Journal of Neuroscience Methods 243 (2015) 84–93

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

EEG artifact elimination by extraction of ICA-component features using image processing algorithms

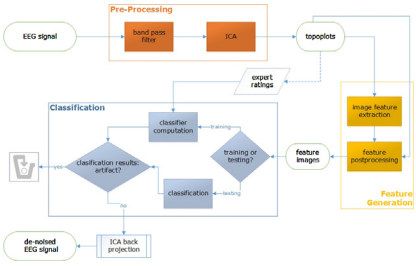
T. Radüntz[a,*], J. Scouten[a], O. Hochmuth[b], B. Meffert[b]

[a] Federal Institute for Occupational Safety and Health, Mental Health and Cognitive Capacity, Nöldnerstr. 40-42, 10317 Berlin, Germany
[b] Humboldt-Universität zu Berlin, Department of Computer Science, Rudower Chaussee 25, 12489 Berlin, Germany

HIGHLIGHTS

- Machine-driven EEG artifact removal through automated selection of ICs is proposed.
- Feature vectors extracted from IC via image processing algorithms are used.
- LDA classification identifies range filter as powerful feature (accuracy rate 88%).
- The method does not depend on direct recording of artifact signals.
- The method is not limited to a specific number of artifacts.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 July 2014
Received in revised form 26 January 2015
Accepted 28 January 2015
Available online 7 February 2015

MSC:
68-04
68U10
68U99
68T10
92C50
92C55

Keywords:
Independent component analysis
ICA
EEG
Artifact elimination
Image processing
Local binary patterns
Range filter
Geometric features

ABSTRACT

Artifact rejection is a central issue when dealing with electroencephalogram recordings. Although independent component analysis (ICA) separates data in linearly independent components (IC), the classification of these components as artifact or EEG signal still requires visual inspection by experts.

In this paper, we achieve automated artifact elimination using linear discriminant analysis (LDA) for classification of feature vectors extracted from ICA components via image processing algorithms.

We compare the performance of this automated classifier to visual classification by experts and identify range filtering as a feature extraction method with great potential for automated IC artifact recognition (accuracy rate 88%). We obtain almost the same level of recognition performance for geometric features and local binary pattern (LBP) features.

Compared to the existing automated solutions the proposed method has two main advantages: First, it does not depend on direct recording of artifact signals, which then, e.g. have to be subtracted from the contaminated EEG. Second, it is not limited to a specific number or type of artifact.

In summary, the present method is an automatic, reliable, real-time capable and practical tool that reduces the time intensive manual selection of ICs for artifact removal. The results are very promising despite the relatively small channel resolution of 25 electrodes.

* Corresponding author. Tel.: +49 30515484418.
E-mail address: raduentz.thea@baua.bund.de (T. Radüntz).
URL: http://www.baua.de (T. Radüntz).

http://dx.doi.org/10.1016/j.jneumeth.2015.01.030

1. Introduction

The electroencephalogram (EEG) as a multi-channel signal of neuronal brain activity can reflect brain states linked to the mental condition of a person. Due to its temporal resolution, it is an excellent and widely used technique for investigating human brain functioning. A major problem though is the contamination of the EEG signal by various physiological and non-biological artifacts such as eye movements, blinks, muscle activity, heartbeat, high electrode impedance, line noise, and interference from electric devices.

The discarding of entire EEG segments due to noise is a widely applied method in research settings and results in the loss of experimental data. This becomes especially problematic if only a few epochs are available and artifacts such as blinks or movements occur too frequently. Moreover, this approach is inappropriate when working with the continuous EEG and non-event-locked activity (Vanhatalo et al., 2004), long-range temporal correlations (Linkenkaer-Hansen et al., 2001), real-time brain–computer interface (BCI) applications, and online mental state monitoring (Jung et al., 2000). Other proposed methods for artifact rejection are based on regression in the time or frequency domain (Kenemans et al., 1991). They concentrate mainly on removing ocular artifacts (Jung et al., 2000), may themselves introduce new artifacts into the EEG recording (Weerts and Lang, 1973; Oster and Stern, 1980; Peters, 1967), and are unsuitable for real-time applications (Jung et al., 2000). A review of BCI-system artifact reduction techniques is given by Fatourechi et al. (2007).

A promising method which has established itself as an important part of EEG analysis is the application of independent component analysis (ICA) for data decomposition (Jung et al., 2000; Makeig et al., 1996) and separation of neuronal activity from artifacts (Fitzgibbon et al., 2007; Romero et al., 2008). The idea central to this method is that the EEG signal is a mixture of linearly independent source components (IC) that can be separated by ICA, visually examined, and classified as artifact or EEG signal components. Once the artifact components have been identified, they can be removed and the remaining EEG signal components can be projected back to the original time domain. This procedure yields the reconstruction of an artifact-free EEG.

The examination and classification of the ICs is time-consuming and requires the rater to have experience and knowledge for decision making. Viola et al. (2009) and Mognon et al. (2010) both worked on mitigating this difficulty by developing an EEGLAB plug-in for finding artifact ICs. The first relies on a user-defined template while the second is completely automatic but limited to four artifacts (Jung et al., 2000; Mognon et al., 2010).

However, EEG data contain an array of artifacts with unknown properties. Hence, there is a further need to develop fully automated machine classification, e.g. of the independent components, for automatic artifact elimination that can deal with all kinds of artifacts.

The idea underlying our method is inspired by the observation of experts during visual classification. For this process they visually inspect the 2D scalp map projections of the ICs, called topoplots. Based on the topoplots' image pattern and on the experts' know-how, they decide if the IC is an artifact or an EEG signal component. Computer vision aims at imitating the abilities of human vision by electronically perceiving and understanding an image for further decision making. Hence, we introduce in this paper a new method for artifact rejection based on machine classification of features derived from topoplot images by applying image processing algorithms for improving performance. Image features can be extracted through range filtering, local binary patterns (LBP) that are originally used as texture features, and geometric approaches like Gaussian curvature. These methods are commonly used in the context of 2D object recognition and include information from the entire image. Combined with a classification method such as linear discriminant analysis (Moghaddam et al., 2000), they can enhance the recognition performance of automated artifact elimination. This groundbreaking method does not depend on direct recording of artifact signals, neither it is limited to a specific number or type of artifact. An additional advantage of this novel method is its adaptability to an arbitrary number of EEG channel configurations. This is due to the nature of the topoplots, which are generated by interpolating the ICA mixing matrix columns onto a fixed-size grid (in our case: 51×63). This interpolation step enables the projection of different numbers of EEG channels and different EEG channel positions on the same grid, precisely by generating images of identical dimension. This permits the classification of images derived from ICs composed of varying column lengths of the mixing matrix, independent of not only the number but also the position of the channels utilized. Re-training the classifier for any further investigations using any other number of channels or any other channel positions is therefore redundant. This can be recognized by looking at Fig. 1. It clearly illustrates the pattern similarity of IC images per artifact despite their different EEG channel configurations.

