PAPER • **OPEN ACCESS**

Robust artifactual independent component classification for BCI practitioners

To cite this article: Irene Winkler *et al* 2014 *J. Neural Eng.* **11** 035013

View the article online for updates and enhancements.

Related content

- EEG artifact removal—state-of-the-art and guidelines
  Jose Antonio Urigüen and Begoña Garcia-Zapirain

- A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update
  F Lotte, L Bougrain, A Cichocki et al.

- EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces
  Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich et al.

Recent citations

- Exploring the effects of visual perspective on the ERP components of empathy for pain
  Carl Michael Galang *et al*

- Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings
  Chi-Yuan Chang *et al*

- Automated artifact rejection algorithms harm P3 Speller brain-computer interface performance
  David E. Thompson *et al*

Robust artifactual independent component classification for BCI practitioners

Irene Winkler[1], **Stephanie Brandl**[2], **Franziska Horn**[1], **Eric Waldburger**[1], **Carsten Allefeld**[2] **and Michael Tangermann**[3]

[1] Machine Learning Laboratory, Technical University of Berlin, Marchstr. 23, D-10587 Berlin, Germany
[2] Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin Berlin, Philippstr. 13, Haus 6, D-10115 Berlin, Germany
[3] BrainLinks-BrainTools Excellence Cluster, University of Freiburg, Albertstr. 23, D-79104 Freiburg, Germany

E-mail: irene.winkler@tu-berlin.de

Abstract

Objective. EEG artifacts of non-neural origin can be separated from neural signals by independent component analysis (ICA). It is unclear (1) how robustly recently proposed artifact classifiers transfer to novel users, novel paradigms or changed electrode setups, and (2) how artifact cleaning by a machine learning classifier impacts the performance of brain–computer interfaces (BCIs). *Approach.* Addressing (1), the robustness of different strategies with respect to the transfer between paradigms and electrode setups of a recently proposed classifier is investigated on offline data from 35 users and 3 EEG paradigms, which contain 6303 expert-labeled components from two ICA and preprocessing variants. Addressing (2), the effect of artifact removal on single-trial BCI classification is estimated on BCI trials from 101 users and 3 paradigms. *Main results.* We show that (1) the proposed artifact classifier generalizes to completely different EEG paradigms. To obtain similar results under massively reduced electrode setups, a proposed novel strategy improves artifact classification. Addressing (2), ICA artifact cleaning has little influence on average BCI performance when analyzed by state-of-the-art BCI methods. When slow motor-related features are exploited, performance varies strongly between individuals, as artifacts may obstruct relevant neural activity or are inadvertently used for BCI control. *Significance.* Robustness of the proposed strategies can be reproduced by EEG practitioners as the method is made available as an EEGLAB plug-in.

Keywords: EEG, artifact removal, independent component analysis (ICA), blind source separation (BSS), brain–computer interface (BCI)

(Some figures may appear in colour only in the online journal)

1. Introduction

Artifacts are omnipresent in recordings of the electroencephalogram (EEG) and other brain signals. For neuroscientific or clinical purposes the interpretation of EEG signals depends on relatively clean recordings. Thus, artifact avoidance during measurement and post-hoc artifact removal are important steps to enhance the signal-to-noise ratio (SNR) before scientific interpretation of the data. While task-independent artifacts may mask an existing effect, artifacts systematically locked to an experimental task are even more problematic: they may lead to misinterpretation of the data and spurious results.

The field of the brain–computer interface (BCI) not only makes use of offline analyses, but strives to interpret mental states on a single-trial basis in real-time and in closed-loop scenarios [1]. BCI research is especially sensitive to

task-locked artifacts, as the decoding of a user's intent by a BCI system should not rely on task-related non-neural signals. This requirement is most important when conducting research with healthy study participants on a novel paradigm or analysis method which should be transferable to severely motor-impaired patients, because they may not be physically capable of producing those artifacts [2–4]. Understandably, the role of artifacts is thus scrutinized during peer-reviewed publication processes.

The exclusive use of brain signals in BCI must typically be dropped when it comes to practical tests with end-users in need, as hybrid BCI approaches [5, 6] provide a richer and more reliable control than pure BCIs. Additionally, interest in novel types of studies is growing amongst EEG researchers. Such studies include users (inter-)acting in space [7–9] like in collaborative and social paradigms (for a review see [10]), the interaction between users and machines [11] and the non-medical use of BCI methods [12, 13].

From an EEG practitioner's point of view, a fully automatic algorithmic solution for the treatment of artifacts is desirable. It would put him or her in control of artifacts and enable him or her to either remove them or check their influence. Ideally, this would be realized by a global classifier which could be trained once and then reliably separates multiple types of artifactual components from neural components. The classifier should work robustly across data from different users and across domains. The latter includes changing experimental paradigms and tasks, different preprocessing methods and varying EEG electrode setups. It should do without any need of re-training, and it should not require separate artifact recordings before it can be applied to novel scenarios.

1.1. State-of-the-art IC artifact classification

For an extensive review of artifact reduction techniques in the context of BCI-systems, we refer the reader to [14]. In our work, we concentrate on a class of popular artifact rejection approaches, which decompose the original EEG into independent source components (ICs) using independent component analysis (ICA). This method exploits the assumption that artifactual signal components and neural activity are generated independently. Artifactual ICs are hand-selected and then discarded. The remaining neural components are used to reconstruct the EEG [15, 16].

While assumptions for the application of ICA methods are only approximately met in practice (no systematic co-activation of artifactual and neural activity, linear mixture of independent components (ICs), stationarity of the sources and the mixture, prior knowledge about the number of components), their application usually leads to a good, albeit not perfect separation for common artifacts such as blinks, eye movements or scalp muscles [17–20]. ICA has successfully been applied to the removal of cochlea implant artifacts [21]. However, gait-related artifacts are reported to remain in most of the ICs in EEG recorded during mobile activities [9, 22].

Because a thorough analysis of the achievable separation performance is out of the scope of this paper, we refer the reader to [17, 23, 24] on the question of which ICA variants are well-suited for artifact rejection. Instead, we focus on practical tools which avoid the time-consuming hand-rating process of ICs by classifying ICs with the help of machine learning methods into artifactual and non-artifactual components. Most approaches concentrate on eye artifacts [25–31], but automatic classification has also been successful for heart-beat artifacts [28, 31], generic discontinuities [29], muscle artifacts [31–34] and even very specialized artifacts such as cochlear implants [21]. As most of these methods have a supervised basis, to some degree they reflect the specific conditions of the training set. The EEG practitioner is now faced with the question of how well supervised methods generalize to his or her data acquired under novel experimental conditions with different preprocessing.

