Psychophysiology, 48 (2011), 229–240. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2010.01061.x

ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features

ANDREA MOGNON,[a,b] JORGE JOVICICH,[a] LORENZO BRUZZONE,[c] AND MARCO BUIATTI[a,d,e,f]

[a]Functional NeuroImaging Laboratory, Center for Mind/Brain Sciences, Department of Cognitive and Education Sciences, University of Trento, Trento, Italy
[b]NILab, Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy
[c]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
[d]INSERM, U992, Cognitive Neuroimaging Unit, Gif/Yvette, France
[e]CEA, DSV/I2BM, NeuroSpin Center, Gif/Yvette, France
[f]Université Paris-Sud, Cognitive Neuroimaging Unit, Gif/Yvette, France

Abstract

A successful method for removing artifacts from electroencephalogram (EEG) recordings is Independent Component Analysis (ICA), but its implementation remains largely user-dependent. Here, we propose a completely automatic algorithm (ADJUST) that identifies artifacted independent components by combining stereotyped artifact-specific spatial and temporal features. Features were optimized to capture blinks, eye movements, and generic discontinuities on a feature selection dataset. Validation on a totally different EEG dataset shows that (1) ADJUST's classification of independent components largely matches a manual one by experts (agreement on 95.2% of the data variance), and (2) Removal of the artifacted components detected by ADJUST leads to neat reconstruction of visual and auditory event-related potentials from heavily artifacted data. These results demonstrate that ADJUST provides a fast, efficient, and automatic way to use ICA for artifact removal.

Descriptors: Electroencephalography, Independent component analysis, EEG artifacts, EEG artefacts, Event-related potentials, Ongoing brain activity, Automatic classification, Thresholding

Due to its excellent temporal resolution, electroencephalography (EEG) is a widely used experimental technique to investigate human brain function by tracking the spatio-temporal neural dynamics correlated to experimentally manipulated events (Niedermeyer & da Silva, 2005). However, a major problem common to all EEG studies is that the activity due to artifacts has typically much higher amplitude than the one generated by neural sources. Artifacts may have a physiological origin as eye movements or muscle contractions, or non-biological causes as electrode high-impedance or electric devices interference (Croft & Barry, 2000).

In a typical event-related potential (ERP) paradigm, data are divided in epochs time-locked to the stimulus, and artifacts are removed by discarding epochs in which the EEG activity exceeds some predefined thresholds either in specific electrodes (e.g., electrooculogram (EOG) signals for ocular movements) or in all electrodes throughout the scalp. Artifact-free ERPs are then obtained by averaging the data over the remaining epochs, thereby increasing the signal-to-noise ratio. However, this procedure is problematic when only a few epochs are available, or when artifacts are very frequent, as in studies involving patients or children. Moreover, it is inapplicable to studies focusing on slow non-event-locked activity arising from the continuous EEG (e.g., slow brain oscillations (Vanhatalo, Palva, Holmes, Miller, Voipio, & Kaila, 2004) or long-range temporal correlations (Linkenkaer-Hansen, Nikouline, Palva, & Ilmoniemi, 2001)).

Alternative procedures consist of modelling the signals generated by blinks or ocular movements and removing them from the data while preserving the remaining activity. Most of these methods are based on regressing out reference signals, usually recorded near the eyes, from the EEG signals with a model of artifact propagation, either in the time domain (Gratton, Coles, & Donchin, 1983; Kenemans, Molenaar, Verbaten, & Slangen, 1991; Verleger, Gasser, & Mocks, 1982) or in the frequency domain (Gasser, Sroka, & Mocks, 1985; Woestenburg, Verbaten, & Slangen, 1983). However, because EEG and ocular activity mix bidirectionally (Oster & Stern, 1980; Peters, 1967), regressing out eye artifacts inevitably involves subtracting relevant neural signals from each recording as well as ocular activity (Croft & Barry, 2002; Jung, Makeig, Westerfield, Townsend, Courchesne,

We thank Mariano Sigman and Stanislas Dehaene for sharing the EEG data, Francesca Bovolo and Michele Dalponte for helpful advice on the use of the thresholding algorithm, and Sara Assecondi for valuable comments on an earlier version of the manuscript.

Address correspondence to: Marco Buiatti, CEA/DSV/I2BM/NeuroSpin, INSERM U992—Cognitive Neuroimaging Unit, Bât 145—Point Courrier 156, Gif sur Yvette F-91191, France. E-mail: marco.buiatti@gmail.com

PSYCHOPHYSIOLOGY

& Sejnowski, 2000). Moreover, these methods do not work without reference signals, which are not always present for ocular movements, and very difficult to obtain for other types of artifacts (muscular, non-biological).

A recent successful approach to this problem is the use of independent component analysis (ICA) (Jung, Humphries, Lee, Makeig, McKeown, Iragui, & Sejnowski, 1998), a statistical tool that decomposes EEG data in a set of sources with maximally independent time courses. ICA proved very efficient in separating activity related to a large number of artifacts from neural activity by automatically segregating the former in specific independent components (ICs) (Jung, Makeig, Humphries, Lee, McKeown, Iragui, & Sejnowski, 2000; Vigário, Särelä, Jousmäki, Hämäläinen, & Oja, 2000). Since the number of sources is potentially much higher than the number of ICs (Baillet, Mosher, & Leahy, 2001; Liu, Dale, & Belliveau, 2002), this separation will never be perfect (Groppe, Makeig, & Kutas, 2009). However, after removing non-stereotyped artifacts by an accurate preprocessing of the data, it is possible to obtain 'clean' ICA decompositions (Onton, Westerfield, Townsend, & Makeig, 2006), such that removing artifacted ICs from the data by simple subtraction generally leads to marginal distortion of the remaining EEG data (Joyce, Gorodnitsky, & Kutas, 2004; Jung, Makeig, Humphries, et al., 2000).

Nevertheless, the practical usability of ICA as a tool for artifact rejection has an important limitation: the detection of the ICs associated with artifacts is time-consuming and involves subjective decision making. Several attempts have been made to guide IC classification by using a number of measures to discriminate artifacted from non-artifacted ICs either in the time domain (Barbati, Porcaro, Zappasodi, Rossini, & Tecchio, 2004; Delorme, Sejnowski, & Makeig, 2007; Mantini, Franciotti, Romani, & Pizzella, 2008), in the space domain (Li, Ma, Lu, & Li, 2006; Viola, Thorne, Edmonds, Schneider, Eichele, & Debener, 2009) or in both (Joyce et al., 2004; Okada, Jung, & Kobayashi, 2007). A single discriminative measure may already be very helpful in detecting specific artifacts (blinks (Li et al., 2006; Okada et al., 2007; Viola et al., 2009), lateral eye movements (Viola et al., 2009), heartbeat artifacts (Viola et al., 2009)) or even a wide variety of biological and non-biological artifacts (Mantini et al., 2008). Multiple measures (Barbati et al., 2004) and additional information from EOG signals (when available) (Joyce et al., 2004; Okada et al., 2007) have been used to improve this detection. However, these algorithms are not completely automatic since they either require a training set (Delorme et al., 2007; Li et al., 2006; Mantini et al., 2008), the arbitrary tuning of the thresholds separating artifacted from non-artifacted components (Barbati et al., 2004; Joyce et al., 2004; Okada et al., 2007) or an initial topography template (Viola et al., 2009).