2. Materials and experiments

The data acquisition took place in the shielded lab of the Federal Institute for Occupational Safety and Health in Berlin. The experiment was fully carried out with each subject in a single day. It

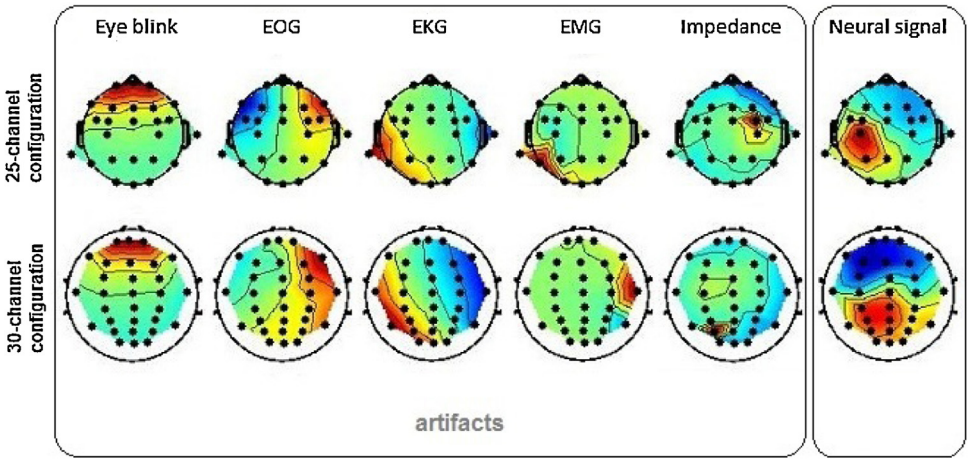


Fig. 1. Examples of similar IC artifact patterns despite the different EEG channel configurations. First row: IC images from a 25 channel configuration; second row: IC images from a 30 channel configuration.

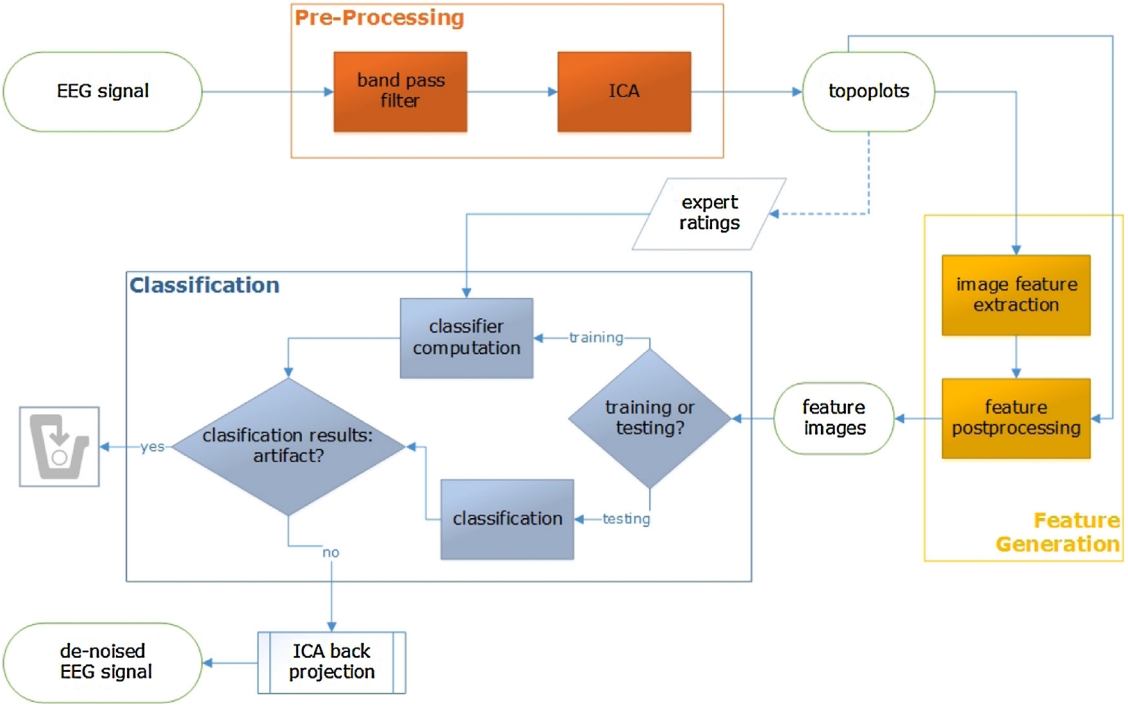


Fig. 2. Processing pipeline.

consisted of two components: a training phase and the main experiment. During the training phase subjects were familiarized with the cognitive tasks, which were identical to the main experiment but with shorter duration. An EEG signal was not recorded at this stage. The main experiment started after a short break subsequent to the training phase. The EEG was captured during the main experiment by 25 electrodes arranged according to the 10–20 system. Data were recorded with reference to Cz and at a sample rate of 500 Hz. For signal recording we used an amplifier from Brain-Products GmbH and their BrainRecorder software. The tasks were presented in a counterbalanced order and were controlled remotely by means of a remote desktop connection, an intercommunication system, and a video monitoring system.

The sample consists of 57 people in paid work and shows high variability with respect to cognitive capacity and age. Table 1 describes the sample set used.

The various cognitive tasks were realized through the implementation of a task battery in the E-Prime application suite (Psychology Software Tools, Pittsburgh, PA). The battery consists of nine tasks with diverse complexity and difficulty that reflect the following executive functions: shifting between tasks, updating and monitoring working memory representations, and inhibition of dominant responses. The implemented tasks are listed in Table 2 and explained in general in Wikipedia (2014a,b,c,d) and Unsworth et al. (2005). Furthermore, we recorded the EEG during a short period of relaxation at the beginning and at the end of the main experiment.

Table 1
Sample set.

Age	Female	Male	Total
30–39	7	7	14
40–49	12	13	25
50–59	9	2	11
60–62	3	4	7
Total	31	26	57

3. Methods

We implemented a MATLAB toolbox consisting of three main modules for pre-processing, feature generation, and classification. To train the classifier, we used visual ratings of the topoplots from two experts. The algorithm then automatically classified the ICs, based either on the new generated image features or the post-processed topoplots. Artifact components are discarded from subsequent processing. Components classified as EEG components are projected back, and in this way reconstruct an artifact-free EEG signal. Fig. 2 summarizes this processing pipeline and is described in-depth in the following.