Unsupervised methods successfully circumvent this problem for example by reverting to automatic thresholding strategies [29]. However, these methods are often limited to the use of one or two features and detect only certain types of artifacts. It is unclear how to extend them to more complex artifacts with a varying physiological fingerprint, such as muscle artifacts. For supervised or template-based approaches, first studies suggest that generalization to novel paradigms is possible [28, 30, 31, 34]; however, efforts have concentrated on eye artifacts [28, 30].

1.2. Robustness under novel paradigms and electrode setups

In this paper, we take a step forward by analyzing the generalization ability of a state-of-the-art supervised IC classification algorithm which we have recently proposed [34]. It is not restricted to the classification of eye or muscle artifacts, but is equally well suited to detect other artifacts such as loose electrodes. By comparing three strategies, we investigate this multi-artifact classifier wrt. new electrode setups and paradigms. We ask the following questions: How does a change of the electrode setup impact the IC classification performance? Is it necessary to hand-label components of the new data set and retrain the classifier based on those? How strong is the deterioration of IC classification performance without re-training? We investigate these questions for three data sets of 6303 labeled ICs from 35 participants in 3 experimental studies: a reaction time (RT) task embedded in a simulated-driving task, an auditory event-related potential study (ERP-BCI) and a study analyzing continuous EEG data (CNT) of subjects instructed to listen to short stories.

1.3. Effect on BCI performance

After having demonstrated the robustness properties of the IC classification, we are interested in the effects of automatic ICA artifact cleaning on the classification of EEG trials in BCI systems. As a first proof-of-concept, Halder *et al* [33] applied artifact cleaning to data from three participants who performed motor imagery. Depending on whether artifacts were systematically co-activated with the task or not, opposite effects of artifact cleaning on BCI classification performance were demonstrated. To the best of our knowledge, only small

data sets of one or two participants have been analyzed since then [35, 36].

To fill this gap, we extend our analysis from [34] by investigating the overall effect of ICA artifact cleaning on BCI performance to data of 101 participants wrt. 3 BCI paradigms: auditory event-related potentials, event-related (de-)synchronization and slow motor-related potentials due to motor imagery tasks.

1.4. Software for the EEG practitioner

Last but not least, we make our IC classification software available as an EEGLAB plug-in 'MARA' (Multiple Artifact Rejection Algorithm). EEGLAB [37] is a popular, Matlab-based open-source tool and used by a growing community of EEG researchers. As existing ICA-based plug-ins primarily focus on the detection of eye artifacts [27–29], we hope this will deliver a substantial contribution to the community by assisting EEG practitioners with the rejection of multiple type of artifacts.

2. Methods and materials

2.1. Processing chain for ICA artifact rejection

The typical process chain for artifact rejection with ICA consists of the following steps: first, a rough pre-cleaning of the data by channel rejection and trial rejection based on variance criteria may be performed. Second, a dimensionality reduction may help to avoid an unnatural splitting of (neural) sources. Unfortunately, the optimal number of components to extract remains unknown and has to be determined either by visual inspection or by a heuristic, such as retaining 99% of the explained variance or a fixed number of components. Third, ICA methods decompose the observed EEG data x into unknown source components s assumed to be mutually independent and following the generative linear model $x = A \cdot s$. Finally, artifactual source components are identified which allows the EEG signals to be reconstructed without them.

In manual classification of ICs, experts ratings are based on a component's time series, its power spectrum and spatial pattern (given by the respective column of A). Unfortunately, ICA frequently results in mixed components containing aspects of both neural and artifactual activity which cannot be rated unambiguously [38]. Consequently, such mixed components tend to be either retained or rejected depending on the specific application. The subjective nature of such expert decisions is reflected by the fact that experts disagree with each other as well as with themselves over time [39]. Nevertheless, the reliability of component classification is often not reported, and if it is, researchers use one of many metrics of inter-rater reliability statistics which are difficult to compare directly (e.g. *Krippendorff's alpha* in [20], *inter-class correlation coefficient* in [40], *degree of association phi* in [28], *mean-squared error (MSE)* or *average agreement* in [34, 39]).

Automatic classification of ICs based on Machine Learning methods offers a well-described algorithm which

rates consistently over time. However, this algorithm, too, is of subjective nature in the sense that it is optimized to predict labels similar to those labeling strategies applied by human raters. The performance of the algorithm thus crucially depends on the quality of the training set and its labels. For all our IC data sets, experts were instructed to identify components which are predominantly driven by artifacts.

In this paper, automatic IC classification is realized by a linear pre-trained classifier. It is based on the following six features which were determined in a feature selection procedure described in [34]. One feature aims to detect outliers in the time series of an IC, three features are extracted from the spectrum, and two features extract information from the scalp pattern of an IC—the latter depending directly on the electrode layout.

(i) *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 the location of the sources can be estimated. We considered 2142 locations arranged in a 1 cm spaced 3D-grid, formulated the forward problem according to [41–43] and sought the source distribution with minimal l_2-norm (i.e. the 'simplest' solution) [44, 45]. Since this source distribution can model cerebral sources only, it is natural that artifactual signals originating outside the brain can only be modeled by rather complicated sources. Those are characterized by a large l_2-norm, which we use as a feature.

(ii) *Range within pattern.* The logarithm of the difference between the minimal and the maximal activation in a pattern.

(iii) *Mean local skewness.* The mean absolute local skewness of time intervals of 15 s duration. This feature aims to detect outliers in the time series.

(iv) λ *and fit error.* These two features describe the deviation of a component's spectrum from a prototypical $1/f$ curve and its shape. The parameters $k_1, \lambda, k_2 > 0$ of the curve

$$f \mapsto \frac{k_1}{f^\lambda} - k_2 \tag{1}$$

are determined by six points of the log spectrum: (1) the log power at 2 Hz, (2) the log power at 3 Hz, (3) the point of the local minimum in the band 5–13 Hz, (4) the point 1 Hz below the third point of support, (5) the point of the local minimum in the band 33–39 Hz, and (6) the point 1 Hz below the fifth point of support. Finally, the logarithm of λ and of the MSE of the approximation of f to the real spectrum in the 8–15 Hz range are used as features for the classifier.

(v) *8–13 Hz.* The average log band power of the α band (8–13 Hz).

2.2. Data sets and experimental paradigms

Data sets of four experimental EEG paradigms (named RT, CNT, MI-BCI, ERP-BCI) were available for this study. For three of them, RT, CNT and ERP-BCI, expert-labeled ICs

(artifacts versus neural sources) were available. Two data sets (MI-BCI, ERP-BCI) stem from BCI experiments. As the trial-wise BCI tasks are known, the estimated single-trial BCI-classification performance provides a metric for the influence of a preceding artifact treatment.

