE. The selection of the artifactual components by visual inspection becomes difficult because the muscle artifact is present in almost all components and even accompanies the ictal component (component 3). Fig. 6 (c) shows the reconstructed EEG based on excluding suspected muscle artifactual components 4, 7, 8, 13, 14, 16-19, 20 and 21. The muscle artifact removal based on ICA is less complete as the CCA result. Fig. 6 (d) shows the result of the low pass filter (10 Hz). As the frequency of the epileptic activity is lower than 10 Hz, the low pass filter was able to remove most of the muscle artifact without altering the epileptic activity, although a flattening of the signal is observed.

## V. Discussion

In this paper a new method for muscle artifact correction in EEG based on CCA was presented. The method belongs to the group of blind source separation techniques. The performance of the method was tested on simulated data with the aim to remove muscle artifacts without altering the known underlying brain signal. In the simulation study, the performance of the proposed method was compared to the performance of other commonly used methods. A low pass filter with different cutoff frequencies and ICA (JADE) were used as contenders. The BSS-CCA method succeeded in removing the superimposed muscle artifact from the brain signal and performed this task better than its competitors. The frequency spectrum of the muscle artifacts overlaps with that of the brain activity signal, leading to the fact that the low pass filters were not capable of removing the whole artifact without altering the underlying brain signal; as shown in Fig. 2 (d-g). Due to mixing of the muscle artifact in ICA components accounting for

brain activity, BSS-CCA outperformed ICA for all considered signal to noise ratios. Especially for low SNRs it significantly outperformed the other methods (Fig. 5).

The method was applied to a real ictal EEG recording. The muscle artifact was removed with success without altering the ictal EEG. Since the BSS-CCA method is designed for finding maximally autocorrelated signals, the method succeeded in extracting the brain activity without cross-talk of the muscle artifacts, which have very different autocorrelation structures. The muscle artifact removal based on ICA (JADE) was more difficult and again less complete than the BSS-CCA results. The muscle artifact appeared on nearly all ICA components and even accompanied the ictal component, as shown in Fig. 6 (c) and Fig. 7 (c). As a consequence, the selection of the ICA components accounting for the muscle artifact that should be excluded in the reconstruction became difficult or even problematic. These results correspond to the findings of [9] and [10]. Both studies reported that the removal of muscle artifacts with ICA was more difficult and less complete than that of eye artifacts because of mixture of the muscle artifact with EEG components.

Compared to ICA, the BSS-CCA method as a blind source separation technique, has three major advantages. First, the CCA components are ordered according to their autocorrelation, whereas the ICA components have no order. Second, the BSS-CCA algorithm has a much smaller computational time, which makes the algorithm applicable for real-time/on-line performance [15], [17]. In addition, the outcome of the BSS-CCA method is unique, while the ICA output can be different for the same input, depending on the algorithm [15].

The muscle activity is present in the lowest components in the CCA decomposition, which indicates that they are a more noise like signal (small autocorrelation) than the brain activity.

A possible explanation for these observations could be found in the origin of the muscle artifact signal. A muscle is composed out of different motor units, each compromising a motor neuron and a group of

innervated muscle fibres. The recruitment of several motor units when contracting a muscle, leads therefore to a complex electrical activity. This complex signal is a summation of asynchronized potentials, caused by the muscle fibres of the different activated motor units, and results in a low autocorrelated signal. This property is not apparent in other EEG activity and consequently, their time course has a higher autocorrelation compared to that of muscle activity, resulting in a good separation between both signals. These observations can be exploited to automatically remove the muscle artifact from the EEG. Further research will indicate whether thresholding on the autocorrelation index will be sufficient to remove muscle artifact automatically in a large subject group.

For now, we have established a semi-automatic removal of muscle artifacts based on a scrolling system as discussed in the Appendix. In this way, the BSS-CCA method becomes practical for people without experience in the field of BSS and avoids manual selection of a fairly large set of components as in ICA. ICA based techniques that try to avoid this need for a selection of artifactual components, like constrained ICA [12], can not be applied for muscle artifact removal. The cICA needs a good reference signal which is lacking for the muscle artifact. An important factor in the clinical applicability of the method is that the user can easily determine the optimal number of components to be remove by observing the EEG of interest while scrolling, hereby avoiding the removal of important information. The proposed semi-automatic removal of muscle artifact is (already) of great importance in the visual evaluation of ictal EEGs.

In future, the method will be further validated on a large data set of ictal EEGs of patients with different types of epileptic seizures. As in [10], a study will be set up in which a subjective evaluation of the performance of the proposed method, ICA and low pass filters on muscle artifact removal will be carried out.

The modelling of the common dynamics to remove muscle artifacts [24], is dependent on the selection

of the time interval and EEG montage used, and is therefore less general applicable than the proposed method. Moreover, this method removes everything that is not common in the different EEG channels, which is more than only muscle artifact.