3.1. Pre-processing

All recorded EEG signal data were imported into MATLAB (MathWorks, R2012b). In our case, the 25 recorded EEG channels were saved in one file per task and per person. The signals were multiplied with a Hamming window function and filtered with a band pass filter (order 100) between 0.5 and 40 Hz. Subsequently, independent component analysis (Infomax algorithm) was applied to the 25-channel EEG signal (Makeig et al., 1996; Delorme and Makeig, 2004):

$$W \cdot x = u \tag{1}$$

with x is the signal of scalp EEG channels, W the unmixing matrix and u the sources (ICs).

In that way, our multi-channel data were decomposed into 25 ICs. Each of the 25-element column vectors of the 25×25 ICA mixing matrix W^{-1} was interpolated onto a 51×63 grid using the inverse distance method of Sandwell (1987). They form the 25 two-dimensional scalp projection maps, here referred to as topoplots. The first plot on the left of Fig. 3(a) illustrates the interpolation on a fine, rectangular Cartesian grid. Examples of such images without the background are depicted in Fig. 5, where a set of topographic maps of the scalp data field are presented in a 2D circular view. The 25 computed ICs are characterized by their activations (the sources'

Table 2
Task battery reflecting executive functions.

Task	0-Back	2-Back	Sternberg	Serial Sternberg	Stroop
Duration (min)	5	5	10	10	5
Task	Switch PAR	Switch NUM	Switch XXX	AOSPAN	Relaxation (start, end)
Duration (min)	5	5	10	20	3 + 3

time course), their activity power spectrum calculated with FFT, and their 2D scalp component maps (interpolated from the columns of the mixing matrix).

All topoplots in combination with their power spectra were visually inspected and rated by two experts as either an artifact or EEG signal component (Fig. 5). Altogether, 15 625 topoplots were labeled, 46% as signal vs. 54% as artifact. Each of the experts rated half of the topoplots. These labels were used later for system training and classification.

Because of the time-consuming procedure and the large number of topoplots to be rated, it was impossible to find more willing experts for labeling. This fact illustrates once more the necessity of an automated solution. However, we compute the correlation of our experts' labeling on a sub-sample of 5500 topoplots that were rated by both. They achieved an agreement of 92.8%.

3.2. Feature generation

The feature generation module consists of algorithms for image feature extraction. Furthermore, it comprises post-processing algorithms that can be applied to the generated features but also directly to the topoplot images. Post-processing algorithms aim to reduce computational effort based on image downsampling and vectorization. In this way, they contribute to enhancing classification performance.

3.2.1. Image feature extraction

Image feature extraction generates features that can be generally categorized in three groups: texture features, gradient and range features, and geometric features.

3.2.1.1. Local binary pattern (LBP). LBP is a powerful feature for 2D texture classification first described in 1994 (Ojala et al., 1994) and refined later in Ojala et al. (1996). The LBP feature values are computed by building the differences between each pixel and its neighbors $P \in \{8, 16, 24\}$ within a prescribed radius R. If the center pixel is greater than its neighbor we note a 1, else a 0, so that the relation to the neighbors is binary coded. In the case of $P = 8$ neighbors, we obtain an eight-digit binary number that can be converted into a decimal number. This constitutes the new value of the central pixel.

A rotation-invariant alternative of the LBP method is described by Ojala et al. (2002). Here, the digits of the binary number are shifted until a minimal value is created. Another variant is the uniform LBP, where thresholding is used to replace rare feature values with a single value. A combination of both results in a rotation invariant, uniform LBP feature.

Huang et al. introduced the 3D LBP method, which computes four distinct LBP layers (Huang et al., 2006). Here coding is done not only for whether the central pixel is less or greater than its neighbor but also for the value of the differences themselves. In a four bit word the sign of the difference is tracked in the first bit and its value in the remaining three bits. In each LBP layer (sign layer, first bit layer, second bit layer, third bit layer) we set the corresponding decimal value at the position of the central pixel.

We implemented and tested all LBP variants described here on the topoplot images and presented the results in Section 4. Some examples of the LBP feature images are presented in Fig. 3(a) and (b).

3.2.1.2. Gradient images, range filter and Laplacian. We also evaluated the recognition performance of several gradient features. The horizontal Sobel operator (HSO) is obtained by spatial convolution with the filter mask

$$HSO = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad (2)$$

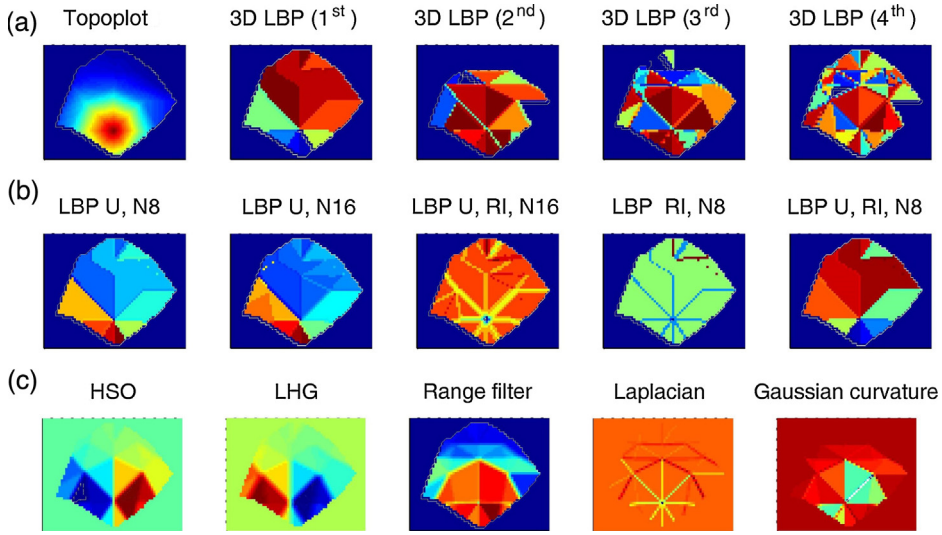


Fig. 3. Examples of feature images generated from the topoplot image (top left) using image processing algorithms introduced in the main text: (a, b) texture features and (c) gradient, range, and geometric features.

and detects vertical edges. To compute the horizontal derivative, typically a 2×1 kernel is applied that calculates the difference of two pixels. The large horizontal gradient operator (LHG) detects horizontal range differences over a greater horizontal distance. Heseltine et al. tested a variety of image processing techniques for face recognition and identified the LHG operator as the most effective surface representation and distance metric for use in application areas such as security, surveillance, and data compression (Heseltine et al., 2004). In consideration of this, we decided to use their recommended size of 5 and compute the LHG via the filter mask

$$LHG = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

We also tested the range filter, which takes the difference between the maximal and minimal value for each topoplot image's pixel in a 3×3 neighborhood (Bailey and Hodgson, 1985). In addition, we applied the Laplacian operator as a second order feature.