Here, a completely automatic ICA-based algorithm for identification of artifact-related components in EEG recordings is proposed. The algorithm is built on the basis of two main observations: (1) for a large number of artifacts, artifact-related ICs are characterized by stereotyped features both in their temporal course and spatial distribution; and (2) while single features may not be accurate enough when discriminating artifact from non-artifact components, their combination can efficiently and systematically achieve this goal. The proposed algorithm is called ADJUST (Automatic EEG artifact Detection based on the Joint Use of Spatial and Temporal features) because it automatically "adjusts" its parameters to the data to compute the set of artifact-specific spatial and temporal features needed for IC classification without any additional information (e.g., EOG channels).

The first step of ADJUST consists of decomposing the EEG data into ICs. Four artifact classes are then considered: three classes are related to ocular artifacts (blinks, vertical, and horizontal eye movements), and a generic artifact class (called discontinuity) is devoted to capturing anomalous activity recorded at single electrodes due to high-impedance conditions or electrical instabilities in the recording device. For each of the four artifact classes, a detector is implemented by computing a class-specific set of spatial and temporal features on all ICs. For each feature, a threshold dividing artifacts from non-artifacts is estimated on the whole set of ICs in a completely automatic way by the Expectation-Maximization automatic thresholding method (Bruzzone & Prieto, 2000). For each detector, ICs are classified as artifacts if all artifact-specific spatial and temporal features belonging to that detector exceed their respective threshold value.

ADJUST spatial and temporal features were optimized on a feature selection EEG dataset. ADJUST was then validated on a validation dataset recorded with a different EEG system in a different laboratory and with a different paradigm with respect to the feature selection dataset. Validation consisted of three steps: first, ADJUST classification accuracy was evaluated by comparing it to a manual classification performed by three independent experts (as in Mantini et al., 2008); second, the advantage of using a combination of features was evaluated; third, ADJUST performance in recovering clean ERP topographies from artifacted data was assessed by comparing ERPs computed after ADJUST correction with the ones obtained from uncorrected data and the ones obtained after the manual classification by experts.

Materials and Methods

EEG Data Acquisition and Experimental Design

Two different datasets were used in this study: a *feature selection dataset* for selecting and implementing the optimal features for the artifact detection algorithm, and a *validation dataset* for testing the accuracy of artifact detection. Relative to the feature selection dataset, the validation dataset uses data recorded with a different EEG system in a different laboratory and with a different experimental design.

The *feature selection dataset* consists of EEG recordings drawn from a study investigating serial and parallel processing during dual-task performance (details in Sigman & Dehaene, 2008). Twenty-one right-handed native French speakers (10 women; mean age 24 years, ranging from 20 to 33 years) participated in the experiment. Data from one subject could not be used for the current study because of data corruption. All participants provided informed written consent to take part in the experiment, which was approved by the Comité Consultatif pour la Protection des Personnes dans la Recherche Biomèdicale, Hôpital de Bicetre (Le Kremlin-Bicetre, France).

In brief, participants were asked to perform two tasks with variable onset delay. Here, only trials with a delay onset of 1200 ms will be considered (240 trials per subject) since shorter delays involve complex effects of interference between the two tasks that were not of interest to the current study. Four more subjects were discarded because of a high rate of artifacts in the selected trials.

Subjects responded to both tasks with key presses, with the right hand for the first task (number-comparison) and with the left hand for the second task (auditory tone discrimination). In the number-comparison task, a number, which varied randomly among four different values (28, 37, 53, and 62), was flashed in the center of the screen for 150 ms, and subjects had to respond whether the number was larger or smaller than 45. In the auditory task, subjects had to respond whether the tone (lasting 150 ms) was high (880 Hz) or low (440 Hz) frequency. Stimuli were presented centrally on a black-and-white display on a 17-inch monitor with a refresh rate of 60 Hz. Subjects sat 1 m from the screen. Auditory stimulation was provided through headphones. During the EEG recording, intertrial intervals (ITIs) were jittered in the range from 3 to 4.2 s (mean ITI, 3.6 s). EEG recordings were sampled at 250 Hz with a 128-electrode geodesic sensor net (EGI, Eugene, OR) referenced to the vertex. Since a high-density channel distribution is not a requirement for the current study, for computationally faster ICA calculation the number of channels was reduced from 128 to 63 while keeping the channels' distribution uniform throughout the scalp.

The *validation dataset* consists of EEG recordings drawn from a study investigating the neural substrates of semantic representations with a semantic priming paradigm.[1] Ten right-handed native Italian speakers participated in the experiment (5 women, mean age 29 years, ranging from 21 to 39 years). On each trial, subjects were presented with two stimuli: a word presented through headphones, and a word presented visually, for a total of 360 auditory-visual pairs. The onset delay between the two stimuli was 800 ms. During the EEG recording, ITIs were jittered in the range from 1.7 to 2.7 s (mean ITI, 2.2 s). Trials comprised equal proportions of semantically related pairs at three different levels of semantic relatedness. Participants were asked to press a button whenever they read the name of a city. Catch trials including a city name were discarded from further analyses. Visual stimuli were presented centrally for 160 ms on a black-and-white display on a 17-inch monitor with a refresh rate of 75 Hz. Subjects sat 1 m from the screen. Auditory stimulation was provided through headphones. All participants provided informed written consent to take part in the experiment, which was approved by the Ethical Committee of the University of Trento (Italy). EEG was recorded at 64 electrodes (BrainAmp, Munich, Germany) referenced to the vertex, and sampled at 500 Hz.

EEG Data Preprocessing

Data were processed using the EEGLAB toolbox (Delorme & Makeig, 2004) and custom-made software running on Matlab 7.5.0 R2007b (MathWorks, Natick, MA) on a CentOS 5.0 Linux system (Xeon CPU X5365 @3.00 GHz quad core, 32 GB of RAM, Intel, Santa Clara, CA).

