

Classification of Artifactual ICA Components

Michael Tangermann^a, Irene Winkler^a, Stefan Haufe^a, Benjamin Blankertz^{a,b}

^aDept. Machine Learning, Berlin Institute of Technology, Berlin, Germany

^bIDA, Fraunhofer Institute FIRST, Berlin, Germany

Correspondence: Michael Tangermann, Berlin Institute of Technology, FR 6-9, Franklinstr. 28/29, 10587 Berlin, Germany.
E-mail: schroedm@cs.tu-berlin.de, phone +49 30 31478624, fax+49 30 31478622

Abstract. The analysis of EEG signals for the use in BCI systems and for mental state monitoring applications is often impeded by artifacts caused by muscular activity or external technical sources. A promising approach for the reduction or removal of artifacts is based on methods of Blind Source Separation (BSS), which transform the original EEG signal into independent source components. In order to avoid the time-consuming hand rating of sources into artifactual and non-artifactual components, an automated method for their classification is proposed. Applying state of the art machine learning algorithms and nonlinear classification with a Support Vector Machine (SVM), the automated method shows a high level of agreement (90.5%) on unseen data with ratings of human experts.

Keywords: Brain-Computer Interface, EEG analysis, Blind Source Separation, Independent Component Analysis, Artifact Removal, Machine Learning

1. Introduction

Signals of the electroencephalogram (EEG) can reflect the electrical background activity of the brain as well as the activity which is specific for a cognitive tasks during an experiment. As the electrical field generated by neural activity is very small, it can only be recognized by EEG if large assemblies of neurons show a similar behavior. Resulting neural EEG signals are in the range of microvolts only and can easily be masked by artifactual sources. Typical artifacts of the EEG are caused either by the non-neural physiological activities of the subject or by external technical sources. Eye blinks, eye movements, muscle activity in the vicinity of the head (e.g. face muscles, jaws, tongue, neck), heart beat, pulse and Mayer waves are examples for physiological artifact sources, while swaying cables in the magnetic field of the earth, line humming, power supplies or transformers can be the cause of technical artifacts.

Brain-Computer Interfaces (BCI) are based on the single trial classification of the ongoing EEG signal. The exclusion of artifacts is of special interest for BCI applications, as the intended or unconscious use of artifacts for BCI control are usually not desirable. Furthermore, as averaging methods have to be avoided in these real-time systems BCIs rely on relatively clean EEG signals. The same holds true for other Mental State Monitoring (MSM) applications, that monitor a subject's mental state continuously and on a fine granular time resolution to detect changes e.g. of wakefulness, responsiveness or mental workload as early as possible [Müller et al., 2008].

If artifactual signal components and neural activity of interest are not systematically co-activated due to a disadvantageous experimental design, methods of Blind Source Separation (BSS) like Independent Component Analysis (ICA) are promising approaches for their separation [Delorme et al., 2007]. A common approach is the transformation of the EEG signals into a space of independent source components, the hand-selection of non-artifactual neural sources and the reconstruction of the EEG without the artifactual components. While assumptions for the application of ICA methods are only approximately met in practice (linear mixture of independent components, stationarity of the sources and the mixture, and prior knowledge about the number of components), their application usually leads to a good separation, with only a small number of hybrid components that contain both, artifacts and neural signals.

Existing semi-automatic approaches require user interaction for ambiguous or outlier components [Barbati et al., 2004; Delorme et al., 2007]. While fully automated methods were proposed for the classification of eye artifacts [Romero et al., 2004; Shoker et al., 2005], these methods do not easily generalize to non-eye artifacts or even require the additional recording of the electrooculogram (EOG) [James et al., 2003; Joyce et al., 2004]. Existing approaches for the general classification of different artifacts types were reported for EEG data of epileptic patients [LeVan et al., 2006], where the authors report an MSE of approx. 20% for their system, and for a system based on non-linear SVM classification for four fixed artifact classes. For the SVM-based system classification error below 10% is reported if a combination of two unmixing methods (AMUSE and Infomax) is applied and if dedicated artifact recordings are available for the classifier training [Halder et al., 2007]. But even if such optimized conditions are present, difficulties of separating muscle artifact components from neural components are common [Halder et al., 2007].