RT.   For this data set, labeled ICs were available. In a simulated-driving study, participants performed a forced-choice left or right key press RT task upon two auditory stimuli in an oddball paradigm [34]. EEG data was recorded from 121 approx. equidistant sensors and high-noise channels were rejected based on a variance criterion. We selected 43 runs of 10 min duration from eight participants that had 104 electrodes in common. Prior to the IC computation via TDSEP [46], a 2 Hz high-pass filter was applied, and dimensionality was reduced to 30 PCA components. Two experts hand-labeled the resulting 30 ICs per run into artifactual and neural components (1290 labeled ICs altogether).

 Of these, 840 ICs (28 runs from 5 participants) were used to train a linear classifier C_{RT} to discriminate artifactual from neural components. Another 450 ICs (15 runs from 3 remaining subjects) were available for estimating the generalization performance of C_{RT}. The training set contained 52% of artifactual ICs, the test set contained 59%.

CNT.   For this data set, labeled ICs were available. Nine participants continuously listened to audio–visual stories during short runs of an average duration of 3.77 min [40]. The resulting 71 recordings contained 62 EEG channels plus one EOG channel. The recording of each run was appended with a short eyes-closed and eyes-open recording and high-pass filtered at 0.16 Hz. No dimensionality reduction was applied, before ICs were estimated by FastICA [47] on the full set of electrodes. This decomposition yielded $63 \times 71 = 4473$ components, which were hand-rated by three experts into 47% artifactual and 53% neural source components.

ERP-BCI.   For this data set, labeled ICs as well as labeled BCI-trials were available. In a spatial auditory BCI study which made use of auditory event-related potentials, participants underwent a calibration run of approx. 30 min duration and an online spelling run [48]. In the online run, subjects were asked to write a sentence while auditory and visual feedback was provided. EEG was recorded from 61 electrodes while the participants listened to a rapid sequence of 6 auditory stimuli and were instructed to silently count the number of appearances of a rare target tone.

 For the classification of artifacts, data of 18 participants was analyzed. Their EEG signals were band-pass filtered between 0.1 and 40 Hz and the dimensionality was reduced to 30 PCA channels. Subsequently 30 ICs were computed per run using TDSEP. The resulting 540 source components were hand-labeled into 72% artifactual and 31% neural source components.

 To assess the influence of artifact correction onto the BCI classification performance, data of the 21 BCI novices participating in the first session of the auditory ERP speller study of Schreuder *et al* [48] was re-analyzed. Their calibration measurement is used to train a shrinkage regularized linear classifier based on spatio-temporal ERP features [48, 49]. BCI performance evaluations are based on the re-analyzed online data of these participants.

MI-BCI.   For this data set, labeled BCI-trials were available, but no labeled ICs. This data set was recorded with 119 EEG channels from 80 healthy BCI novices, who first performed motor imagery tasks (left hand, right hand and both feet) in a calibration run (i.e. without feedback). Every 8 s, the requested BCI task of the current trial was indicated by a visual cue. A CSP-based BCI-classifier (see below) was trained on the labeled calibration trials using the pair of classes which provided best discrimination. During the three online runs of 100 trials each participant controlled an application which provided continuous visual feedback in the form of a horizontally moving cursor [50].

 Motor imagery data can be exploited by two different types of EEG features.

 (i) CSP-MI-BCI: the most common strategy makes use of oscillatory features which describe event-related (de)-synchronization (ERD/ERS) in the alpha- and beta band of the EEG. After enhancing the SNR of these effects by individual data-driven spatial filters, which are derived by the common spatial patterns (CSP) analysis [51], CSP-features can be classified by a shrinkage-regularized linear classifier.
(ii) LRP-MI-BCI: the second strategy is based on slow motor-related potentials (e.g. the lateralized readiness potential (LRP)). Different classes of imagined movements are distinguished with an ERP-type analysis [49, 52]: EEG is band-pass filtered between 4 and 8 Hz, before a small number of class-discriminative intervals is determined on the calibration data. The average activity per interval and channel is used as features for a binary shrinkage-regularized linear classifier.

 While the original online runs were performed with the CSP-MI-BCI classifier, without artifact rejection, the offline re-analysis makes use of both types of features in order to assess the influence of a preceding artifact removal.

2.3. Robustness under novel paradigms and electrode setups

For the classification of artifactual IC components, three classification strategies—*fixed*, *adapted* and *study-specific*—were compared on the ERP-BCI and the CNT data set. Figure 1 visualizes the strategies. In the *fixed* scenario, classifier C_{RT} is trained once on features of labeled ICs of the RT data set, and furthermore applied to ICs of any other data set. Neither hand-labeling of novel ICs nor re-calculation of features or any re-training of the classifier is necessary in this simplest scenario. While hand-labeling of novel ICs is also avoided successfully in the *adapted* strategy, a channel adaptation on the RT-data is performed by cutting the training patterns to the specific electrode layout of the test data set. Features then need to be re-calculated based on the reduced patterns and a
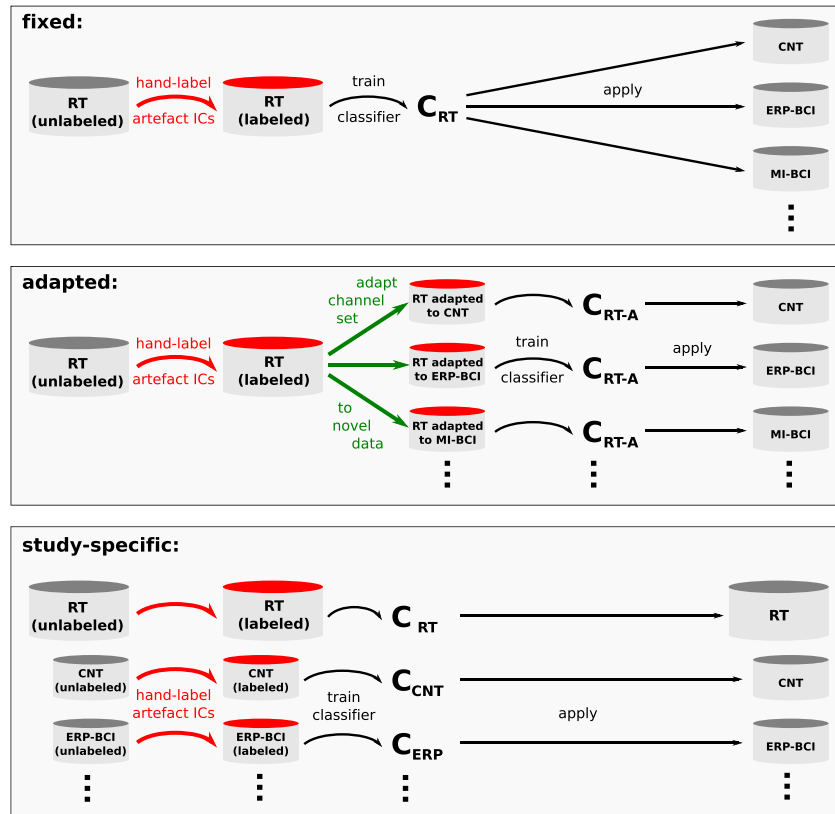
Figure 1. Schematic plot of the three transfer strategies *fixed*, *adapted* and *study-specific*. Expensive hand-labeling steps of ICs are marked with red arrows, cheap channel reduction and classifier training steps in green and black. Note that any self-application of classifiers in the *study-specific* strategy was performed exclusively in a leave-one-subject-out validation scenario.