Fig. 3(c) illustrates examples for gradient images, range filtering and the Laplacian operator.

3.2.1.3. Gaussian curvature. The Gaussian curvature belongs to the geometric features and is defined as follows. The curvature K of a point on a surface is the product of the principal curvatures k_1 and k_2:

$$K = k_1 \cdot k_2 = \frac{1}{r_1} \cdot \frac{1}{r_2} \tag{4}$$

with main curvature radius r_1 and r_2.

We implemented the Gaussian curvature of the topoplot surface points as a further feature and tested its classification accuracy (Horn, 1984). An example is shown in Fig. 3(c).

3.2.2. Post-processing

Each feature image consists of $51 \times 63 = 3213$ pixels. We cropped the inner area containing the scalp map and discarded the frame, thus obtaining 1489 pixels (approx. 46% of the original feature image). We continued downsampling the image by selecting either each second pixel in horizontal and vertical direction (4:1), each third (9:1), or each fourth (16:1) pixel. In this way, we generated feature images with 372 pixels for the 4:1 alternative (approx. 11.58% of the original feature image), 164 pixels for the 9:1 alternative (approx. 5.10% of the original feature image) and 94 pixels for the 16:1 alternative (approx. 2.93% of the original feature image). The main purpose of downsampling was to investigate how much pixel information is still needed for achieving reasonably good classification results and to possibly reduce computational effort for the classification. For the sake of convenience we transformed the image matrix in a feature vector for further classification.

3.3. Classification

We used linear discriminant analysis (LDA) as an easily comprehensible and reproducible method for classification. LDA, a widely used method for face recognition (Kong et al., 2015), image retrieval (Swets and Weng, 1996), text classification (Ye et al., 2004), etc. was applied to all computed features. LDA minimizes the within-class variance S_W while maximizing the between-class variance S_B by computation of the projection matrix W_{LDA}.

$$S_W = \sum_{i=1}^{C} \sum_{n=1}^{l_i} (x_{in} - m_i) \cdot (x_{in} - m_i)^T \tag{5}$$

$$S_B = \sum_{i=1}^{C} l_i \cdot (m_i - m) \cdot (m_i - m)^T \tag{6}$$

with C is the number of classes, m_i the mean of class i and l_i the number of samples in class i. Hence the ratio of S_B to S_W has to be maximized for maximal separability of the classes:

$$W_{LDA} = \mathrm{argmax}_W \frac{\left| W^T \cdot S_B \cdot W \right|}{\left| W^T \cdot S_W \cdot W \right|} \tag{7}$$

In contrast to principal components analysis (PCA), LDA aims to maximize the class-discriminatory information. Hence, it provides class separability and a decision boundary between the classes by computing the axes' directions that maximizing the separation between them. Euclidean distance is used to classify the data.

A well-known challenge of LDA for many applications is that it requires the within-class covariance matrix to be nonsingular, which is not always given. In order to overcome the possible singularity of S_W, several approaches have been proposed (Swets and Weng, 1996; Dai and Yuen, 2003; Kong et al., 2015) including computation of the pseudo-inverse LDA (Raudys and Duin, 1998; Skurichina and Duin, 1996). The latter approach was used for this work whenever the computation of the inverse failed.

Altogether, the data consist of 625 sets (subjects $\times$ tasks) with 25 topoplots in each set. We randomly selected 60% of the sets for the training and the remaining 40% for testing. Hence, for each feature generation method applied we used 9375 image feature vectors (375 sets) for training the classifier and 6250 image feature vectors (250 sets) for testing it. In addition to the features generated, we also trained and classified the topoplots themselves.

4. Results

We evaluated our artifact removal algorithm by means of a close inspection of the classification results, the signal-to-noise ratio and the performance time. While the first two are standard for the assessment of performance, the performance timing accounts for the real-time feasibility of the algorithm.

4.1. Classification results

Determination of the minimum number of eigenvectors was done for each feature and sampling rate. The main purpose was to reduce computational effort. The identification of this minimum was done empirically using the receiver operating characteristic (ROC) curve (Fig. 4) which illustrates quite well the performance of a binary classifier system as its discrimination threshold is varied. Hence, it provides a way to select possibly optimal models and to discard suboptimal ones. Based on our ROC analysis of diagnostic decision making, a cut-off of 70% of the total number of eigenvectors, sorted in descending order by the size of their eigenvalues, was set as threshold. The entire classification procedure with random selection, training and testing was cross validated by executing it 50 times for each feature. Results were averaged and the variance was calculated (Table 3).

To further validate our results, we compared the recognition rates from the LDA with 70% of the total number of eigenvectors (Table 3) with the classification results from an LDA classifier with all eigenvectors. The extremely small differences (in the range of 0–0.7%) in the recognition rates between the two systems suggest the use of the 70%-classifier in favor of a minimal computing time, which is always advantageous when dealing with the development of real-time systems.

The recognition rates for feature images vary between 73% for the fourth layer of 3D LBP and 88% for range filter. Downsampling of the feature images does not have a large impact on most of the features' accuracy rates. Looking at Table 3 we notice as expected that recognition rates decline differently. They decline least for the topoplots, the range filter and the Gaussian curvature. They