2. Material and Methods

2.1 Data

Data from 12 healthy right-handed male subjects were used to train and to test the proposed automated component classification method. Every subject participated in one EEG recording session of approx. 5 hours duration. EEG was recorded from 121 approx. equidistant sensors and high pass filtered at 2 Hz. During a session, 4 repeated blocks of 3 different conditions (C0, C1, C2) were performed. During all three conditions, subjects performed a forced-choice left or right key press reaction time task upon two auditory stimuli in an oddball paradigm. The key press actions were performed with micro switches attached to the index fingers. During condition C0 subjects had to gaze at a fixation cross without any further visual task. Condition C1 introduced an additional distraction, as a video of a driving scene had to be watched passively on a screen. Condition C3 introduced an additional second task: subjects infrequently had to follow simple lane change instructions and control a steering wheel. By design, EEG recordings under condition C3 were inevitably more prone to muscle and eye artifacts, while C2 possibly stimulated eye movement artifacts, but not muscle artifacts. However, all subjects had been instructed during all conditions to avoid producing artifacts.

2.2 Preprocessing, unmixing and data split

To avoid the artificial split of signal components due to the high dimensionality of the data, the separation of the EEG signals by ICA was preceded by a dimensionality reduction from 121 EEG channels in the sensor space into $k=30$ PCA components. This choice of k was based on previous experience, but was probably not the optimal choice. The TDSEP algorithm [Ziehe et al. 2004] was used to transform the 30 PCA components into 30 independent source components. Compared to other ICA methods, TDSEP takes the temporal structure of components into account. Given a multivariate time series X , TDSEP amounts to finding a demixing $WX=S$ which leads to minimal cross-covariances between all pairs of components of S . Each source time series s_n is now characterized by a spatial filter w_n and a spatial pattern a_n , where

$$[s_1, \dots, s_n]^T = S, [w_1, \dots, w_n]^T = W, [a_1, \dots, a_n] = A = W^{-1}$$

All resulting source components were hand labeled by 2 experts into artifactual and non-artifactual components. For the training of the proposed automated classification method, 23 EEG recordings of 10 minutes duration were taken from the first experimental blocks only, leading to 690 labeled source components. Neural components and artifact components were approx. equally distributed (46% vs. 54%). Figure 1 shows typical examples of two artifacts and a neural component.