Continuous data from both datasets were visually inspected to discard paroxysmal portions of artifacted data and high-pass filtered at 0.5 Hz with a basic FIR filter to remove linear trends (EEGLAB tutorial, http://sccn.ucsd.edu/eeglab/) and improve the reliability of ICA decomposition (Groppe et al., 2009). Epochs of 2.5-s duration were extracted starting 500 ms before the onset of the first stimulus. To further remove non-stereotyped artifacts that would significantly affect the quality of the ICA decomposition (Onton et al., 2006), epochs in which recordings at any channel exceeded $\pm$ 150 μV were rejected (average number of rejected epochs: 31 $\pm$ 9 for the feature selection dataset, 36 $\pm$ 10 for the validation dataset). The remaining epochs were low-pass filtered at 40 Hz for the feature selection dataset and at 25 Hz for the validation dataset to minimize muscular artifacts (more frequent in the validation dataset), a class of artifacts that is not considered in the current study.

[1]Buiatti M., Finocchiaro, C., Mognon, A., Caramazza, A., Dehaene, S., and Piazza, M., in preparation.

ADJUST: Implementation

The three main steps of ADJUST are illustrated in the scheme of Figure 1, and described in the following.

Independent Component Analysis

ICA is a well known technique in signal processing literature that detects and separates the information sources associated with multidimensional signals (Hyvarinen, Karhunen, & Oja, 2001). ICA can be used for identifying the information sources mixed in the EEG data (Lee, Girolami, & Sejnowski, 1999). Let us assume that a set of q measured observations of random variables $g(t) = [g_1(t), \ldots, g_q(t)]^T$ is given by linear combination of p independent source signal components $s(t) = [s_1(t), \ldots, s_p(t)]^T$, whose number is at most equal to the number of observations ($p \leq q$); source activity is supposed to be non-Gaussian (Lee et al., 1999) or non-white in time (Belouchrani, Abed-Meraim, Cardoso, & Moulines, 1997).

The ICA model can be expressed in the general case as

$$g(t) = As(t) + n(t) \tag{1}$$

indicating that the observations $g(t)$ can be obtained by mixing the sources $s(t)$ via a constant $[q \times p]$ matrix $\boldsymbol{A}$ called mixing matrix and adding the vector of white noise $n(t)$ (which is not considered in some implementations). The mixing matrix is full column-rank ($r(\boldsymbol{A}) = p$).

Given these hypotheses, a solution to the problem of the identification of the ICA components can be implemented, and

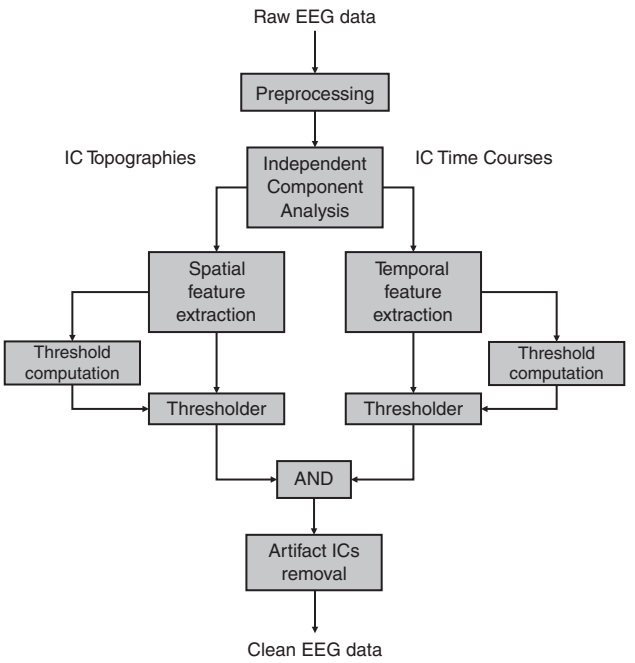

Figure 1. Architecture of the ADJUST algorithm for a generic detector with one spatial and one temporal feature. Any supplementary spatial or temporal feature can be added in parallel to the existing ones within the same architecture.

the ICs can be estimated by determining a $[p \times q]$ matrix $\boldsymbol{W}$ called *unmixing matrix* for which the vector

$$\hat{s}(t) = Wg(t) \quad (2)$$

is the best estimate of $s(t)$.

In this work, the INFOMAX algorithm (Bell & Sejnowski, 1995) implementation for ICA included in the EEGLAB toolbox was used. The INFOMAX algorithm is based on a learning rule which minimizes the mutual information between the source signals estimates, which is equivalent to maximizing the joint entropy between the estimates, in order to estimate the sources, which are assumed to be super-Gaussian. INFOMAX ICA estimates q ICs from a set of q observation vectors.

ICA decomposition was computed on all datasets, separately for each subject. The number of epochs was sufficiently large to ensure a good performance of the ICA algorithm, as the (number of time points)/(number of electrodes)^2 (considered as a predictor of ICA reliability (Groppe et al., 2009)) ranged between 18.70 and 50.07 (38.8 ± 7.2 average $\pm$ std) for the feature selection dataset, and between 68.02 and 120.93 (91.4 ± 12.6 average $\pm$ standard deviation) for the validation dataset.

Features Computation

We searched for spatial and temporal features that best captured the behavior of the ICs associated with four different artifact classes: eye blinks, vertical eye movements, horizontal eye movements, and generic discontinuities (see Figure 2 for examples of IC topographies and time courses typical of each artifact class and of a neural component).

Since artifact-specific IC topographies are characterized by a particular spatial shape, which is independent of the overall scale of the topography, IC topography weights were normalized with respect to their norm across the scalp: $a(n) = a_{original}(n)/\sqrt{\sum_m a^2_{original}(m)}$, where $a(n)$ is the topography weight at sensor n, $a_{original}(m)$ is the topography weight originally computed by ICA (the vector of all topography weights $a_{original}(n)$ corresponds to one column of the mixing matrix A (Equation 1)) and the sum is computed across all sensors m. For coherence with the ICA model (Equation 1), each corresponding IC activation was multiplied by the same factor $\sqrt{\sum_m a^2_{original}(m)}$.

Ideal features should maximally discriminate artifact from non-artifact ICs, resulting in a bimodal distribution of values, or in case of few artifacts, in artifact IC values falling on the tails (outliers) of the non-artifact IC values distribution. For each artifact class, several measures were tested on the feature selection EEG dataset in a trial-and-error approach following this criterion, and the most effective artifact-specific features were selected for artifact classification.

Hereafter we describe the selected features for each artifact class.