re-training yields the *adapted* classifier C_{RT-A}. All steps can be performed automatically and do not require user input. The third strategy, *study-specific*, requires the effort of experts every time a novel study is performed. The ICs of at least some subjects need to be hand-labeled, before a study-specific classifier (e.g. C_{CNT} or C_{ERP}) can be trained and applied to novel subjects. It's performance was evaluated by leave-one-subject-out cross-validation.

To explore the robustness of the artifact classifier against reduced EEG channel sets, we compared the *fixed* IC-classifier C_{RT} with the *adapted* IC-classifier C_{RT-A} on the RT and ERP-BCI test data sets with reduced setups (varying from 16 to 104 resp. 61 EEG channels). All electrode setups were approximately equidistant and covered the whole scalp.

2.4. Effect on BCI performance

This offline re-analysis of three BCI paradigms described in section 2.2 compares standard BCI performance with and without a preceding ICA artifact cleaning. In both cases, artifactual channel and trial rejection based on a variance criterion was performed prior to BCI training. Training of the BCI-classifiers is based on the calibration runs only, and BCI performance tests are performed with the online runs of the participants.

ICA artifact cleaning is included in a manner that allows for real-time BCI applications. Prior to TDSEP, we estimated whether a PCA pre-processing to 99% explained variance would be useful via cross-validation on the calibration data. This was the case only for the LRP-MI paradigm. IC components were then derived by TDSEP and classified with the *adapted* classifier C_{RT-A} on the calibration data. The BCI is set up on the remaining ICs. On the online runs, un-mixing and component rejection is performed according to the de-mixing determined on the calibration data. The BCI classifier is applied to features extracted from the remaining components of the online runs.

3. Results

3.1. Robustness under novel electrode setups

Figure 2 shows the classification error for the *fixed* classifier C_{RT} and the *adapted* classifier C_{RT-A} for different channel setups on both the RT and the ERP-BCI test sets. On the RT test data with the full 104 channel setup, a classifier using all six features achieves a MSE of 9.3% only, which slightly outperforms the use of only four pattern-independent features (12.4% MSE). While C_{RT} generalizes robustly over the range of 104 to 48 electrodes in the RT test sets, its error increases up to 31.8% for the smallest set of 16 electrodes. On the ERP-BCI data set, the use of only four pattern-independent features is already outperforming the fixed classifier C_{RT} on the full 61 electrode setup. Classification performance of C_{RT} then breaks down to 50% on the smallest set of 16 electrodes. In both the RT and the ERP-BCI data set, the drop in overall performance is due to the bad performance of both pattern-based features of over 50%.
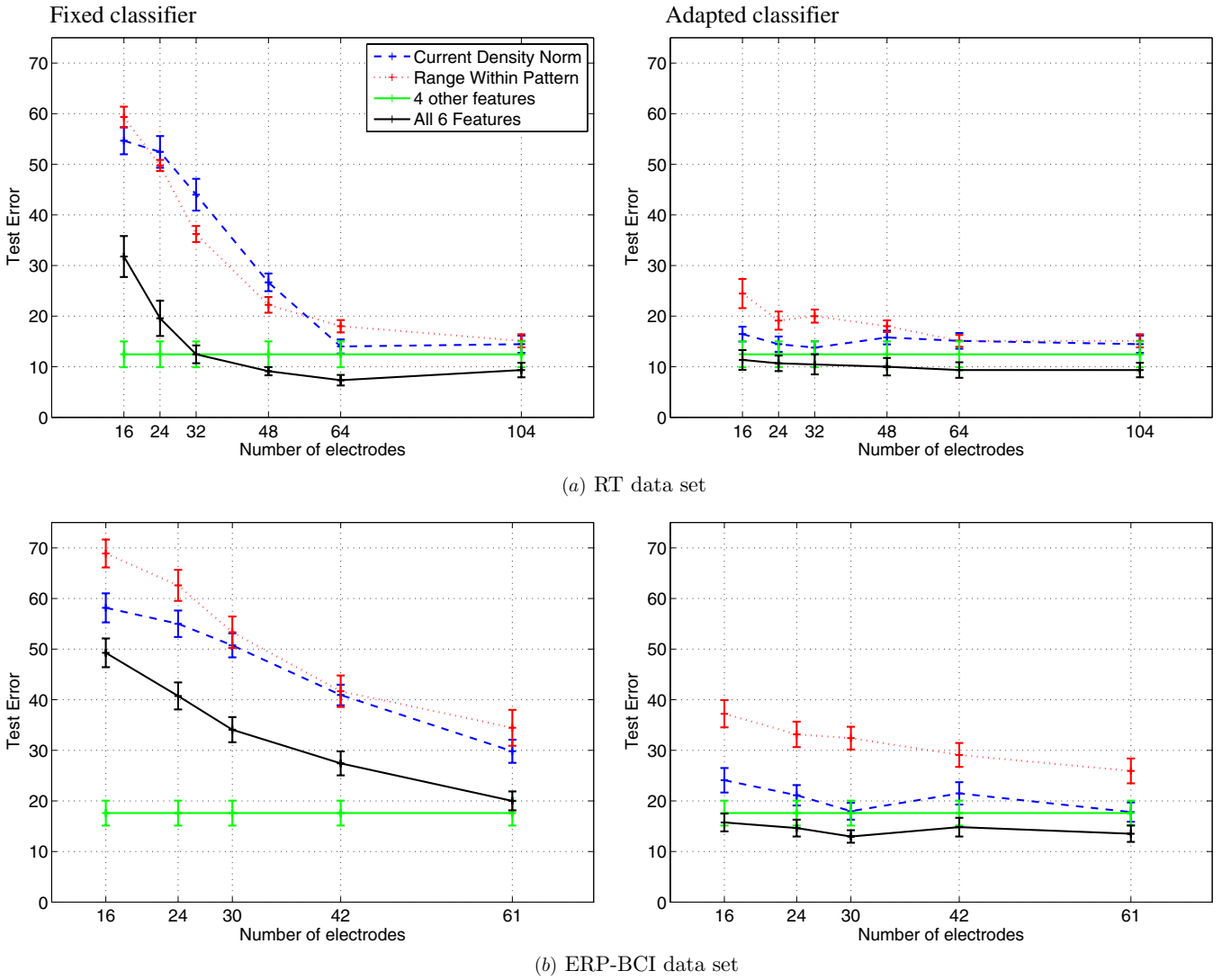
Figure 2. Mean classification error ± standard error estimated on (*a*) the RT and (*b*) the ERP-BCI test sets for different channel setups. The left plot shows the results for a fixed classifier, the right plot for a classifier adapted to each channel setup.