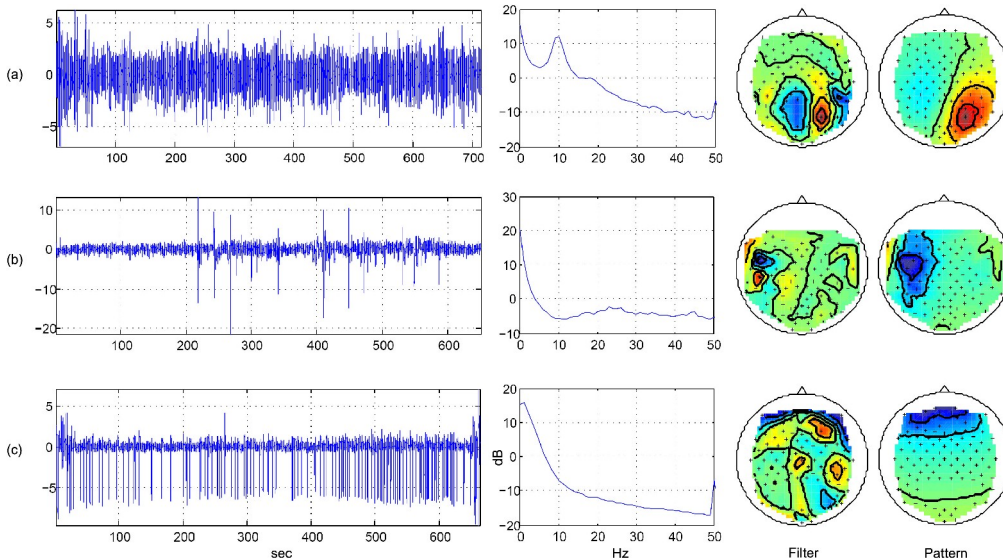


Figure 1: Time series (first column), spectrum (second column), filter map (third column) and filter pattern (fourth column) of three components. The first row (a) shows an alpha generator in the occipital lobe. The second row (b) shows a rare muscle artifact component with an increased spectrum in higher frequencies. The third line (c) shows an eye artifact component that appears regularly, has an increased spectrum in lower frequencies and a typical front-back distribution in the pattern.

The trained classifier was tested on 36 unseen EEG recordings from the third experimental blocks. Among these 1080 source components were 47% neural components and 53% artifact components.

3. Classification and Features

For classification, a non-linear SVM with Gaussian kernel was chosen. A suitable regularization parameter $C=1$ and the kernel width $\gamma=1$ were selected in a 5x10 cross validation setup. The final classifier was trained on the full training data (690 examples) on the following features, which were extracted either from the time series of the components, from their frequency spectrum or from the spatial distribution of the component's patterns.

3.1 Features derived from a component's time series

Average Local Skewness

This feature is calculated as the logarithm of the average local skewness of time intervals of 15s duration.

3.2 Features derived from a component's spectrum

λ and Fit Error

These two features describe the deviation of a component's spectrum from a prototypical "1/frequency" curve and its shape. The parameters $k_1, k_2, \lambda > 0$ of the curve

$$f : f \rightarrow \frac{k_1}{f^\lambda} - k_2$$

are determined by three points of the log spectrum: (1) value at 2 Hz, (2) local minimum in the band 5-13 Hz, (3) local minimum in the band 33-39 Hz. The logarithm of λ and of the Fit Error of f to the real spectrum are used as features.

Band Power 8-13 Hz

The average log band power of the band between 8 and 13 Hz.

3.3 Features derived from a component's pattern

Current Density Norm

ICA itself does not provide information about the locations of the sources S . However, ICA patterns can be interpreted as EEG potentials for which a physical model is given by

$$a = Fz \quad (2)$$

Here, z are unknown source dipoles at M locations in the brain and F describes the mapping from sources to sensors, which is determined by the shape of the head and the conductivities of brain, skull and skin tissues. We consider $M=2142$ sources which are arranged in a 1cm grid. Source estimation can only be done under additional constraints since $M \gg N$. Commonly, the source distribution with minimal l_2 -norm (i.e., the "simplest" solution) is sought [Hämäläinen et al., 1994]. This leads to estimates

$$\min_z \|Fz - a\|^2 + \lambda \|\Gamma s\|^2 = (F^T F + \lambda \Gamma^T \Gamma)^{-1} F^T a := J_\lambda a,$$

where Γ approximately equalizes the cost of dipoles at different depths [Haufe et al., 2008] and λ defines a trade-off between the simplicity of the sources and the fidelity of the model. Since Eq. (2) models only cerebral sources, it is natural that noisy patterns and patterns originating outside the brain can only be described by rather complicated sources, which are characterized by a large l_2 -norm. We propose to use $f = \log \|\Gamma J_\lambda \tilde{a}\|$ as a feature for dividing physiological from noisy or artifactual patterns. Here $\tilde{a} = a/\|a\|$ are normalized ICA patterns and λ is chosen from $\{0, 1, 10, 100, 1000\}$.

Range Within Pattern

The logarithm of the difference between the minimal and the maximal activation in a pattern.