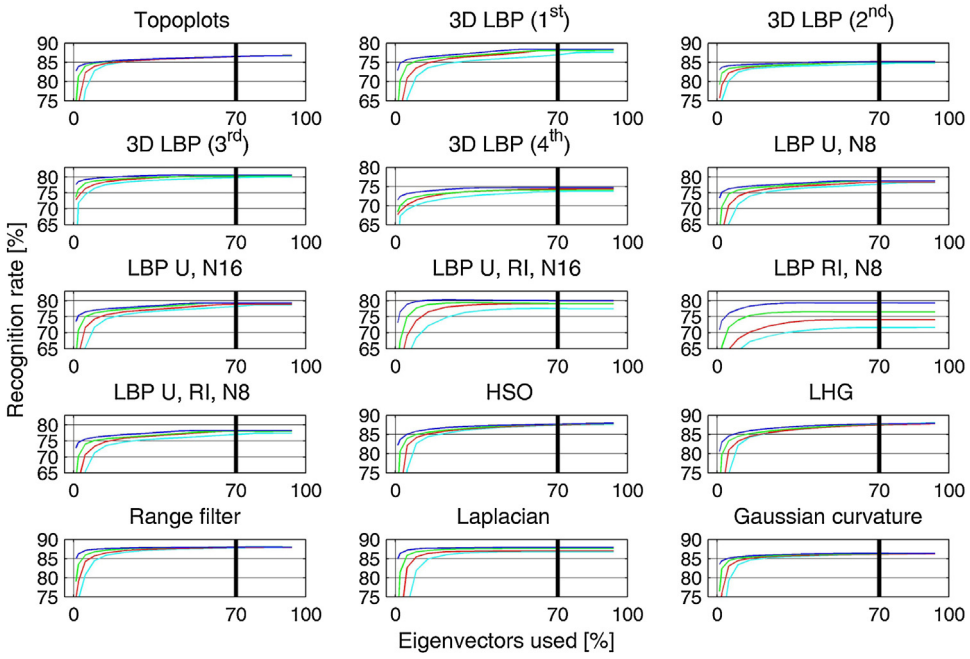


Fig. 4. Dependence of recognition rates on amount of eigenvectors used for each of the 15 features (subplots) and the four sampling rates (blue 1:1, green 4:1, red 9:1, cyan 16:1); the vertical line indicates the cut-off by 70% of total number of eigenvectors (eigenvectors are sorted by the size of their eigenvalues).

decline most for the uniform, rotation invariant LBP computed for 16 neighbors and for the rotation invariant LBP computed for eight neighbors, both suffering from downsampling on the order of 4–8%. This appears reasonable when looking at Fig. 3(b).

Additionally, by applying dimensionality reduction, we not only tried to reduce the computational costs of classification but also tested for over-fitting. In general, dimensionality reduction can help minimize the error in parameter estimation due to redundant information (the curse of dimensionality, also known as the Hughes effect in machine learning).

For improving classification we also mixed different kinds of features (e.g. 3D LBP (1st/2nd) and Laplacian, Laplacient and Gaussian curvature, range filter and gradients, etc.) but the results did not outperform the 88% recognition rate of range filter alone.

On a final note, we considered benchmarking our algorithm with an existing removal method implemented in EEGLAB. Two of the best-known EEGLAB implementations for artifact removal are ADJUST (Mognon et al., 2010) and CORRMAP (Viola et al., 2009). ADJUST distinguishes between stereotyped artifacts (e.g. ocular eye movements, blinks, heart beats) and non-stereotyped artifacts (e.g. movements, external sources). The authors note in their tutorial (Buiatti and Mognon, 2014) that ADJUST aims to automatically identify and remove components of the stereotyped artifacts, while the non-stereotyped artifacts are problematic: "ADJUST does not attempt to remove these artifacts and relies on a suitable pre-processing for removing them before the ICA decomposition" (Buiatti and Mognon, 2014). They advise the users to "remove from the continuous data all segments containing paroxysmal artifacts" (Buiatti and Mognon, 2014), after filtering. Due to this pre-processing step of segments' removal we deemed ADJUST as inappropriate for benchmarking. We then switched over to CORRMAP, ran it in automatic mode, and manually

Table 3
Recognition rates of features tested: mean (%) over 50 cross-validated results and variance (in parentheses).

Sampling	1:1	4:1	9:1	16:1
Eigenvectors	1042	260	114	65
Features				
Topoplots	86.6 (0.16)	86.6 (0.17)	86.5 (0.17)	86.7 (0.19)
3D LBP (1st)	78.4 (0.20)	78.0 (0.16)	78.1 (0.16)	77.0 (0.25)
3D LBP (2nd)	85.1 (0.14)	84.9 (0.16)	85.3 (0.17)	84.6 (0.16)
3D LBP (3rd)	80.3 (0.12)	80.1 (0.11)	80.4 (0.12)	79.8 (0.11)
3D LBP (4th)	74.5 (0.21)	74.0 (0.18)	74.2 (0.22)	73.5 (0.22)
LBP rotation invariant, 8 neighbors	79.3 (0.25)	76.4 (0.23)	74.0 (0.14)	71.5 (0.28)
LBP uniform, 8 neighbors	78.6 (0.16)	78.7 (0.18)	78.4 (0.15)	77.8 (0.17)
LBP uniform, 16 neighbors	79.3 (0.21)	79.4 (0.22)	79.0 (0.19)	78.3 (0.25)
LBP uniform, rotation invariant, 8 neighbors	78.3 (0.20)	78.0 (0.22)	78.0 (0.20)	77.1 (0.22)
LBP uniform, rotation invariant, 16 neighbors	80.1 (0.15)	79.1 (0.14)	79.1 (0.19)	77.6 (0.27)
LHG	87.9 (0.10)	87.7 (0.11)	87.5 (0.14)	87.6 (0.09)
HSO	87.8 (0.11)	87.7 (0.13)	87.5 (0.10)	87.4 (0.12)
Range filter	**88.0** (0.10)	**88.1** (0.09)	**88.0** (0.11)	87.8 (0.11)
Laplacian	87.9 (0.11)	87.6 (0.10)	87.0 (0.12)	86.8 (0.14)
Gaussian curvature	86.3 (0.13)	86.2 (0.18)	86.0 (0.16)	86.1 (0.15)

Recognition rates greater than 88% in bold.