For the *adapted* strategy (i.e. re-training the classifier on the patterns cut to the specific electrode setup), the error of the pattern features (*range within pattern* and *current density norm*) was much less pronounced in both data sets. The overall error of C_{RT-A} for 16 electrodes remained at 11.3% on the RT data set (compared with 9.3% on 104 channels) and at 15.9% for the ERP-BCI data set (compared with 13.3% on 61 channels). In both data sets, we slightly gain from using the pattern features. On the reduced electrode setup, the classifier weight of the *range in pattern* dropped, while the weight for *current density norm* remained stable.

3.2. Robustness under novel paradigms

The results for the three proposed classification strategies on the three labeled IC data sets are summarized in table 1. The *adapted* classifier C_{RT-A} (trained on the RT data set cut to the specific electrode montage of the ERP-BCI or CNT data set) achieves an error of 13.3% on the ERP-BCI data and an error of 14.0% on the CNT data set.

The classification performance can be improved by a re-training on labeled data from the same study, but the effect is small. We observe an error of 9.3% on the RT data set, an error of 9.6% on the ERP-BCI data set and an error of 13.1% on the CNT data set. This improved performance is due to two effects: first, adjusting feature thresholds for the specific study may improve the performance of each feature. For example, a re-training of the 8–13 Hz feature of the CNT data set decreased its error from 33.3% to 18.0%. Second, feature weights adjust such that more discriminative features obtain a higher weight. Interestingly, after re-training both C_{ERP} and C_{CNT} primarily use one of the two pattern features—C_{ERP} focuses mostly on the *current density norm* feature, while C_{CNT} is strongly based on the *range within pattern* feature.

3.3. Effect on BCI performance

The upper plots of figure 3 show scatter plots of BCI performance with and without preceding ICA artifact cleaning for the three analyzed BCI paradigms. For ERP-BCI, BCI performance decreased slightly from 69.4% to 68.3% ($t(20) = -2.43$, $p = 0.03$, $d = 0.21$). On average, 44 components were retained and 16 artifactual components were removed. There was no significant change in overall MI-CSP performance
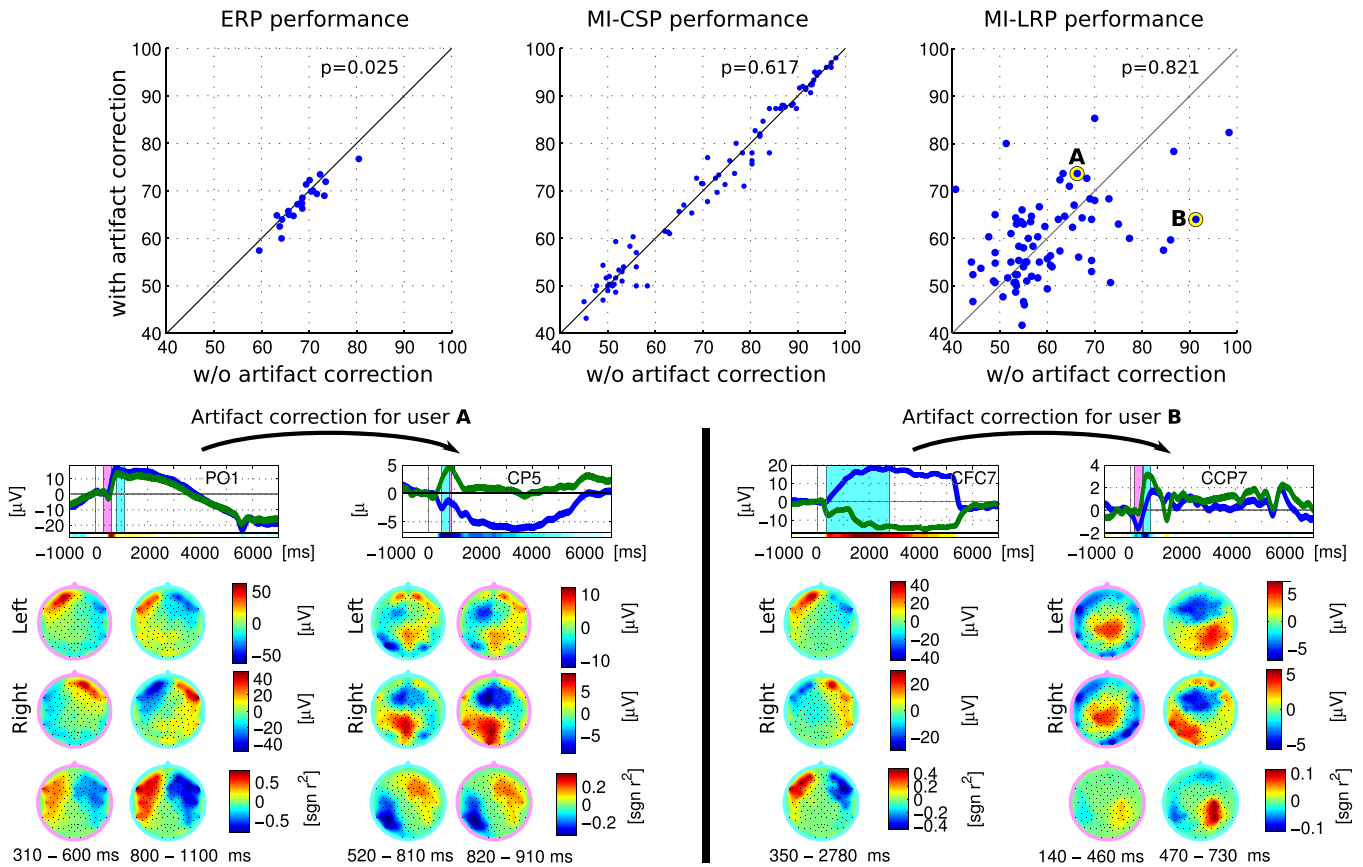
Figure 3. Upper plots: effect of artifact correction for three BCI paradigms. Dots over the diagonal indicate participants, whose data improved in classification performance (in per cent correct trials), dots below indicate participants whose performance decreased by the correction. Changes are strongest for the paradigm MI-LRP, which is most sensitive to eye artifacts. For this paradigm, participants (A) and (B) are highlighted, which undergo relatively strong changes. Lower plots: effect of artifact cleaning for participants (A) and (B). Top row: average activity of selected channels for left trials (blue) and right trials (green). The four upper scalp plots indicate the spatial distribution of average activity (in μV) for one or two time intervals (in columns) and for left and right trials (upper and lower scalp plots). Lowest scalp plots indicate the spatial distribution of class-discriminative information (as signed r^2 values) per interval. For participant A, a dominating eye artifact could be removed, which lead to an increase in the SNR and of classification performance. For participant B, very little class-discriminant signal remained after artifact cleaning.