Border Activation

This binary feature captures the spatial distribution of patterns. It is defined as 1 if either the global maximum of the pattern is located at one of the outmost electrodes of the setup in Fig. 2 (right), or if the local maximum of an electrode group in Fig. 1 (left) is located at the outmost electrode of the group and if that local maximum deviates at least 2 standard deviations from the group average. Otherwise the feature is defined as -1.

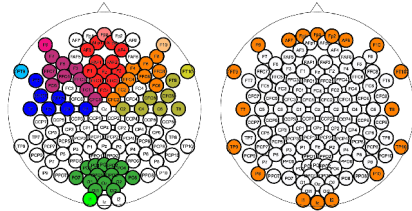


Figure 2: Electrode groups (left) and outer electrodes (right) used to determine the feature Border Activation.

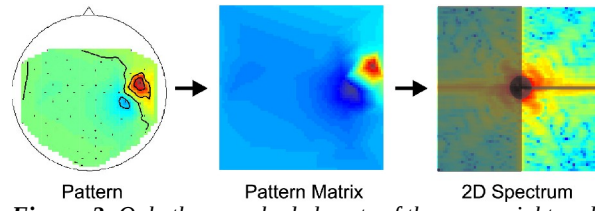


Figure 3: Only the non-shaded parts of the upper-right and lower-right quadrant of the 2D-spectrum are averaged for 2DDFT.

2DDFT

The spatial frequency of a pattern is described by means of a two-dimensional discrete Fourier transformation. As a first step, the pattern is linearly interpolated to a quadratic 64x64 pattern matrix. The feature 2DDFT is the average logarithmic band power of higher frequencies of the 1st and 4th quadrant (see Fig. 3) of the 2D-Fourier spectrum of the pattern matrix.

Central Activation

This feature describes the characteristic of the spatial distribution of a component's pattern. It is calculated as the logarithm of the average activation of a group of 13 central electrodes that were places symmetrically around the vertex.

4. Results

The kernel width $\gamma=1$ of the trained SVM is relatively small compared to the scale of the data, which was within within (-10,10) units for every feature, which underlines the usefulness of a non-linear method compared to a linear solution (if linear classification is sufficient for solving the task, much larger kernel widths tend to deliver good solutions).

Table 1. Test error (MSE) and SVM hyper parameters (selected by 5x10 cross validation) for the 9 single features and for the combined classification.

Features	Regularization Parameter C	Kernel Width γ	MSE of SVM
Current Density Norm	1	1	0.143
Range Within Pattern	1	0.1	0.150
Avg. Local Skewness	1	1	0.293
λ	0.1	1	0.182
Band Power 8-13 Hz	10000	1	0.168
Fit Error	1	1	0.397
Border Activation	0.1	0.001	0.382
2DDFT	100	100	0.188
Central Activation	10	1	0.436
9 Pooled Features	1	1	0.094

Testing the trained classifier ($C=1, \gamma=1$) on unseen data (1080 examples from experimental blocks 3) leads to an error of 9.5% only (MSE), which corresponds to a high agreement with the expert's labeling. It is interesting to take a closer look on the performance of single features. Allowing a (potentially slightly over-fitted) optimal choice of hyper parameters for every single feature, the best one, *Current Density Norm*, leads to a MSE of 14.3% on the test data. The selected hyper parameters and classification results for every single feature and for the combined classification are given in Table 1. The combination of the 9 features from all three domains improves the error substantially compared to even the best single feature. This shows that features which are far from optimal in single classification have a positive contribution in combination with other features.

Looking at the complete test set of 1080 components, 83 of them were misclassified as artifacts and 19 components were misclassified as neural sources. A detailed visual analysis of these cases reveals, that most of them were mixed components that contained both, artifacts and brain activity. In some rare cases, mislabeled examples could be identified. There was no systematic bias in the question whether

eye artifacts or muscle artifacts were subject to misclassification. Both artifact types were handled equally well by the classifier.