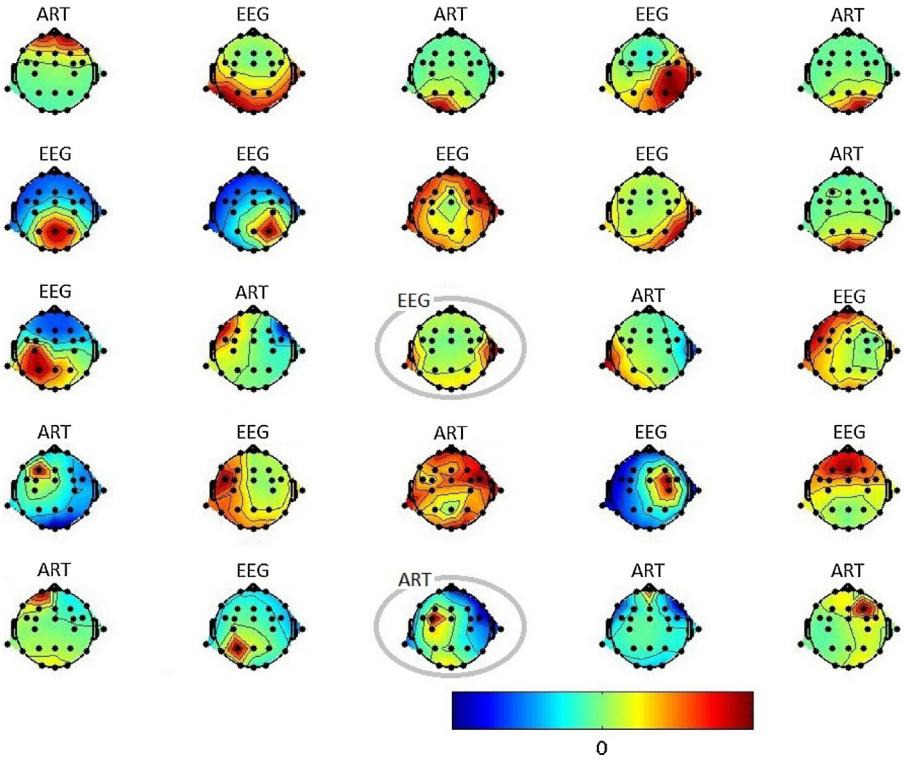


Fig. 5. Example of a complete IC set of 25 topoplots with one expert's labeling printed on the top of each topoplot. Diverging machine ratings are circled in gray.

provided artifact templates as requested by the system. We tested 250 ICs and achieved a 62% agreement with the experts' ratings. The CORRMAP results are not surprising due to CORRMAP's focus on eye blinks and lateral eye movements, as mentioned by the authors (Viola et al., 2009). CORRMAP similar to ADJUST is engineered to capture in particular stereotyped artifacts like blinks, eye movements and/or heart beats. Their acknowledged lack of ability to capture non-stereotyped artifacts is a challenge that we accept.

4.2. Signal-to-noise ratio

After automated classification, the components classified as artifact were discarded from the subsequent process. The remaining ones, classified as signal components, were back projected. The result is the machine de-noised signal. A representative complete IC set of 25 topoplots is shown in Fig. 5. The expert's labeling is printed above each topoplot. Diverging machine ratings are circled in gray, e.g. the image in the third row, third column was classified by the machine as artifact. For the sake of completeness, we should mention that our experts used the activity power spectrum as additional information for the classification procedure. An example of the original band pass filtered signal of electrode Fp1, the expert de-noised signal, and the signal de-noised automatically using range filtering (sampling 9:1) is shown for one subject in Fig. 6.

We used the signal-to-noise ratio (SNR) as a criterion to characterize the quality and similarity of the artifact rejection procedure when comparing the experts' visual classification with the image pattern approach. SNR values were computed on the basis of the definition in Strutz (2000).

$$SNR = 10 \cdot \log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \text{(dB)} \tag{8}$$

with σ_x^2 is the variance of the signal and σ_e^2 the variance of the noise. For zero mean signals as is the case here, this results in

$$SNR = 10 \cdot \log_{10}\frac{\sum_{(i=1)}^{N} x_i^2}{\sum_{(i=1)}^{N}(s_i - x_i)^2} \tag{9}$$

with N is the number of sample points, x_i the noise reduced signal at time i, and s_i the band pass filtered signal at time i.

For the noise signal in the denominator, we employed the same value for the expert de-noised SNR_{EXP} and the machine de-noised SNR_M. Hence, we used as residual noise for both SNR calculations the difference between band pass filtered signal and experts' noise reduced signal. Under this assumption the performance of both approaches can be compared. The signals of each task were appended one after another, resulting in one long signal for each person, and the SNR_{EXP} and SNR_M were computed for each electrode. Results were averaged over the subjects.

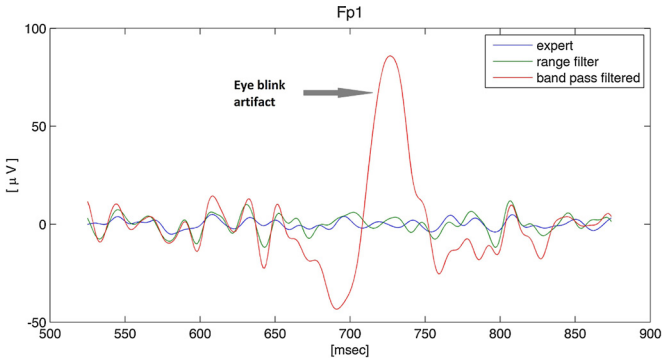


Fig. 6. Band pass filtered signal of electrode Fp1 (red) with a pronounced eye-blink artifact. Expert de-noised signal (blue) and automatically de-noised signal (green) by range filtering (sampling 9:1) without the artifact.

Table 4
Performance times for the processing pipeline.

Pre-processing		Band pass filter	0.164 s
		ICA	4.731 s
Feature generation		Topoplot	0.121 s
		Range filter	0.002 s
		Feature postprocessing	0.00007 s
Classification	Training (offline)	1:1 (worst case)	1.7 s
		16:1 (best case)	0.025 s
	Testing (online)	1:1 (worst case)	0.03 s
		16:1 (best case)	0.002 s

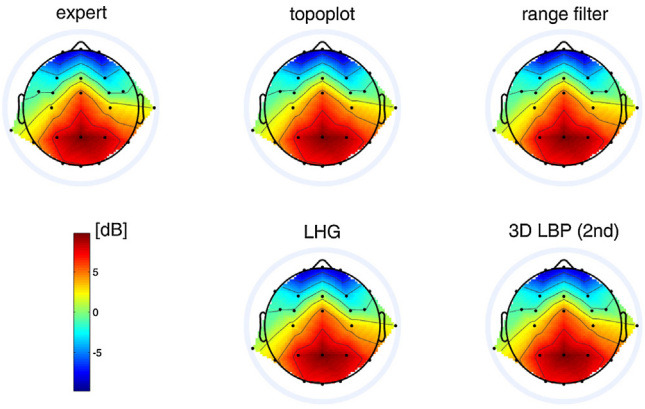


Fig. 7. SNR maps (group average) for expert de-noising and for the machine based de-noising of four selected features (sampling 9:1).

The SNR range is very low, from −9.8 to 9.8 dB, well within the range found in the literature. SNRs are, as expected, lower in the frontal areas (Goldenholz et al., 2009; Mishra and Singla, 2013) that are contaminated by eye artifacts. The difference between SNR_{EXP} and SNR_M is almost zero indicating that we achieved similar quality between expert based and machine based de-noising. SNR scalp maps for experts' noise reduced signal and for de-noising of four selected image features are shown in Fig. 7. They illustrate the similarity of the time-consuming, manual classification for artifact rejection and the automatic, machine de-noising for all EEG channels averaged over subjects and tasks.