1. *Eye Blinks:* Eye blinks typically generate abrupt amplitude jumps in frontal electrodes. Their time course is well captured by the kurtosis (Barbati et al., 2004; Delorme et al., 2007), a measure that is very sensitive to outliers in the amplitude distribution. Since its sensitivity to abrupt jumps would be hampered by slow amplitude drifts on the whole IC time course, here the kurtosis is computed within each epoch after removing the epoch mean, and then averaged over epochs:

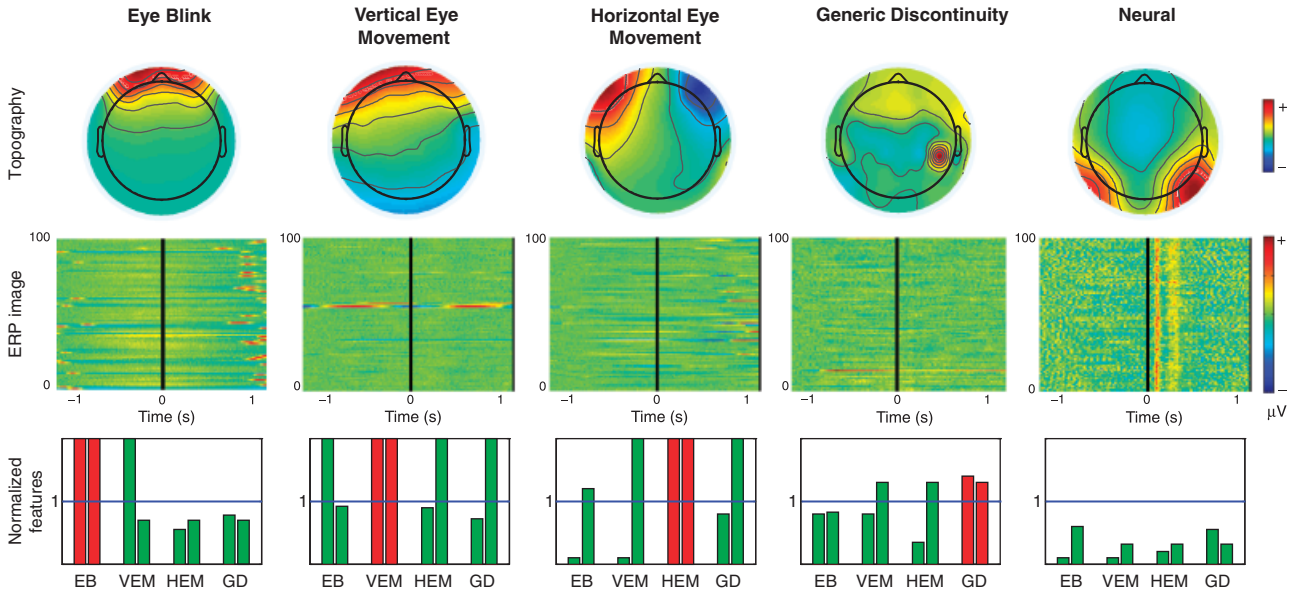

Figure 2. Examples of typical ICs from each artifact class and of a typical neural IC drawn from the validation dataset. Top row: IC topography. Middle row: ERP image illustrating the color-coded amplitude fluctuations (arbitrary units) of the IC in 100 contiguous epochs (time relative to visual word presentation on the x-axis, epochs on the y-axis). Bottom row: histograms of feature values (Spatial Average Difference and Temporal Kurtosis for eye blinks (EB); Spatial Average Difference and Maximum Epoch Variance for vertical eye movements (VEM); Spatial Eye Difference and Maximum Epoch Variance for horizontal eye movements (HEM); Generic Discontinuities Spatial Feature and Maximum Epoch Variance for generic discontinuities (GD)) normalized by the corresponding automatically calculated threshold value. Bars of features belonging to the same artifact class are grouped together, and are marked in red color if they all cross the threshold, indicating that ADJUST classifies the IC as a component of that artifact class.

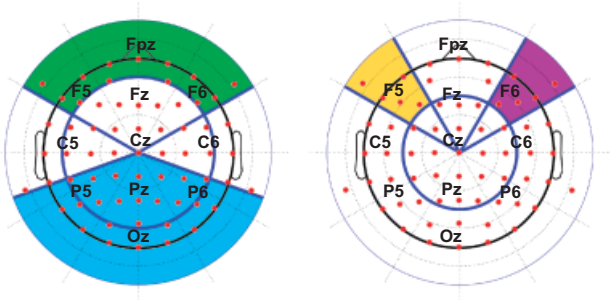

Figure 3. Scalp areas used in ADJUST spatial features computation. Left-hand panel: Frontal Area (in green) and Posterior Area (in blue); Right-hand panel: Left Eye area (in yellow) and Right Eye area (in purple). Red dots indicate channel positions in the validation dataset.

$$Temporal\ Kurtosis = trim\_and\_mean\left(\frac{\left\langle s_i(t)^4\right\rangle_{ep}}{\left\langle s_i(t)^2\right\rangle_{ep}^2} - 3\right)_i, \tag{3}$$

where $s_i(t)$ indicates the time course of the IC as defined by Equation (2) within the epoch i, $\langle\ldots\rangle_{ep}$ indicates the average within an epoch, and $trim\_and\_mean(\ldots)_i$ denotes the average across epochs computed after the top 1% of the values have been removed. This measure was preferred to the simple average because the latter would be too sensitive to spurious outliers.

To capture the spatial topography of blink ICs, we used a measure specifically sensitive to higher amplitude in frontal areas compared to posterior areas:

$$Spatial\ Average\ Difference = \left|\langle a\rangle_{FA}\right| - \left|\langle a\rangle_{PA}\right|, \tag{4}$$

where a is the vector of normalized IC topography weights defined above, $\langle\ldots\rangle_{FA}$ denotes the average over all channels in the frontal area (FA) (radial range: $0.4<r<1$; angular range from medial line: $0<|\theta|<60^\circ$ (if present, electrodes below the eyes would not be included), see Figure 3), and $\langle\ldots\rangle_{PA}$ denotes average over all channels in the posterior area (PA) (radial range: $0<r<1$; angular range: $110^\circ<|\theta|<180^\circ$, see Figure 3).

Two additional controls were imposed:

a) The average IC topography weights across the left eye area (LE) (radial range $0.3<r<1$ and angular range $-61^\circ<\theta<-29^\circ$, see Figure 3) must have the same sign of the average IC topography weights across the right eye area (RE) ($0.3<r<1$ and $29^\circ<$theta$<61^\circ$, see Figure 3) (to distinguish blinks from horizontal eye movements);
b) Variance of scalp weights included in the FA (defined above) must be higher than the variance of scalp weights included in the PA; this control is quantified by the feature

$$Spatial\ Variance\ Difference = \left(\langle a^2\rangle_{FA} - \langle a\rangle_{FA}^2\right) - \left(\langle a^2\rangle_{PA} - \langle a\rangle_{PA}^2\right), \tag{5}$$

which should be positive for eye blink components. This control is useful against false positives in cases where IC weights across the PA span both positive and negative values such that their average is very low, resulting in a spuriously high value of the spatial average difference (SAD).