5. Discussion

Compared to the classification results of Halder et al. [Halder et al., 2007], who reported 8% error for muscle artifacts and 1% error for eye artifacts, the classification error of our solution (9.5%) is slightly higher. A major difference between the two studies rests on the way the training data was generated. Halder et al. reported, that subjects had specifically been instructed to *produce* a number of artifacts under controlled conditions. It can be speculated that such a training set contains stronger artifacts and less erroneous labels. Nevertheless, the results of Halder et al. were generated based on EEG recordings of only 16 electrodes. Contrary to this study, our approach for the classifier training was based on data from an ongoing EEG experiment, where subjects had been instructed to *avoid* the generation of artifacts. However, the performance of a system on the classification task has to be judged in the light of the fact that inter-expert disagreements on EEG signals are above 10% [Klekowicz et al. 2008].

The optimal rejection level for hybrid components varies dependent on the application. For the application in BCI, a rather heterogeneous result can be expected in terms of BCI performance. Subjects that (unconsciously) use artifacts to control a BCI system might show decreased performance if artifacts or hybrid components are removed, while other subjects could profit from clearer brain signals after the removal of artifactual components and show higher BCI performance.

The method presented in this study delivers a substantial contribution for the BCI community and the EEG community in general, as a reliable tool for the removal of artifacts is proposed that works without explicit recordings of artifacts. The method is suitable for application in offline analysis and in online experiments due to its low computational demands.

Acknowledgments

This work is supported by a BMBF grant No. 01GQ0850 and No. 01IB001A, a DFG grant MU 987/3-1, and by the European ICT Programme Project FP7-224631 and ICT-216886. This abstract only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

References

- Barbati G., Porcaro C., Zappasodi F., Rossini PM., Tecchio F. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. In: *Clinical Neurophysiology* 115 (2004), 1220-1232.
- Delorme A., Sejnowski T., Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. In *NeuroImage*, vol. 34, no. 4, pp. 1443-1449, 2007.
- Hämäläinen MS, Ilmoniemi RJ. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.*, 32:35-42, 1994.
- Halder S., Bensch M., Mellinger J. Bogdan M., Kübler A., Birbaumer N., Rosenstiel W. Online Artifact Removal for Brain-Computer Interfaces Using Support Vector Machines and Blind Source Separation. In: *Computational Intelligence and Neuroscience* 7 (2007), Nr. 3, 1-10.
- Haufe S., Nikulin VV., Ziehe A., Müller KR., Nolte G. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage*, 42(2):726-738, 2008.
- James CJ., Gibson OJ. Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis. In: *IEEE Trans Biomed Eng* Bd. 50, 2003, S. 1108-1116.
- Joyce CA., Gorodnitsky IF., Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. In: *Psychophysiology* 41 (2004), 331-325.
- Klekowicz H, Malinow U, Niemcewi S, Wakarow A, Wolynczyk-Gmaj D, Piotrowski T and Durka P. Automatic analysis of sleep EEG. In *Frontiers in Neuroinformatics. Conference Abstract: Neuroinformatics 2008*.
- LeVan P., Urrestarazu E., Gotman J. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification. In: *Clin. Neurophysiology* 117 (2006), 912-927.
- Müller KR., Tangermann M., Dornhege G., Krauledat M., Curio G., and Blankertz B. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods* 167(1):82-90, 2008.
- Romero S., Mañanas MA., Riba J., Giménez S., Clos S., Barbanoj MJ. Evaluation of an automatic ocular filtering method for awake spontaneous EEG signals based on Independent Component Analysis. In: 26th Annual Int. Conf. of the Engineering in Medicine and Biology Society (EMBS), 2004, S. 925-928.
- Shoker L, Sanei S., Chambers J. Artifact Removal from Electroencephalograms using a hybrid BSS-SVM algorithm. In: *IEEE Signal Processing Letters* 12 (2005), Nr. 10, S. 721-724.
- Ziehe A., Laskov P., Nolte G., Müller KR. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. In *Journal of Machine Learning Research*, 5:777-800, Jul 2004.