4.3. Performance time

We tested our system on an Intel Core i5-3320M Processor (2.6 GHz) with 8 GB DDR3-SDRAM (2 × 4 GB). As minimum input signal for performance time testing of the algorithm we chose a short 25-channel-EEG signal of 94.34 s (500 Hz sample rate). The selection was made with reference to the fact that the ICA computation forms the bottleneck for the system's timing performance and depends on the signal length. According the EEGLAB Wiki tutorial (Delorme and Makeig, 2014) finding N stable components requires more than kN^2 data sample points (N denotes the number of EEG channels, N^2 is the number of weights in the ICA unmixing matrix, and k is a multiplier that increases as the number of channels increases). Hence, in our case the pre-processing module needs approx. 5 s for filtering and performing ICA, while the generation of one feature takes only 2.7 ms. System training can be done offline and actually does not need to be considered for the evaluation of real-time performance. However, for the 9375 image feature vectors used, the computation time was quite small and varied between 1.7 s (1:1 image, worst case) and 0.025 s for the downsampled image (16:1 image, best case). Classification testing of one image feature vector needs in best case only 0.002 s and in worst case 0.03 s. Table 4 outlines the performance times and demonstrates that the system is real-time capable.

As a side note, we call attention to hardware-level programming, which can enhance timing performance of the pre-processing module, in particular by consideration of parallelism inherent in hardware design.

5. Conclusions

In this paper we described a novel approach for robust, automated EEG artifact rejection inspired by computer vision. It is based on machine classification of ICs as artifact or EEG signal via features gained with image processing algorithms. We implemented, tested, and validated several image features, e.g. range filter, local binary patterns, and geometric features (Fig. 3). Linear discriminant analysis was applied for classification. Accuracy rates as well as signal-to-noise ratios were calculated for assessing the features' achievement potential.

Existing methods rely either on a user-defined template or are limited to a few artifacts (Viola et al., 2009; Mognon et al., 2010; Jung et al., 2000). Practitioners on the other hand well know that real-world EEG data contain an array of artifacts with unknown properties. Our goal therefore was to develop an automated method capable of distinguishing between the pure EEG signal and all types of artifacts. One could suggest a simpler approach that uses just the columns of the mixing matrix instead of their interpolation onto a fixed-size grid (topoplots). To justify our method, we conducted classification tests utilizing only the mixing matrix columns and received recognition rates of 87.7%. Indeed, the result does not outperform the 88% recognition rates of range filter but supports our idea of using the interpolated images, which inherently utilize the columns of the mixing matrix to train our classifier. The major difference between both procedures is that our approach is universal. This means: it is independent of the number of used EEG channels, it performs for any EEG channel positions, and it does not need to be retrained anymore. In this paper we want to highlight all these additional advantages arising from our method's adaptability to an arbitrary number of EEG channel configurations that constitute its universality. It is worth pointing out that experts generally proceed during visual classification of the topoplots without taking into account the number of EEG channels used. This fact supports our approach of using the 2D scalp map projections of the ICs instead of the mixing matrix columns. In short, the spatial pattern of the image contains information crucial for the experts' decisions.

Currently, we are working on evaluating our method by testing it with EEG data from another lab. This EEG data has a different EEG channel setup, it is conducted with another EEG registration system, and its experimental setting and tasks differ from ours as well. Nevertheless, the classifier trained with our data will not in any way be retrained on the new EEG data. Specifically, this means

that neither the large number of topoplots nor the time-consuming procedure for expert ratings is required for the purposes of re-training the classifier. The results will be presented in a future paper.

To conclude, we achieved robust, automatic, and real-time recognition performance with a best recognition rate of 88% for range filtering, which is very good compared with the currently existing methods. Furthermore, the proposed method is not limited to specific types of artifacts and can be applied online over different experiments and subjects. It is known that manual classification by experts can slightly differ between each other (Winkler et al., 2011). Hence, the differences of the SNRs between expert and machine de-noising are extraordinarily good. However, for more accurate classification results an average rating between more than two experts should be considered for successful system training. To reduce computational effort downsampling can be applied without loss of classification accuracy.

In summary, the presented method is an automatic, reliable, real-time capable, and practical tool that reduces the time-intensive manual selection of ICs for artifact removal. The results are very promising despite the relatively small channel resolution of 25 electrodes. We expect an improvement of the algorithm by combining the image features with frequency information and applying non-linear classifiers (e.g. support vector machine). This will be done in future work and presented in a separate paper.

Acknowledgments

We would like to thank Dr Sergei Schapkin, Dr Patrick Gajewski, and Prof Michael Falkenstein for selection of the battery's tasks. We would like to thank Mr Ludger Blanke for technical support during the timing tests for the tasks. In addition, we would like to thank Ms Xenija Weißbecker-Klaus, Mr Robert Sonnenberg, Dr Sergei Schapkin, and Ms Marion Exner for general task testing and for conducting the laboratory experiments. Furthermore, we would like to thank Ms Marion Exner for daily operational support and Dr Gabriele Freude for her valuable suggestions.

More information about the project where our EEG data were acquired can be found under http://www.baua.de/de/Forschung/Forschungsprojekte/f2312.html?nn=2799254.

References

Bailey D, Hodgson R. Range filters: local intensity subrange filters and their properties. Image Vis Comput 1985;3(3):99–110, http://dx.doi.org/10.1016/0262-8856(85)90058-7, URL http://www.sciencedirect.com/science/article/pii/0262885685900587

Buiatti M, Mognon A. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features – a tutorial; 2014, URL http://www.unicog.org/pm/uploads/MEG/TutorialADJUST1_1.pdf [accessed 19.09.14].

Dai D, Yuen P. Regularized discriminant analysis and its application to face recognition. Pattern Recognit 2003;36(3):845–7.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. J Neurosci Methods 2004;134:9–21.

Delorme A, Makeig S. EEGLAB wiki; July 2014, URL http://sccn.ucsd.edu/wiki/EEGLAB#The_EEGLAB_Tutorial_Outline [accessed 10.09.14].

Fatourechi M, Bashashati A, Ward R, Birch G. EMG and EOG artifacts in brain computer interface systems: a survey. Clin Neurophysiol 2007;118(3):480–94.