Table 1. Feature weight vectors w and test errors (MSE) for three data sets (RT, ERP-BCI and CNT) and three classification strategies (fixed classifier C_{RT}, adapted classifier C_{RT-A} and study-specific classifiers C_{ERP}, C_{CNT}). Test errors are reported for the 6 single features and for the combined classification. The fixed classifier is trained on the RT train data set. The adapted classifier is trained on the RT train data set cut to the specific electrode montage. The study-specific classifiers are trained on data from the same study and evaluated with leave-one-subject-out CV.

			Current density norm	Range within pattern	Local skewness	λ	8–13 Hz	FitError	Combined
RT	C_{RT}	w	0.485	0.511	0.404	0.155	−0.522	−0.210	
		MSE	0.144	0.151	0.355	0.158	0.171	0.173	**0.093**
ERP-BCI	C_{RT}	MSE	0.296	0.289	0.459	0.244	0.154	0.357	**0.185**
	C_{RT-A}	w	0.454	0.463	0.384	0.235	−0.563	−0.247	
		MSE	0.178	0.259	0.459	0.244	0.154	0.357	**0.133**
	C_{ERP}	w	0.533	0.085	0.363	0.359	−0.650	−0.009	
		MSE	0.244	0.289	0.376	0.237	0.150	0.298	**0.096**
CNT	C_{RT}	MSE	0.421	0.198	0.275	0.190	0.323	0.489	**0.167**
	C_{RT-A}	w	0.341	0.498	0.417	0.234	−0.587	−0.251	
		MSE	0.265	0.214	0.275	0.190	0.323	0.489	**0.140**
	C_{CNT}	w	0.035	0.589	0.459	0.259	−0.602	−0.010	
		MSE	0.234	0.196	0.232	0.163	0.180	0.569	**0.131**

($t(79) = -0.50$, $p = 0.62$, $d = 0.04$) which remained constant at $\approx 72\%$ after the removal of on average 18 artifactual components (69 components were kept). In both BCI systems, the effect per subject was small.

The strongest changes were observed for the MI-LRP paradigm, which is most prone to eye artifacts due to the focus on low-frequency signal components. Note that as feedback was provided with a moving cursor, eye activity
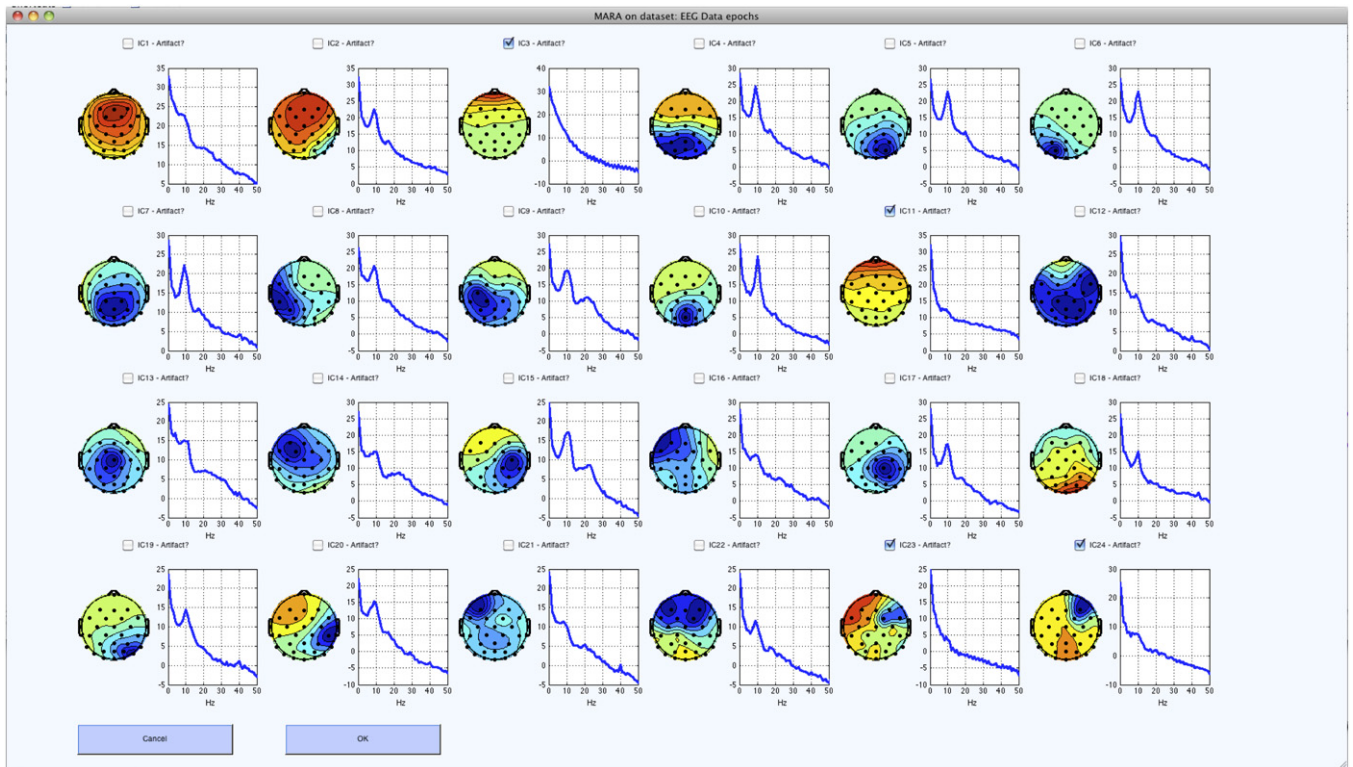
Figure 4. Screen shot of the MARA plug-in applied to EEGLAB sample data.

may be correlated with the two classes. On average, nine components were retained and ten artifactual components were removed. While the mean BCI accuracy remained constant at $\approx 60\%$ ($t(79) = 0.23$, $p = 0.82$, $d = 0.03$), the performance of each participant varied considerably. The lower plots of figure 3 exemplarily highlight the effect of the artifact rejection for two participants. Without artifact rejection, both participants mainly use eye artifacts for BCI control (frontal class-discriminative activation). The effect of artifact removal can be twofold. For participant A, eye artifacts obstruct the underlying neural activity, and the system's accuracy improved upon artifact cleaning from 66.3% to 73.6% due to an improved signal-to-noise level. In participant B, very little class-discriminant activity remained after the eye activity was removed. BCI classification dropped considerably from 91.3% to 64.0%.