2. *Vertical Eye Movements:* Vertical eye movements generate large amplitude fluctuations in frontal channels that are typically slower than those generated by blinks, therefore not efficiently identifiable by the kurtosis. They are well captured by a temporal feature based on the variance of the signal within each epoch:

$$Maximum\ Epoch\ Variance = \frac{trim\_and\_\max\left(\left\langle s_i(t)^2\right\rangle_{ep} - \langle s_i(t)\rangle_{ep}^2\right)_i}{trim\_and\_mean\left(\left\langle s_i(t)^2\right\rangle_{ep} - \langle s_i(t)\rangle_{ep}^2\right)_i}, \tag{6}$$

where $trim\_and\_max(\ldots)_i$ indicates the maximum of the trimmed vector of variance values over the epochs (as for the kurtosis, this measure was preferred to the simple maximum because the latter would be too sensitive to spurious outliers); this measure is normalized with respect to the average of trimmed variance values ($trim\_and\_mean(\ldots)_i$, see Temporal Kurtosis definition above) in order to better capture the difference from the baseline behavior of the time course. Trim was performed as explained for Temporal Kurtosis.

Since the spatial distribution of vertical eye movement artifacts is similar to that of blink artifacts, the same spatial feature (SAD) was used, together with the same additional controls.

3. *Horizontal Eye Movements:* Since the time course of artifacts caused by horizontal eye movements is similar to the one generated by vertical eye movements, the temporal feature used is the same (Maximum Epoch Variance (Equation 6)).

The spatial distribution is characterized by large amplitudes in frontal channels near the eyes, typically in anti-phase (one negative and one positive). A spatial feature sensitive to this pattern is

$$Spatial\ Eye\ Difference = \left|\langle a\rangle_{LE} - \langle a\rangle_{RE}\right|, \tag{7}$$

where $\langle\ldots\rangle_{LE}$ ($\langle\ldots\rangle_{RE}$) denotes average overall channels in the LE area (RE area) defined above, respectively.

To check that amplitudes are in anti-phase, one additional control is added: the average of IC topography weights in the LE and RE must have a different sign.

4. *Generic Discontinuities:* Artifacts generated by impedance fluctuations or electronic device interference typically involve sudden amplitude fluctuations in one channel, with no spatial preference. The time course of this artifact is captured by Maximum Epoch Variance (Equation 6). Its spatial distribution is captured by a feature sensitive to local spatial discontinuities:

$$Generic\ Discontinuities\ Spatial\ Feature = \max\left(\left|a_n - \langle k_{mn}a_m\rangle_m\right|\right)_n, \tag{8}$$

where a_n is the nth topography weight, $k_{mn} = \exp(-||y_m - y_n||)$ decays exponentially with the distance $||y_m - y_n||$ between channel m and channel n, $\langle\ldots\rangle_m$ denotes the average over all channels $m\neq n$, and $\max(\ldots)_n$ indicates the maximum over all channels n of the scalp.

Automatic Classification

For each feature included in the detectors, the threshold value was computed by means of a completely automatic image processing thresholding algorithm based on the Expectation-Maximization (EM) technique (Bruzzone & Prieto, 2000). This algorithm is expected to work in a 1-dimensional feature space where a set $\Omega = \{\omega_n,\omega_a\}$ of two information classes ω_n and ω_a is defined. The classes ω_n and ω_a represent the cases in which an IC component is not associated to an artifact or is associated to an artifact, respectively. The EM algorithm estimates the *a priori* probabilities of the classes and their probability density functions. The former model the probability that a random sample belongs to a given class, while the latter describe the distribution of each class's random variable. Probability density functions are assumed normally distributed, and thus they are modelled by mean values μ_n, μ_a and variances ${\sigma_n}^2$, ${\sigma_a}^2$ of the two Gaussian distributions.

In the first step, Expectation, an approximation for the two Gaussian distributions is computed from the data. The overall distribution is initially divided into two clusters that approximately contain entries from ω_n and entries from class ω_a; given the middle value of the histogram $M_D = (\max\{X_D\}+\min\{X_D\})/2$, where X_D indicates the vector of feature values, two thresholds T_n and T_a equally distant from M_D (distance: 0.01 $(\max\{X_D\} - M_D)$ from M_D) are used in this initial step to separate the clusters: entries in the interval $[min\{X_D\},\ T_n]$ are included in cluster 1, and entries in the interval $[T_a,\ max\{X_D\}]$ are included in cluster 2. Mean, variance and prior probability computed from the clusters are assumed to be the statistics of the classes' distributions at step zero of the following iteration process.

The iteration process, named Maximization, refines the statistics of the classes' distributions by maximizing a log-likelihood measure. At each iteration, the statistics prior probability, mean, and variance of the distributions are updated as:

$$P^{t+1}(\omega_n) = \left(\sum_{X(i)\in X_D} [P^t(\omega_n)p^t(X(i)/\omega_n)/p^t(X(i))] \right) / I, \quad (9)$$

$$\mu_n^{t+1} = \left(\sum_{X(i)\in X_D} \frac{P^t(\omega_n)p^t(X(i)/\omega_n)}{p^t(X(i))} X(i) \right) / \left(\sum_{X(i)\in X_D} \frac{P^t(\omega_n)p^t(X(i)/\omega_n)}{p^t(X(i))} \right), \quad (10)$$

$$(\sigma_n^2)^{t+1} = \left(\sum_{X(i)\in X_D} \frac{P^t(\omega_n)p^t(X(i)/\omega_n)}{p^t(X(i))} [X(i) - \mu_n^t]^2 \right) / \left(\sum_{X(i)\in X_D} \frac{P^t(\omega_n)p^t(X(i)/\omega_n)}{p^t(X(i))} \right), \quad (11)$$

where the superscripts t and $t+1$ indicate the current and successive iteration, respectively, $X(i)$ denotes the ith feature value and I is the length of the feature vector X_D. Analogous equations can be written for class ω_a.

This iterative process is repeated until the difference between any of the statistics at step i and the same statistic at step $I+1$ is lower than 10^{-4} times the statistic at step zero. At convergence, the threshold value is computed as the intersection between the estimated Gaussian distribution of class ω_n and the estimated Gaussian distribution of class ω_a, where the Gaussian distributions are computed from the statistics estimated in Equations (9)–(11).