Fitzgibbon S, Powers D, Pope K, Clark C. Removal of EEG noise and artifact using blind source separation. J Clin Neurophysiol 2007;24(3):232–43, http://dx.doi.org/10.1097/WNP.0b013e3180556926.

Goldenholz D, Ahlfors S, Hämälä inen M, Sharon D, Ishitobi M, Vaina L, et al. Mapping the signal-to-noise ratios of cortical sources in magnetoencephalography and electroencephalography. Hum Brain Mapp 2009;30(4):1077–86, http://dx.doi.org/10.1002/hbm.20571.

Heseltine T, Pears N, Austin J. Three-dimensional face recognition: a fishersurface approach. In: Campilho A, Kamel M, editors. Image analysis and recognition, vol. 3212 of lecture notes in computer science. Springer Berlin Heidelberg; 2004. p. 684–91, http://dx.doi.org/10.1007/978-3-540-30126-4_83 [accessed 1.09.14].

Horn B. Extended Gaussian images. Proc IEEE 1984;72(12):1671–86, http://dx.doi.org/10.1109/PROC.1984.13073.

Huang Y, Wang Y, Tan T. Combining statistics of geometrical and correlative features for 3D face recognition. In: Proceedings of British machine vision conference; 2006. p. 879–88.

Jung T, Makeig S, Humphries C, Lee T, Mckeown M, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 2000;37(2):163–78.

Kenemans J, Molenaar P, Verbaten M, Slangen J. Removal of the ocular artifact from the EEG: a comparison of time and frequency domain methods with simulated and real data. Psychophysiology 1991;28(1):114–21, http://dx.doi.org/10.1111/j.1469-8986.1991.tb03397.x.

Kong H, Li X, Wang J-G, Teoh EK, Kambhamettu C. Discriminant low-dimensional subspace analysis for face recognition with small number of training samples. In: Clocksin WF, Fitzgibbon AW, Torr PHS, editors. BMVC. British Machine Vision Association; 2005, ISBN 1-901725-29-4, http://www.bibsonomy.org/bibtex/2ed5a880a804cbb1335a14747db4609cf/dblp, http://www.bmva.org/bmvc/2005/papers/paper-57-165.html, http://dblp.uni-trier.de/db/conf/bmvc/bmvc2005.html#KongLWTK05

Linkenkaer-Hansen K, Nikouline V, Palva J, Ilmoniemi R. Long-range temporal correlations and scaling behavior in human brain oscillations. J Neurosci 2001;21(4):1370–7, arXiv: http://www.jneurosci.org/content/21/4/1370.full.pdf+html. URL http://www.jneurosci.org/content/21/4/1370.abstract

Makeig S, Bell A, Jung T, Sejnowski TJ. Independent component analysis of electroencephalographic data. In: Advances in neural information processing systems. MIT Press; 1996. p. 145–51.

Mishra P, Singla S. Artifact removal from biosignal using fixed point ICA algorithm for pre-processing in biometric recognition. Meas Sci Rev 2013;13:7–11, http://dx.doi.org/10.2478/msr-2013-0001.

Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. Pattern Recognit 2000;33:1771–82.

Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 2010:1–12.

Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit 1996;29(1):51–9.

Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24(7):971–87, http://dx.doi.org/10.1109/TPAMI.2002.1017623.

Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of the 12th IAPR international conference on pattern recognition, computer vision and image processing, vol. 1; 1994. p. 582–5, http://dx.doi.org/10.1109/ICPR.1994.576366.

Oster P, Stern J. Measurement of eye movement electrooculography. In: Matin I, Venables P, editors. Techniques in psychophysiology. Chichester, UK: Wiley; 1980. p. 275–309.

Peters J. Surface electrical fields generated by eye movements. Am J EEG Technol 1967;7(2):27–40.

Raudys S, Duin R. Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. Pattern Recognit Lett 1998;19(5–6):385–92, http://dx.doi.org/10.1016/S0167-8655(98)00016-6.

Romero S, Ma nanas M, Barbanoj M. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case. Comput Biol Med 2008;38(3):348–60, http://dx.doi.org/10.1016/j.compbiomed.2007.12.001.

Sandwell D. Biharmonic spline interpolation of GEOS-3 and SEAS, altimeter data. Geophys Res Lett 1987;14(2):139–42, http://dx.doi.org/10.1029/GL014i002p00139 [accessed 26.11.14].

Skurichina M, Duin RPW. Stabilizing classifiers for very small sample sizes. In: Proceedings of the 13th international conference on pattern recognition, vol. 2; 1996. p. 891–6, http://dx.doi.org/10.1109/ICPR.1996.547204.

Strutz T. Bilddatenkompression. Braunschweig: Vieweg; 2000.

Swets D, Weng J. Using discriminant eigenfeatures for image retrieval. IEEE Trans Pattern Anal Mach Intell 1996;18(8):831–6, http://dx.doi.org/10.1109/34.531802.

Unsworth N, Heitz R, Schrock J, Engle R. An automated version of the operation span task. Behav Res Methods 2005;37(3):498–505, http://dx.doi.org/10.3758/BF03192720, URL http://webs.wofford.edu/boppkl/coursefiles/Thesis/articles/UnsworthHeitz_2005_autoospan.pdf [accessed 02.04.14].

Vanhatalo S, Palva J, Holmes M, Miller J, Voipio J, Kaila K. Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. Proc Natl Acad Sci U S A 2004;101(14):5053–7.

Viola F, Thorne J, Edmonds B, Schneider T, Eichele T, Debener S. Semi-automatic identification of independent components representing EEG artifact. Clin Neurophysiol 2009;120(5):868–77, http://dx.doi.org/10.1016/j.clinph.2009.01.015.

Weerts T, Lang P. The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). Biol Psychol 1973;1(1):1–19, http://dx.doi.org/10.1016/0301-0511(73)90010-0, URL http://www.sciencedirect.com/science/article/pii/0301051173900100

Wikipedia. n-Back; August 2014, URL http://en.wikipedia.org/wiki/N-back [accessed 01.09.14.

Wikipedia. Task switching (psychology); July 2014, URL http://en.wikipedia.org/wiki/Task_switching_(psychology) [accessed 01.09.14].
Wikipedia. Stroop effect; August 2014, URL http://en.wikipedia.org/wiki/Stroop_effect [accessed 01.09.14].
Wikipedia. Saul sternberg; April 2014, URL http://en.wikipedia.org/wiki/Saul_Sternberg [accessed 01.09.14].
Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav Brain Funct 2011., http://dx.doi.org/10.1186/1744-9081-7-30.
Ye J, Janardan R, Park C, Park H. An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Trans Pattern Anal Mach Intell 2004;26:982–94.