As described above, the proposed EMG removal technique can be used to clean the EEG resulting in a better visual interpretation. Next to this visual interpretation tool, the removal of muscle artifact can be used as a preprocessing step to facilitate the signal processing of scalp EEG signals. As the muscle artifacts give rise to false positives in current seizure detection methods [2], preprocessing the EEG with CCA could improve the sensitivity and specificity of the detectors. In the field of seizure prediction, muscle artifacts obscure the interpretation of nonlinear measures when looking for a pre-ictal state in scalp EEG recordings [3]. Removing the artifacts could again improve the performance of the nonlinear measures. Furthermore, ictal dipole source localization would benefit considerably when muscle artifacts were removed. In the domain of the event related potentials, heavily EMG contaminated epochs will no longer need to be rejected. This could lead to a better interpretability of the ERP, especially in cases where a small number of epochs is available, as for example in complex memory tasks.

The method could also be applied to MagnetoEncephaloGraphic (MEG) recordings. In MEG recordings the number of channels is usually much larger than in EEG. As a consequence, the number of components derived by ICA is also much larger and the selection of the artifactual ICA components on visual inspection becomes even more demanding than in EEG. For this reason, the ordered CCA components, could be of major use to MEG signals too.

A previous pilot study has demonstrated that BSS-CCA can be used for separating ocular artifacts that are introduced in background EEG by an artificial mixing process [16]. It is for now, not clear yet whether

BSS-CCA will outperform ICA on eye-blink removal from the EEG. Therefore, a study needs to be undertaken in which the performance of the eye blink removal based on BSS-CCA is compared with the performance of ICA on a large data set containing EEG recordings with eye blinks.
Furthermore, it is not clear whether BSS-CCA will be the method of choice in separation of EEG brain signals in their underlying sources.

## VI. Conclusion

This paper proposed a new method based on canonical correlation analysis as a blind source separation technique for muscle artifact removal. The performance was tested on both synthetic data and ictal EEG. In both cases the method outperformed the low-pass filter and an ICA based technique for muscle artifact removal. Moreover, the fast computational time and the ordering of the CCA components make the proposed method applicable in clinical practice.

## Appendix

### Semi-automatic removal of muscle artifacts

For now, we established a semi-automatic removal of muscle artifacts. CCA components are gradually removed from bottom upwards based on the ordering of the components. An example of use is given below. Fig. 8 (a) shows a 10 s EEG epoch containing, in the opinion of an experienced neurologist, sharp epileptic activity at channel $O2$. This EEG epoch is severely contaminated by repeated eye blinking and muscle artifacts. The activity on channel $O2$ is used to inspect the performance of the semi-automatic artifact removal of muscle artifacts. Fig. 8 (b-c) shows, respectively, the reconstructed EEG with exclusion of the 7, 14 lowest autocorrelated CCA components. By gradually excluding CCA components the muscle artifact is progressively removed, while the background EEG and the sharp epileptic activity on channel $O2$ remain unaltered. Fig. 8 (d) shows the EEG reconstruction with exclusion of the 15 lowest autocorrelated CCA components. Here, too many components were removed as the clinical relevant

activity was removed on channel $O2$. This means that the maximal number of CCA components which can be removed is equal to 14.

This proposed semi-automatic method for muscle artifact removal can be implemented in a Graphical User Interface (GUI) with a scrolling system where moving the scroll bar corresponds to removing/adding CCA components.

## ACKNOWLEDGMENT

The authors would like to thank Guido Van Driel for his assistance in the preparation of the data. W. De Clercq is a research assistant of the Fund for Scientific Research-Flanders (Belgium) (FWO-Vlaanderen). Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. B. Vanrumste is a postdoctoral fellow at the Katholieke Universiteit Leuven in Belgium. Currently he is funded by the Programmatorische Federale Overheidsdienst Wetenschapsbeleid of the Belgian Government. Research is supported by Research Council KUL: GOA-Mefisto 666, IDO /99/003 and /02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc and fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM); IWT: PhD Grants; Belgian Federal Science Policy Office IUAP P5/22 (‘Dynamical Systems and Control: Computation, Identification and Modelling’); EU: PDT-COIL, BIOPATTERN, ETUMOUR.

## REFERENCES

[1] M. Stecker, “The effects of automatic artifact rejection on evoked potential recordings.” *Computers in Biology and Medicine*, vol. 32, pp. 247–259, 2002.

[2] M. Saab and J. Gotman, “A system to detect the onset of epileptic seizures in scalp EEG,” *Clinical Neurophysiology*, to appear.

[3] W. De Clercq, P. Lemmerling, S. Van Huffel, and W. Van Paesschen, “Correspondence : Anticipation of epileptic seizures from standard EEG recordings,” *The Lancet*, vol. 361, pp. 970–971, 2003.

[4] I. Goncharova, D. McFarland, T. Vaughan, and J. Wolpaw, “EMG contamination of EEG: Spectral and topographical characteristics,” *Clinical Neurophysiology*, vol. 114, pp. 1580–1593, 2003.

[5] G. Worell, E. So, J. Kazemi, T. O’Brien, R. Mosewich, G. Cascino, F. Meyer, and W. Marsh, “Focal ictal beta discharge on scalp EEG predicts excellent outcome of frontal lobe epilepsy surgery,” *Epilepsia.*, vol. 43, no. 3.

[6] R. Croft and R. Barry, “EOG correction: Which regression should we use?” *Psychophysiology*, vol. 37, pp. 123–125, 2000.

[7] G. Wallstrom, R. Kass, A. Miller, J. Cohn, and N. Fox, “Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods,” *International Journal of Psychophysiology*, vol. 53, pp. 105–119, 2004.