In the last step of ADJUST, each detector checks whether each IC feature value is above the respective threshold; if this occurs for *all* the features belonging to that detector, the IC is marked as artifacted IC (it belongs to ω_a) for that artifact class. Virtually artifact-free EEG data are thus obtained by simply subtracting the artifacted ICs from the data.

A free version of the ADJUST software with sample data used in this study will be publicly released in the form of a plug-in toolbox to be run under the EEGLAB software (http://sccn.ucsd.edu/eeglab/).

ADJUST: Validation Procedure

The validation procedure was divided into three steps: (1) determination of the accuracy of IC classification; (2) evaluation of the benefit of combining spatial and temporal features compared to their separate use; and (3) determination of the accuracy in ERP reconstruction after removal of the ICs detected as artifacts by ADJUST.

Artifact Classification Accuracy

In the first step, ADJUST's IC classification was compared to manual IC classification performed by three independent scorers with proven expertise in the field of EEG analysis and familiarity with ICA decomposition of EEG data. Experts manually classified ICs from the feature selection datasets and the validation datasets by visualizing IC properties via the EEGLAB software package (Delorme & Makeig, 2004). Experts were invited to inspect the IC topography, power spectrum, and 'ERP image' (Jung, Makeig, Humphries, et al., 2000), a useful graphic representation displaying the IC time course of all epochs within the same figure by coding amplitudes in color scale (see Figure 2, middle row). Experts were asked to mark the ICs relative to the four classes of artifacts defined above (blinks, vertical and horizontal eye movements, discontinuities), and were invited to do so by looking both at the topography and at the time course of each component. Discontinuities were defined as sudden jumps with localized, non-biological spatial distribution. Experts were asked to mark only components that clearly belonged to one artifact class, and to not mark components also containing some presumably neural portion, as well as ambiguous components. Experts classified a total of 1008 ICs of the feature selection dataset and 630 ICs of the validation dataset. A unique classification, further referred to as 'manual classification,' was generated from the three scorers' classifications by using a majority criterion.

Manual and ADJUST classifications were then compared by computing an agreement measure for each class-specific detector. An additional agreement measure was generated for the detection of all types of artifacts (an IC was considered to be artifacted if detected by at least one single detector), which we will refer to as 'general artifact detection.'

The agreement measure γ was computed as the ratio between the variance accounted for by the ICs for which the two classifications agree (IC marked as artifact or non-artifact in both classifications) and the total variance of all ICs:

$$\gamma = \frac{\sum_{AgreementICs} \nu_i}{\sum_i \nu_i} \quad (12)$$

where v_i indicates the variance accounted for by the ith IC computed (using the EEGLAB function *eeg_pvaf*()) as the average variance of the IC activations back-projected into the electrode space, and *AgreementICs* is the list of ICs for which the two classifications agree. This measure is very similar to the agreement measure used in Li et al. (2006), the only difference being that in γ each IC is weighted by the variance it accounts for. The rationale behind this difference is that the more variance is explained by the IC, the more it is important to correctly classify that IC as artifacted or not.

Effects of Combining Spatial and Temporal Features

ADJUST detectors can be thought of as AND-detectors because they identify an artifact only when all associated temporal and spatial features have values higher than the decision threshold. To evaluate the advantage of combining features in an exclusive rather than inclusive way, ADJUST classification was compared to the one obtained by using OR-detectors, which identify an artifact when *any* of the associated features exceeds the threshold.

To evaluate the statistical significance of the difference, a paired t-test was computed between the accuracies of AND-detectors and OR-detectors for each type of artifact, and for general artifact detection.

ERP Accuracy from Artifact Corrected Data

The third step consisted of testing the efficiency of ADJUST in reconstructing artifact-free topographies of well known ERPs from sets of artifacted epochs. In order to have a reliable reference, ERPs computed from artifacted epochs after ADJUST correction were compared with artifact-free ERPs from the same subjects in the same experimental sessions. For this purpose, two sets of epochs were extracted for each dataset: 'Most Contaminated Epochs,' in which there was at least one channel exceeding 65 μV, and 'Least Contaminated Epochs,' selected as artifact-free epochs in the same number of the Most Contaminated Epochs.

Three different ERP topographies were computed by averaging the spatiotemporal ERP within an interval centered on the latency of the peak of the grand-averaged ERP:

Auditory N1: latency [90–120] ms after the auditory stimulus;
Visual P1: latency [80–110] ms after the visual stimulus;
Visual N1[2]: latency [160–190] ms after the visual stimulus.

These latencies are well matched with the typical latencies found in the literature (e.g., Hine & Debener, 2007, for the Auditory N1; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2001, for the visual P1 and N1). The distortions caused by artifacts on each ERP topography were quantified by the topography error ε, computed as the square root of the sum squared difference between the ERP topography calculated from the Least Contaminated Epochs and the one calculated from the Most Contaminated Epochs, normalized with respect to the square root of the sum squared amplitude of the Least Corrupted ERP:

$$\varepsilon = \sqrt{\sum_{scalp} (g_{LCE} - g_{MCE})^2} \Big/ \sqrt{\sum_{scalp} (g_{LCE})^2} \qquad (13)$$

[2]Posterior Visual N1.

where g_{LCE} (g_{MCE}) indicates the ERP topography map calculated from Least Corrupted Epochs (Most Corrupted Epochs), and the sum is computed over all channels of the scalp.

Normalization is performed in order to scale the difference by the size of the least polluted topography map; this was done because the degree of distortion introduced by artifacts depends on the magnitude of the "clean" ERP wave.

Results

EEG recordings from all datasets contained ocular artifacts and artifactual discontinuities in a variable amount across subjects. Several artifactual ICA components could be easily identified by visual inspection in all subjects. Figure 2 (top row) shows examples drawn from the validation set of artifact-specific IC topographies for each artifact class defined in the Methods sections (blinks, horizontal and vertical eye movements, discontinuities). In contrast, neural ICs typically display a smooth dipolar topography (see example of an IC representing a visual evoked potential in Figure 2, last column). The time course of artifactual signals is also archetypical: all artifacted ICs exhibit low-amplitude fluctuations interspersed with high-amplitude jumps occurring in only a few trials, represented by red or blue color spots in the ERP images of Figure 2 (second row). In contrast, the neural IC presents an activity distributed across all trials comprising a clear event-related potential, which is visible in most trials.

Once ICA weights are computed, ADJUST is very fast: it takes about 12 s to run the algorithm on a dataset of about 200 MB and display the classification results for further inspection on a standard PC (Microsoft Windows XP Professional SP3, Intel Core 2 Duo CPU E4600 @2.40 GHz, 2.00 GB RAM).