4. Discussion

To summarize, we have analyzed the robustness properties of our recently proposed artifact classification method and proposed a strategy to handle a wide range of electrode setups. The proposed *adapted* strategy fully automates the time-consuming rating of artifactual ICs and reliably identified multiple types of artifacts from 35 participants and 3 EEG paradigms.

IC classification performance of three strategies was evaluated against expert ratings. We showed that our simplest automatic *fixed* strategy (train the classifier once, then apply to other setups) exhibits sensitivity to drastically reduced electrode setups. As a solution, we proposed the *adapted*

strategy which recomputes the training features based on the specific electrode montage of the test sets. Using this relatively inexpensive strategy—no hand-labeling is involved—artifact classification generalizes well even on very reduced electrode setups.

For comparison reasons, a re-training of the classifier using labor-intensively gained hand-labeled ICs from every new study was analyzed (strategy *study-specific*). While avoiding some generalization issues in theory, it is prohibitively expensive in most practical situations and only achieved a performance gain of a few per cent compared with the *adapted* strategy.

We therefore recommend the *adapted* strategy for artifact classification. It generalized robustly even to completely novel EEG paradigms, with its IC classification performance (13.3% MSE on auditory ERP data and 14.0% MSE on auditory listening data) staying on a similar level as inter-expert disagreements (often above 10% [34, 39]). This classification error is remarkably low given that the studies have been recorded with half the number of electrodes, used different ICA methods and contained different proportions of artifactual components.

We provide the ready-to-use artifact classifier to the community as an open-source EEGLAB plug-in called MARA (multiple artifact rejection algorithm). MARA automatically adapts to novel channel setups and its output is designed to support the experimenter in his or her decisions: a semi-automatic mode allows for visual inspection of components and for changing the classifier's proposed ratings. Figure 4 shows an example screen shot of the visual inspection menu. The plug-in is published under the

General Public License (GPL) and can be downloaded from www.user.tu-berlin.de/irene.winkler/artifacts/.

BCI practitioners may find the application of MARA on BCI data sets of particular interest. We used the *adapted* strategy to analyze how ICA artifact cleaning impacts on single-trial BCI performance of three different BCI paradigms. In all three paradigms, we were able to remove artifactual activity while maintaining the average BCI performance.

On the single subject level the effect of artifact cleaning depends on whether artifacts mask the relevant neural activity or serve as a control signal for BCI. While artifact cleaning had little influence on an auditory ERP speller and on oscillatory motor imagery data analyzed with CSP, we observed strong effects for a paradigm known to be heavily affected by eye artifacts, the use of slow motor-related potentials. Here our analysis suggests that artifact removal by MARA or similar tools may drastically improve the safety and reliability of results, as they guarantee that rejected artifacts are not utilized mistakenly to control the BCI system.

Acknowledgments

References

[1] Tangermann M *et al* 2012 Review of the BCI competition IV *Front. Neurosci.* **6** 55

[2] Kübler A, Nijboer F, Mellinger J, Vaughan T M, Pawelzik H, Schalk G, McFarland D J, Birbaumer N and Wolpaw J R 2005 Patients with ALS can use sensorimotor rhythms to operate a brain–computer interface *Neurology* **64** 1775–7

[3] Hill N J *et al* 2006 Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 183–6

[4] Conradi J, Blankertz B, Tangermann M, Kunzmann V and Curio G 2009 Brain–computer interfacing in tetraplegic patients with high spinal cord injury *Int. J. Bioelectromagn.* **11** 65–8

[5] Millán J del R *et al* 2010 Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges *Front. Neurosci.* **4** 161

[6] Pfurtscheller G, Allison B Z, Brunner C, Bauernfeind G, Solis-Escalante T, Scherer R, Zander T O, Mueller-Putz G,

Neuper C and Birbaumer N 2010 The hybrid BCI *Front. Neurosci.* **4** 42

[7] Gramann K, Gwin J T, Ferris D P, Oie K, Jung T-P, Lin C-T, Liao L-D and Makeig S 2011 Cognition in action: imaging brain/body dynamics in mobile humans *Rev. Neurosci.* **22** 593–608

[8] Debener S, Minow F, Emkes R, Gandras K and de Vos M 2012 How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology* **49** 1617–21

[9] Castermans T, Duvinage M, Petieau M, Hoellinger T, De Saedeleer C, Seetharaman K, Bengoetxea A, Cheron G and Dutoit T 2011 Optimizing the performances of a P300-based brain–computer-interface in ambulatory conditions *IEEE J. Emerg. Sel. Top. Circuits Syst.* **4** 566–77

[10] Hari R and Kujala M V 2009 Brain basis of human social interaction: from concepts to brain imaging *Physiol. Rev.* **89** 453–79

[11] Tangermann M, Krauledat M, Grzeska K, Sagebaum M, Vidaurre C, Blankertz B and Müller K-R 2009 Playing pinball with non-invasive BCI *Advances in Neural Information Processing Systems 21, December 8–11, 2008* ed D Koller, D Schuurmans, Y Bengio and L Bottou (Vancouver, BC: MIT Press) pp 1641–8

[12] Blankertz B *et al* 2010 The Berlin brain–computer interface: non-medical uses of BCI technology *Front. Neurosci.* **4** 198

[13] van Erp J, Lotte F and Tangermann M 2012 Brain–computer interfaces: beyond medical applications *Computer* **45** 26–34

[14] Fatourechi M, Bashashati A, Ward R K and Birch G E 2007 EMG and EOG artifacts in brain computer interface systems: a survey *Clin. Neurophysiol.* **118** 480–94

[15] Makeig S, Bell A J, Jung T-P and Sejnowski T J 1996 Independent component analysis of electroencephalographic data *Advances in Neural Information Processing Systems* vol 8 ed D S Touretzk, M C Mozer and M E Hasselmo (Cambridge, MA: MIT Press) pp 145–51

[16] Jung T-P, Makeig S, Humphries C, Lee T-W, Mckeown M J, Iragui V and Sejnowski T J 2000 Removing electroencephalographic artifacts by blind source separation *Psychophysiology* **37** 163–78

[17] Fitzgibbon S P, Powers D M W, Pope K J and Richard Clark C 2007 Removal of EEG noise and artifact using blind source separation *Clin. Neurophysiol.* **24** 232–43