[8] T.-P. Jung, C. Humphries, T. Lee, S. Makeig, M. McKeown, V. Iragui, and T. Sejnowski, “Extended ICA removes artifacts from electroencephalographic recordings.” *Advances in Neural Information Processing Systems*, vol. 10, pp. 894–900, 1998.

[9] H. Nam, T.-G. Yim, S. Han, J.-B. Oh, and S. Lee, “Independent component analysis of ictal EEG in medial temporal lobe epilepsy,” *Epilepsia*, vol. 43, no. 2, pp. 160–164, 2002.

[10] E. Urrestarazu, J. Iriarte, M. Alegre, M. Valencia, C. Viteri, and J. Artieda, “Independent component analysis removing artifacts in ictal recordings,” *Epilepsia*, vol. 45, no. 9, pp. 1071–1078, 2004.

[11] S. Mozaffar and D. Petr, “Artifact Extraction from EEG Data Using Independent Component Analysis,” Information Telecommunication and Technology Center, University of Kansas, Lawrence, KS, Tech. Rep. ITTC-FY2003-TR-03050-02, Dec. 2002.

[12] C. J. James and J. Gibson, “Temporally constrained ICA: An applicaton to artifact rejection in electromagnetic brain signal analysis,” *IEEE Trans. Biomed. Eng.*, vol. 50, no. 9, pp. 1108–1116, Sept. 2003.

[13] W. Lu and J. Rajapakse, “ICA with reference,” in *Proceedings 3rd In. Conf. Independent Component Analysis and Blind Signal Separation:ICA 2003*, 2003, pp. 120–125.

[14] H. Hotelling, “Relations between two sets of variates.” *Biometrika*, vol. 28, pp. 321–377, 1936.

[15] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, “Exploratory fMRI analysis by autocorrelation maximization.” *NeuroImage*, vol. 16, no. 2, pp. 454–464, 2002.

[16] J. Knight, “Signal fraction analysis and artifact removal in EEG,” M. Sc. thesis, Colorado State University, 2003.

[17] M. Borga and H. Knutsson, “A canonical correlation approach to blind source separation.” Dept. of Biomedical Engineering, Linköping University, Sweden, Tech. Rep. LiU-IMT-EX-0062, 2001.

[18] O. Friman, “Adaptive analysis of functional MRI data.” *PhD-Thesis, Dept. of biomedical engineering, Linköpings universitet, Linköping, Sweden*, 2003.

[19] O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson, “Detection of neural activity in functional MRI using canonical correlation analysis.” *Magnetic Resonance in Medicine*, vol. 45, pp. 323–330, 2001.

[20] R. Plonsey and D. Heppner, “Considerations of quasistationarity in electrophysiological systems.” *Bulletin of Mathematical Biophysics*, vol. 29, no. 4, pp. 657–664, 1967.

[21] M. Nuwer, G.Comi, R. Emmerson, *et al.*, “IFCN standards for digital recording of clinical EEG,” *Electroencephalogr Clin Neurophysiol*, vol. 106, pp. 259–261, 1998.

[22] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique usingsecond order statictics,” *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, 1997.

[23] J.-F. Cardoso, “Higher-order contrast for independent component analysis,” *Neural Comput*, vol. 11, pp. 157–192, 1999.

[24] W. De Clercq, J.-M. Papy, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, “Moddeling common dynamics in multichannel signals with application to artifact and background removal in EEG recordings,” *accepted for publication in IEEE Trans. Biomed. Eng.*

## LIST OF FIGURES

1 (a), (b) and (c) The three selected independent components and their corresponding field distributions. (d) The resulting EEG epoch containing all three independent muscle activities. 21

2 (a) Simulation matrix $\mathbf{X}$ containing: alpha activity $\mathbf{B_2}$ and superimposed muscle activity $\mathbf{M}$ with SNR = 0.65; (b) Original EEG epoch $\mathbf{B_2}$; (c) BSS-CCA processed data; (d), (e), (f) and (g) low pass filtered data with cutoff frequency 10Hz, 15Hz, 20Hz and 30Hz respectively; (h) JADE processed data . . . . . 22

3 (a) CCA components (normalized) obtained by applying BSS-CCA on the 10 sec EEG epoch shown in Fig 2 (a); (b) The autocorrelation of the CCA components; (c-e) Distribution of the three lowest autocorrelated CCA components ((c) component 21, (d) component 20, (e) component 19). . . . . 23

4 The RRMSE as a function of the number of lowest autocorrelated CCA components excluded in the reconstruction of the EEG. . . . . 24

5 The RRMSE as a function of SNR on the EEG epoch $\mathbf{B_2}$ containing alpha (8-13Hz) activity. 25

6 (a) Original 10 s EEG epoch; (b) EEG reconstruction after excluding artifactual CCA components. Notice that the muscle artifact was removed completely, while preserving the sharp quality ictal theta activity in the right temporal lobe ($T2$ (=right spenoidal electrode), $F8$, and lesser degree $T4$ and $T6$). (c) EEG reconstruction based on JADE components. (d) EEG filtered by the low pass filter with cutoff frequency equal to 10 Hz. Notice that the 10 Hz filter has removed not only the muscle artifact, but also the sharp, ictal quality of the recruiting right temporal lobe theta activity. . . . . 26