The joint use of spatial and temporal features revealed crucial for artifact identification: even though single features may not follow a bimodal distribution, the combination of more features together led to a cluster of artifact-specific ICs clearly separate from the rest of the ICs (see Figure 4 for an example).

ADJUST's artifact-specific detector selectivity is evident from the example of Figure 2 (bottom row): for each artifact-specific IC, all features associated to that artifact visibly cross the threshold (blue horizontal line), triggering the classification of that IC as artifacted (red bars); in contrast, for the detectors specific to the other types of artifacts there is always at least one feature that has a value lower than the threshold, so that ADJUST does not classify that IC as belonging to that artifact class (green bars). The neural IC is not classified as artifact because none of the artifact-specific groups of features crosses the threshold.

ADJUST Validation

As mentioned in the previous section, the effectiveness of ADJUST was assessed on a validation dataset completely different and uncorrelated with the feature selection dataset used to optimize ADJUST. The validation dataset is different from the feature selection datasets in the EEG recording system, in the laboratory in which it was recorded, and in the experimental paradigm (see Methods for details). This validation was performed by using the same spatial and temporal features that were optimized on the feature selection datasets.

Artifact Classification Accuracy

In the first validation step, ADJUST's IC classification was compared to manual IC classification made by three independent expert scorers (see Methods for details). Agreement between scorers' classification was high (95.3% on all artifacts). Among

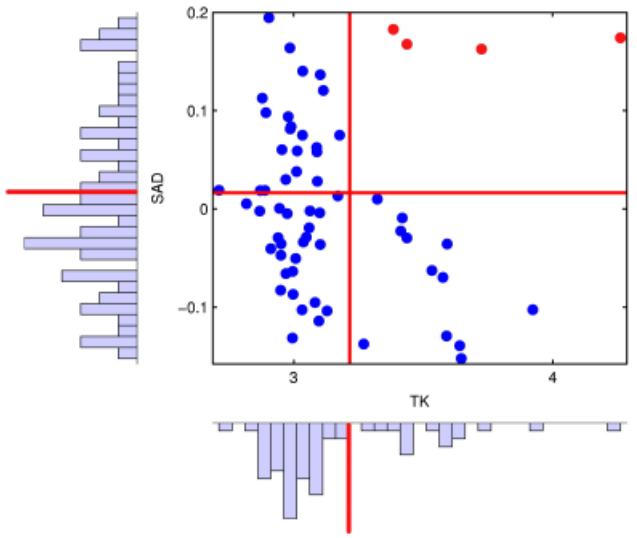

Figure 4. Scatter plot showing the values of the two features composing the Blink detector (Temporal Kurtosis on the *x*-axis and Spatial Average Difference on the *y*-axis) for the ICs belonging to one subject (validation dataset). Distributions of ICs values for each feature are superposed on the relative axis. Red lines indicate the thresholds computed automatically for each feature. Red points indicate the values of ICs marked as artifacted for the Blink detector.

the 630 ICs presented to the experts, 43 were classified as ocular artifacts, 74 as generic discontinuities, and the remaining as neural or low-amplitude noise components. Even though ocular artifact ICs were fewer than discontinuity ICs, they explained a much larger amount of the total data variance (51.5% vs 1.7% of discontinuity ICs), suggesting that most artifacts are captured by a few ICs associated with ocular movements.

ADJUST accuracy was computed for each artifact class and for the general artifact detection by means of the agreement measure defined in Equation (12) as the ratio between the variance accounted for by the ICs classified in the same way by ADJUST and the independent scorers and the variance of all ICs (see Methods). Accuracy relative to ocular artifacts was excellent: 99.0% for blinks, 96.0% for vertical eye movements, 99.2% for horizontal eye movements. Despite the more heterogeneous spatiotemporal features of generic discontinuities compared to ocular artifacts, the associated accuracy was also high (97.7%). Overall, the accuracy of the detection of any type of artifact (general artifact detection) was 95.2%.[3] This result is homogeneous across subjects (see black error bars for AND-detectors in Figure 5). Figure 6 summarizes the agreement between ADJUST classification and the manual classification for all artifacts. It is worth noting that false positive errors (components classified as artifact by ADJUST but considered non-artifact by manual classification) are very infrequent for both datasets, meaning that

[3]The agreement measure relative to the general artifact detection (i.e., computed on all types of artifacts) is generally lower than the one relative to a single type of artifact because the former comprises disagreements (false alarms and missed alarms) relative to all types of artifacts, while the latter is only penalized by disagreements relative to that type of artifact (see Equation 12). This effect is partially compensated by occasional artifact mislabeling (e.g., a blink labeled as a generic discontinuity by ADJUST), which affects the single artifact agreement measure but not the general artifact detection one.

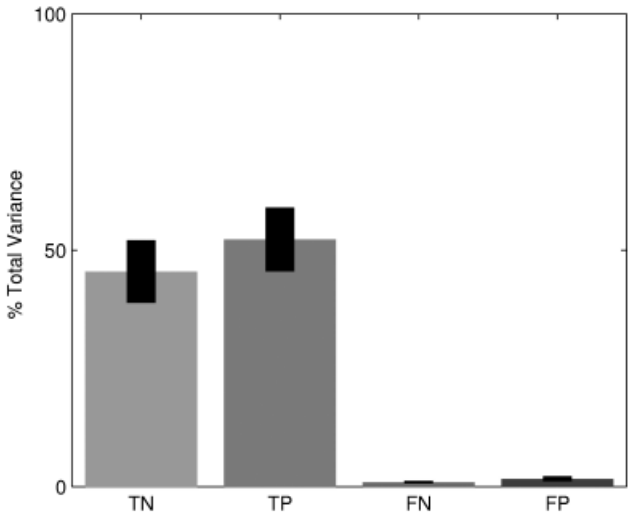

Figure 5. Comparison of classification accuracies obtained by AND-detectors (dark gray bars) and OR-detectors (light gray bars) for each artifact detector and for the general artifact detection for the validation dataset. Black error bars indicate the standard error of the mean.

the probability of removing a neural component is very small (2.5% of the total variance).