[18] Romero S, Mañanas M A and Barbanoj M J 2008 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.* **38** 348–60

[19] Crespo-Garcia M, Atienza M and Cantero J L 2008 Muscle artifact removal from human sleep EEG by using independent component analysis *Ann. Biomed. Eng.* **36** 467–75

[20] McMenamin B W, Shackman A J, Maxwell J S, Bachhuber D R W, Koppenhaver A M, Greischar L L and Davidson R J 2010 Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG *NeuroImage* **49** 2416–32

[21] Campos Viola F, De Vos M, Hine J, Sandmann P, Bleeck S, Eyles J and Debener S 2012 Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials *Hear. Res.* **284** 6–15

[22] Gwin J T, Gramann K, Makeig S and Ferris D P 2010 Removal of movement artifact from high-density EEG recorded during walking and running *J. Neurophysiol.* **103** 3526–34

[23] Meinecke F, Ziehe A, Kawanabe M and Müller K-R 2002 Resampling approach to estimate the stability of one- or multidimensional independent components *IEEE Trans. Biomed. Eng.* **49** 1514–425

[24] Choi S, Cichocki A, Park H-M and Lee S-Y 2005 Blind source separation and independent component analysis: a review *Neural Inform. Process. Lett. Rev.* **6** 1–57

[25] Romero S, Mañanas M A, Riba J, Morte A, Giménez S, Clos S and Barbanoj M J 2004 Evaluation of an automatic ocular filtering method for awake spontaneous EEG signals based on independent component analysis *EMBS: 26th Annu. Int. Conf. Engineering in Medicine and Biology Society* pp 925–8

[26] Shoker L, Sanei S and Chambers J 2005 Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm *IEEE Signal Process. Lett.* **12** 721–4

[27] Gómez-Herrero G, De Clercq W, Anwar H, Kara O, Egiazarian K, Van Huffel S and Van Paesschen W 2006 Automatic removal of ocular artifacts in the EEG without an EOG reference channel *NORSIG: Proc. 7th Nordic Signal Processing Symp.* pp 130–3

[28] Campos Viola F, Thorne J, Edmonds B, Schneider T, Eichele T and Debener S 2009 Semi-automatic identification of independent components representing EEG artifact *Clin. Neurophysiol.* **120** 868–77

[29] Mognon A, Jovicich J, Bruzzone L and Buiatti M 2010 ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features *Psychophysiology* **48** 229–40

[30] Bigdely-Shamlo N, Kreutz-Delgado K, Kothe C and Makeig S 2013 Eyecatch: data-mining over half a million EEG independent components to construct a fully-automated eye-component detector *Proc. 35th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society* pp 5845–8

[31] Frøhlich L, Andersen T S and Mørup M 2013 Classification of independent components of EEG into multiple artifact classes *BaCI Conf. Abstract: Conf. on Basic and Clinical Multimodal Imaging*

[32] LeVan P, Urrestarazu E and Gotman J 2006 A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification *Clin. Neurophysiol.* **117** 912–27

[33] Halder S, Bensch M, Mellinger J, Bogdan M, Kübler A, Birbaumer N and Rosenstiel W 2007 Online artifact removal for brain–computer interfaces using support vector machines and blind source separation *Comput. Intell. Neurosci.* **7** 1–10

[34] Winkler I, Haufe S and Tangermann M 2011 Automatic classification of artifactual ICA-components for artifact removal in EEG signals *Behav. Brain Funct.* **7** 30

[35] Asadi Ghanbari A, Kousarrize M R N, Teshnehlab M and Aliyari M 2009 An evolutionary artifact rejection method for brain computer interface using ICA *Int. J. Electr. Comput. Sci.* **9** 48–53

[36] Bartels G, Shi L-C and Lu B-L 2010 Automatic artifact removal from EEG—a mixed approach based on double blind source separation and support vector machine *EMBC'10: Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society* pp 5383–6

[37] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

[38] Shackman A J, McMenamin B W, Slagter H A, Maxwell J S, Greischar L L and Davidson R J 2009 Electromyogenic artifacts and electroencephalographic inferences *Brain Topography* **22** 7–12

[39] Klekowicz H, Malinowska U, Piotrowska A J, Wolynczyk-Gmaj D, Niemcewi S and Durka P 2009 On the robust parametric detection of EEG artifacts in polysomnographic recordings *Neuroinformatics* **7** 147–60

[40] Kuhlen A K, Allefeld C and Haynes J-D 2012 Content-specific coordination of listeners' to speakers' EEG during communication *Front. Human Neurosci.* **6** 266

[41] Fonov V, Evans A C, McKinstry R C, Almli C R and Louis Collins D 2009 Unbiased nonlinear average age-appropriate brain templates from birth to adulthood *NeuroImage* **47** (Suppl. 1) S102

[42] Fonov V, Evans A C, Botteron K, Almli C R, McKinstry R C and Louis Collins D 2011 Unbiased average age-appropriate atlases for pediatric studies *NeuroImage* **54** 313–27

[43] Nolte G and Dassios G 2005 Analytic expansion of the EEG lead field for realistic volume conductors *Phys. Med. Biol.* **50** 3807–23

[44] Hämäläinen M S and Ilmoniemi R J 1994 Interpreting magnetic fields of the brain: minimum-norm estimates *Med. Biol. Eng. Comput.* **32** 35–42

[45] Haufe S, Nikulin V V, Ziehe A, Müller K-R and Nolte G 2008 Combining sparsity and rotational invariance in EEG/MEG source reconstruction *NeuroImage* **42** 726–38

[46] Ziehe A, Müller K-R, Nolte G, Mackert B-M and Curio G 2000 Artifact reduction in magnetoneurography based on time-delayed second-order correlations *IEEE Trans. Biomed. Eng.* **47** 75–87

[47] Hyvärinen A and Oja E 1997 A fast fixed-point algorithm for independent component analysis *Neural Comput.* **9** 1483–92

[48] Schreuder M, Rost T and Tangermann M 2011 Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI *Front. Neurosci.* **5** 112

[49] Blankertz B, Lemm S, Treder M S, Haufe S and Müller K-R 2011 Single-trial analysis and classification of ERP components—a tutorial *NeuroImage* **56** 814–25

[50] Blankertz B, Sannelli C, Halder S, Hammer E M, Kübler A, Müller K-R, Curio G and Dickhaus T 2010 Neurophysiological predictor of SMR-based BCI performance *NeuroImage* **51** 1303–9

[51] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K-R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Proc. Mag.* **25** 41–56

[52] Krauledat M, Dornhege G, Blankertz B, Losch F, Curio G and Müller K-R 2004 Improving speed and accuracy of brain–computer interfaces using readiness potential features *IEMBS'04: 26th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society* vol 2 pp 4511–5