7 (a) Components (normalized) obtained by applying BSS-CCA on the ictal EEG recording shown in Fig. 6 (a). (b) The autocorrelation of the CCA components. (c) ICA components obtained by applying JADE on the same recording. . . . . 27

8 (a) Original EEG with sharp epileptic activity on channel $O2$; (b) EEG after exclusion of components (15-21); (c) EEG after exclusion of components (8-21); (d) EEG after exclusion of components (7-21). . . . . 28

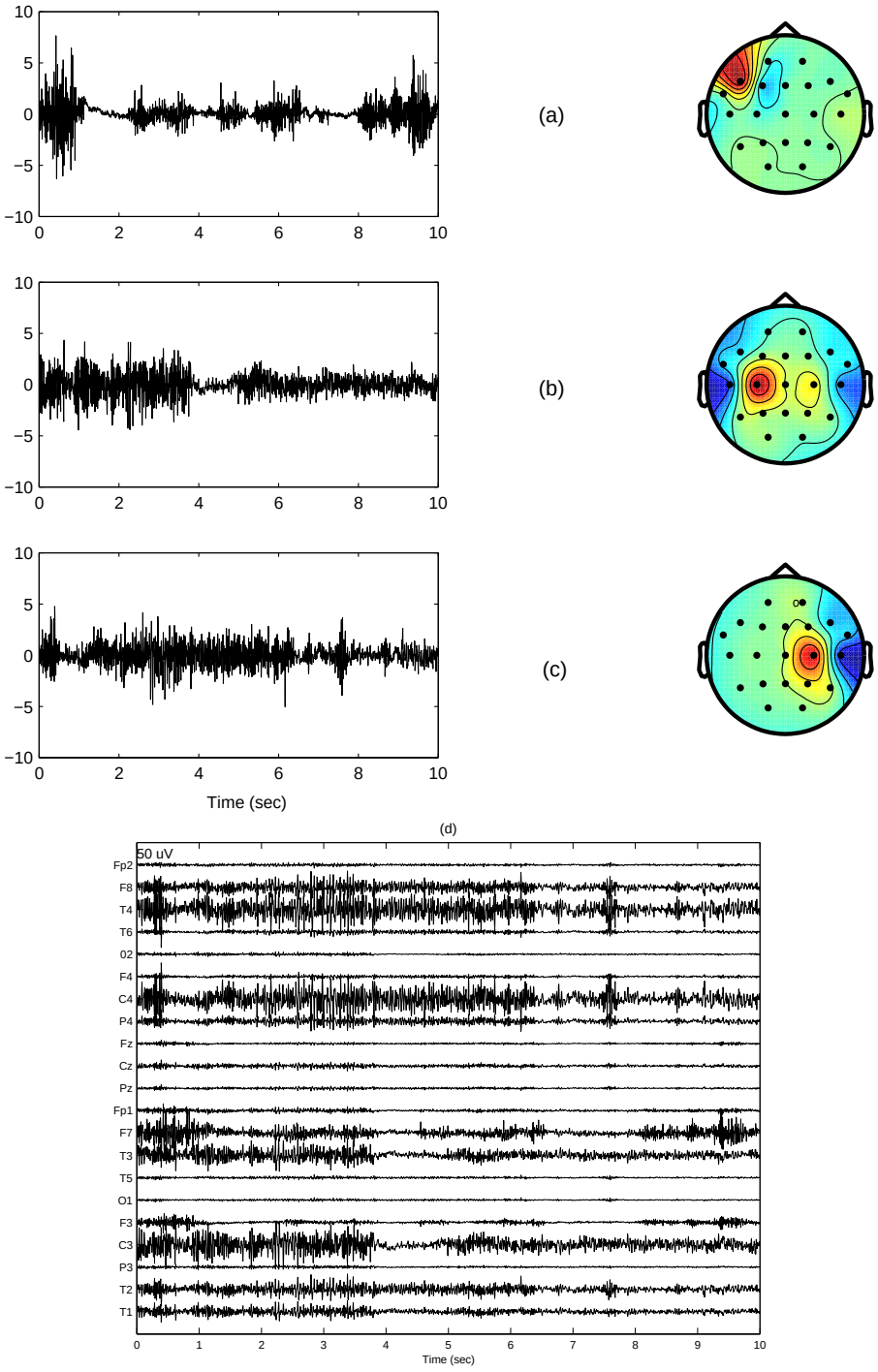



Fig. 1. (a), (b) and (c) The three selected independent components and their corresponding field distributions. (d) The resulting EEG epoch containing all three independent muscle activities.

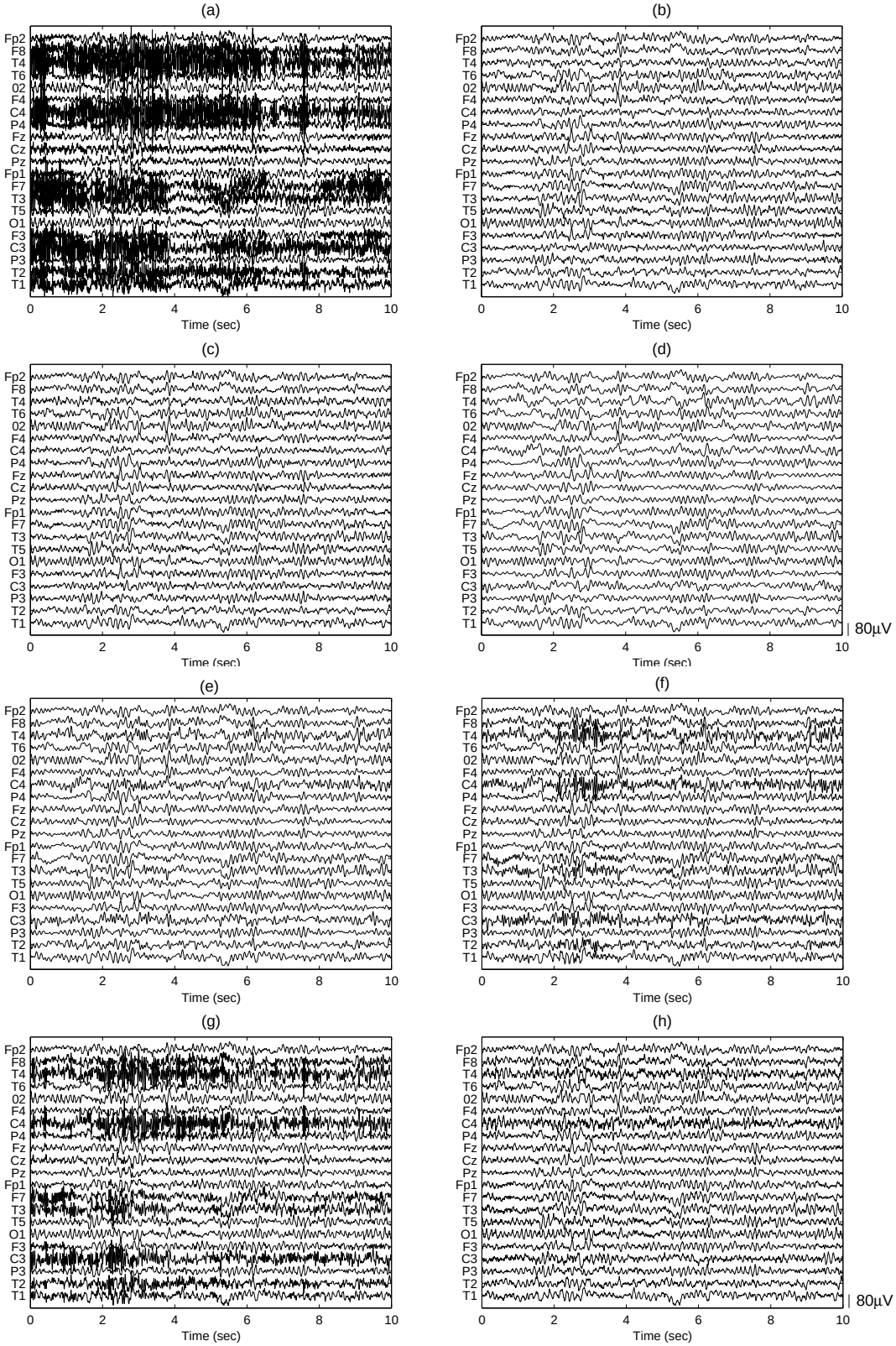



Fig. 2. (a) Simulation matrix $\mathbf{X}$ containing: alpha activity $\mathbf{B_2}$ and superimposed muscle activity $\mathbf{M}$ with SNR = 0.65; (b) Original EEG epoch $\mathbf{B_2}$; (c) BSS-CCA processed data; (d), (e), (f) and (g) low pass filtered data with cutoff frequency 10Hz, 15Hz, 20Hz and 30Hz respectively; (h) JADE processed data

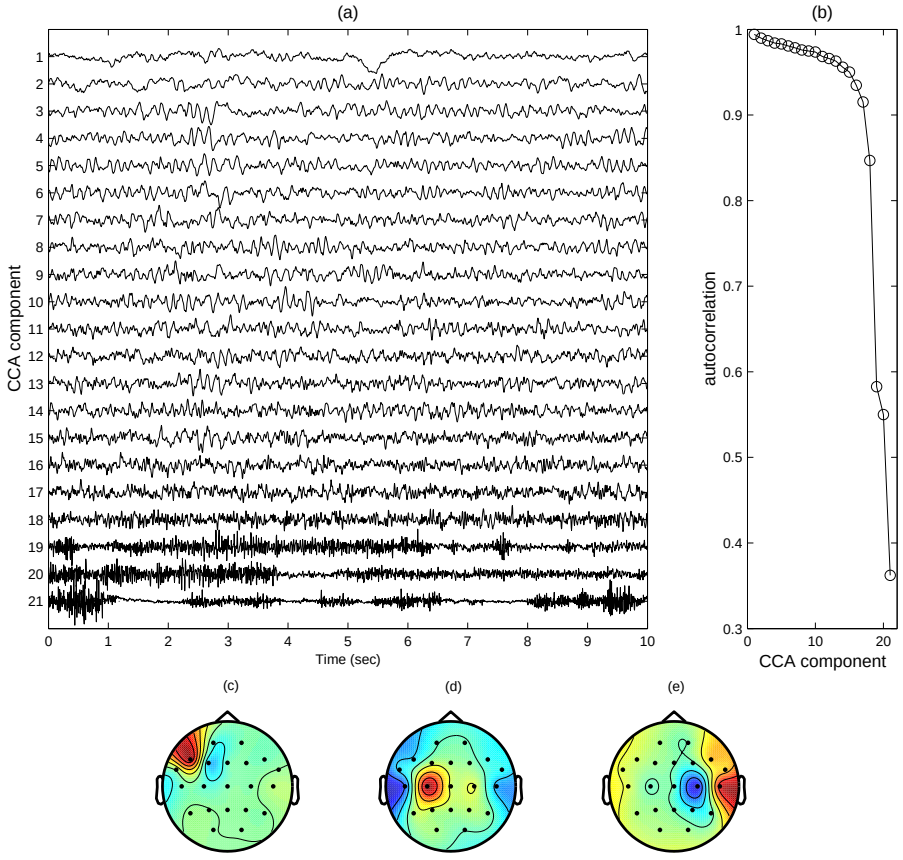



Fig. 3. (a) CCA components (normalized) obtained by applying BSS-CCA on the 10 sec EEG epoch shown in Fig 2 (a); (b) The autocorrelation of the CCA components; (c-e) Distribution of the three lowest autocorrelated CCA components ((c) component 21, (d) component 20, (e) component 19).

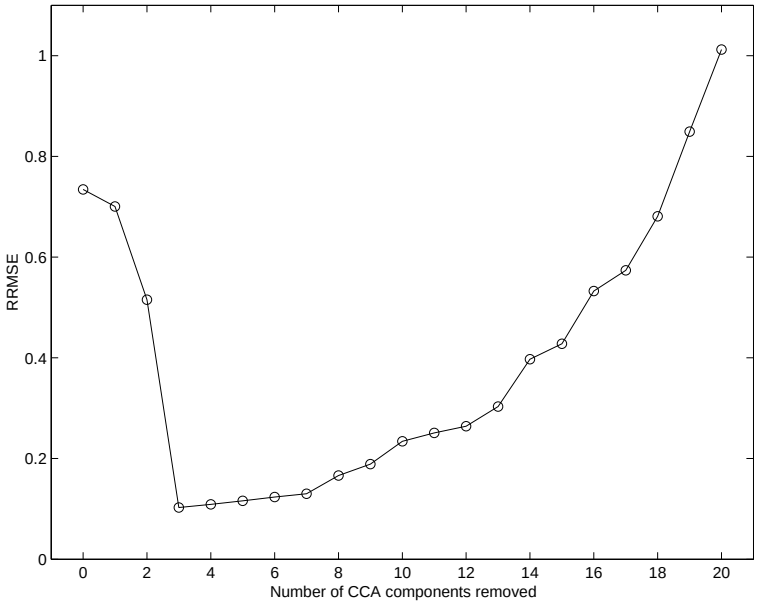



Fig. 4. The RRMSE as a function of the number of lowest autocorrelated CCA components excluded in the reconstruction of the EEG.

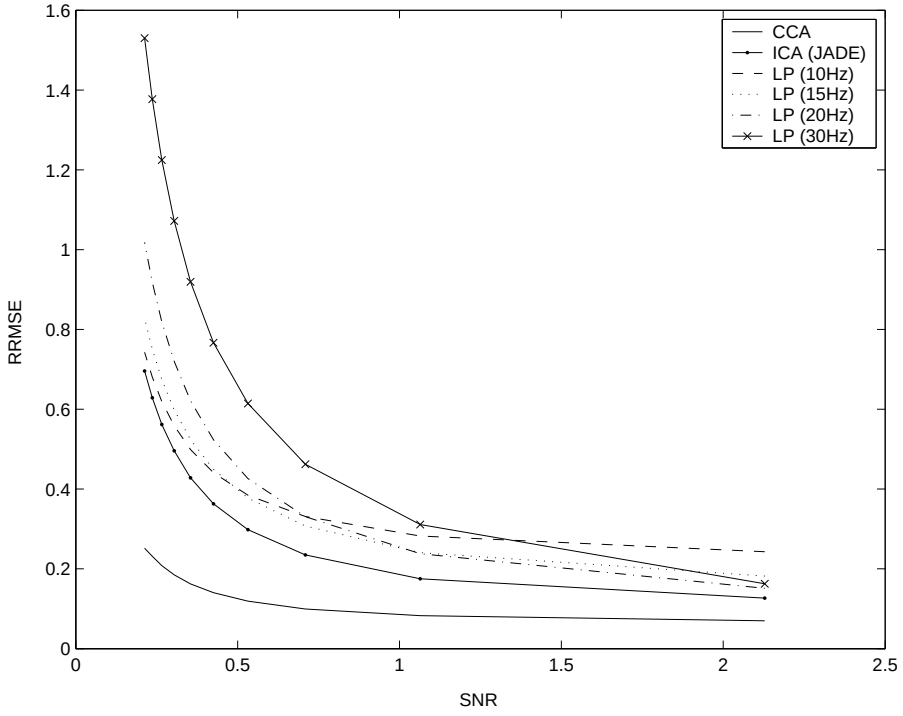



Fig. 5. The RRMSE as a function of SNR on the EEG epoch $\mathbf{B_2}$ containing alpha (8-13Hz) activity.

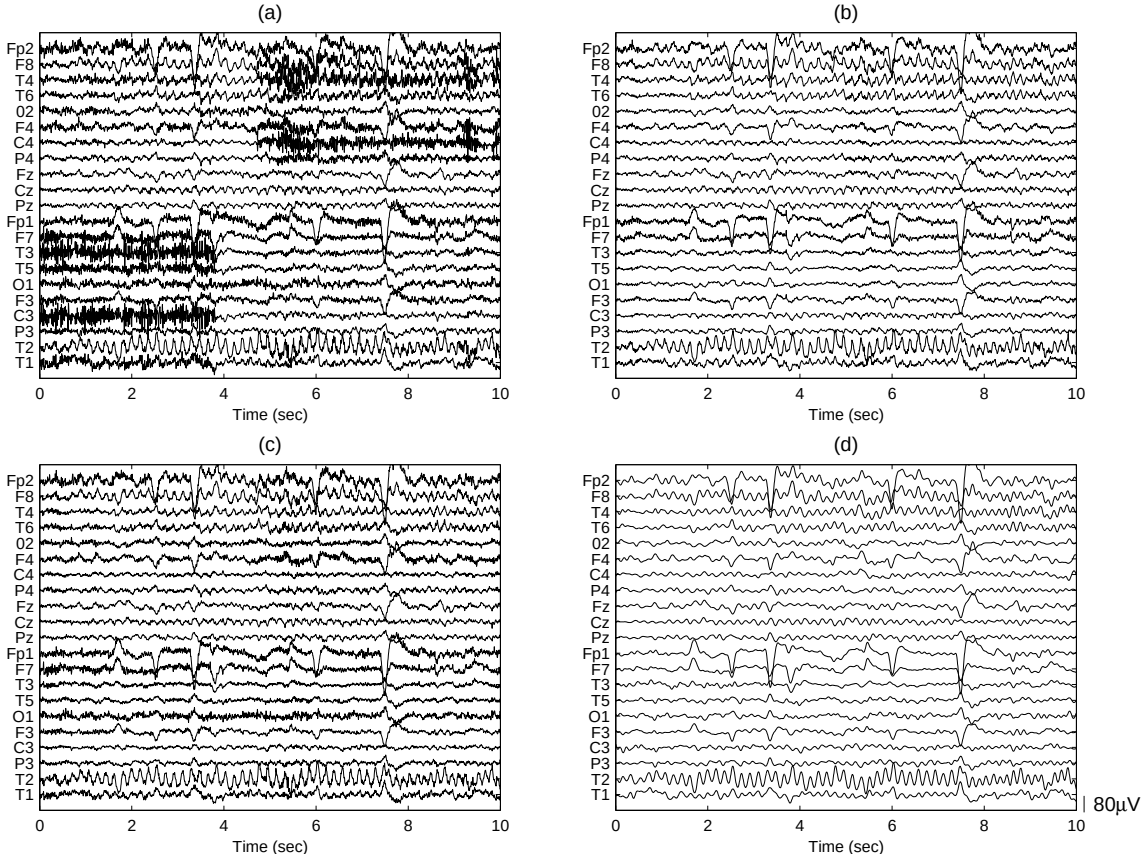



Fig. 6. (a) Original 10 s EEG epoch; (b) EEG reconstruction after excluding artifactual CCA components. Notice that the muscle artifact was removed completely, while preserving the sharp quality ictal theta activity in the right temporal lobe ($T2$ (=right spenoidal electrode), $F8$, and lesser degree $T4$ and $T6$). (c) EEG reconstruction based on JADE components. (d) EEG filtered by the low pass filter with cutoff frequency equal to 10 Hz. Notice that the 10 Hz filter has removed not only the muscle artifact, but also the sharp, ictal quality of the recruiting right temporal lobe theta activity.

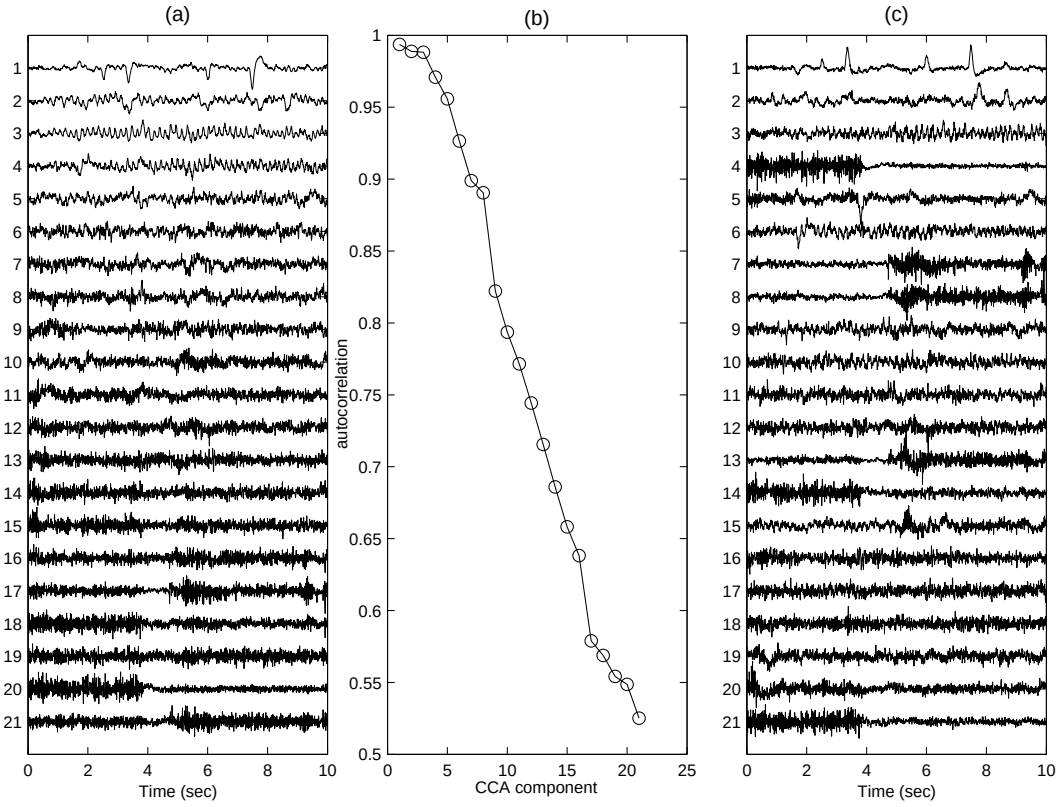



Fig. 7. (a) Components (normalized) obtained by applying BSS-CCA on the ictal EEG recording shown in Fig. 6 (a). (b) The autocorrelation of the CCA components. (c) ICA components obtained by applying JADE on the same recording.

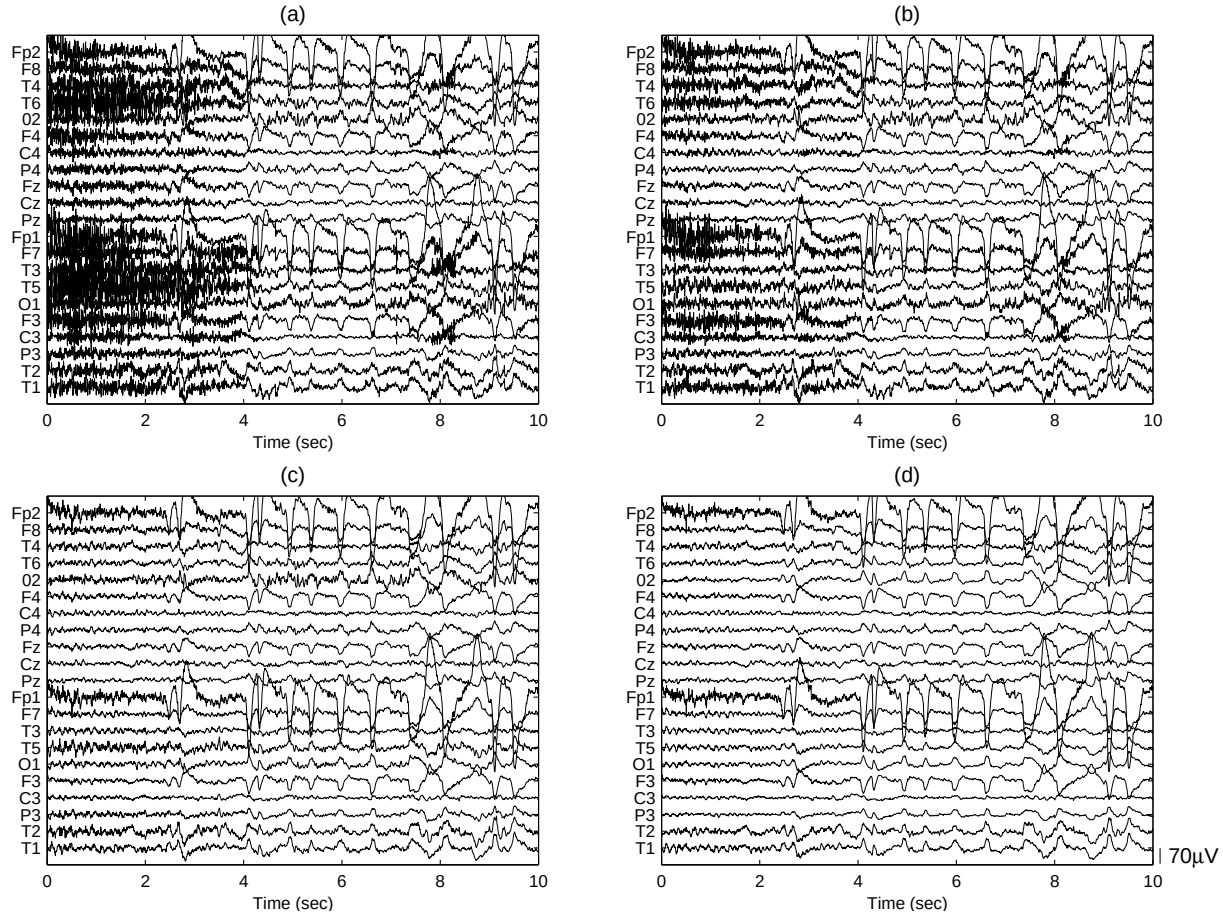



Fig. 8. (a) Original EEG with sharp epileptic activity on channel $O2$; (b) EEG after exclusion of components (15-21); (c) EEG after exclusion of components (8-21); (d) EEG after exclusion of components (7-21).