Effects of Combining Spatial and Temporal Features

In the second validation step, we evaluated the benefits of characterizing each artifact class by the combination of spatial and temporal features together with respect to using each feature as a separate artifact-specific detector. To this purpose, performance of ADJUST detectors (here indicated as AND-detectors because they classify an IC as artifacted only when all associated temporal and spatial features have a value higher than their respective threshold) was compared with that obtained by OR-

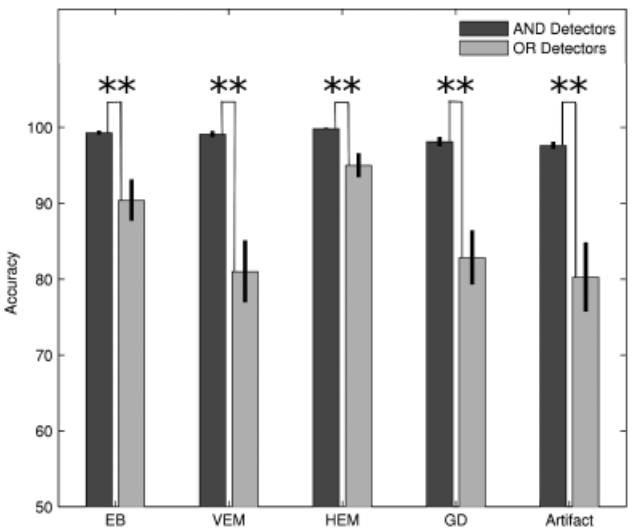

Figure 6. Classification performance of the general artifact detector (an IC is labeled as an artifact if marked as such by any of the four artifact detectors) compared to the one provided by three independent experts for the validation dataset. Bars represent the amount of True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) ICs weighted by the percent of total variance they account for, respectively. Black error bars represent the standard error of the mean.

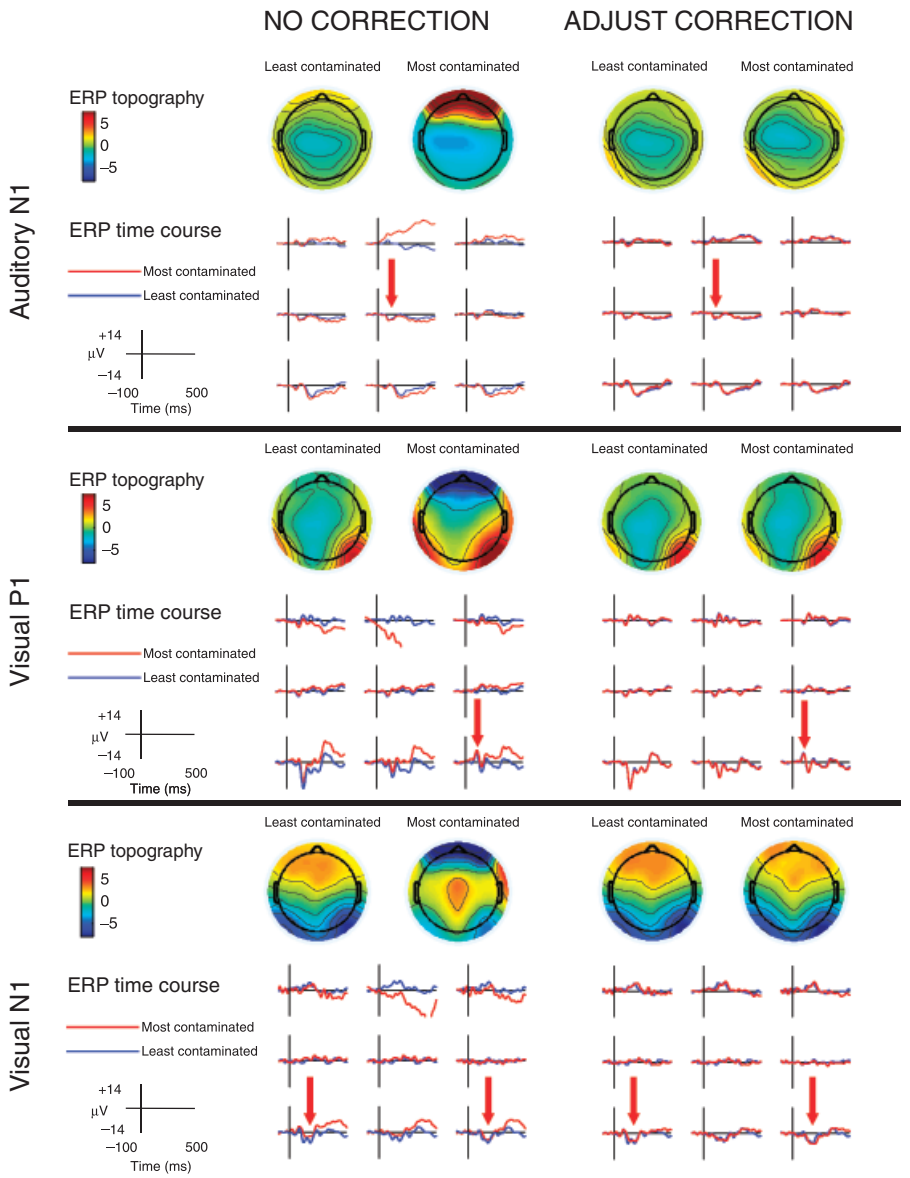

Figure 7. Examples of ERP reconstruction at the three selected latencies drawn from three representative subjects. Each panel shows: 1) ERP topographies at peak latency (top row) computed on Least Contaminated epochs (odd columns) and Most Contaminated epochs (even columns) with no correction (left-hand columns) and after ADJUST correction (right-hand columns); 2) ERP time courses (bottom row) at representative electrodes (F3, Fz, F4, C3, Cz, C4, O1, Oz, O2) before and after ADJUST correction averaged over Least Contaminated epochs (blue lines) and Most Contaminated epochs (red lines). Vertical red arrows mark the latency of the ERP components at the electrodes showing highest amplitude.

detectors, which classify an IC as artifacted when any of the associated features exceeds the threshold value. AND-detectors reach significantly higher accuracies than the respective OR-detectors for all types of artifacts and for the general artifact detector (for all t-tests, $t(9) > 19.68$, $p < .001$) (Figure 5). The performance of OR-detectors is sometimes near to that of AND-detectors for the most stereotyped artifact (e.g., horizontal eye movements), but drastically drops for more heterogeneous artifacts, resulting in accuracy for the general artifact detector of about 80%.

ERP Accuracy from Artifact Corrected Data

The last validation step consisted of testing the efficiency of ADJUST in reconstructing artifact-free topographies of well-known ERPs from sets of artifacted epochs. In order to have a reliable reference, ERPs computed from artifacted epochs after ADJUST correction were compared with artifact-free ERPs from the same subjects in the same experimental sessions. For this purpose, two sets of epochs were extracted for each dataset: 'Least Contaminated Epochs' were virtually artifact-free, while 'Most Contaminated Epochs' were the most contaminated by artifacts (see Methods for selection criteria). The average number of epochs in each set was 90 ± 12 (mean $\pm$ standard deviation).

ADJUST was evaluated on the efficiency of reconstruction from the Most Contaminated Epochs of the topographies of three different ERPs: auditory N1 (latency 100 ms), visual P1 (latency 100 ms), and visual N1 (latency 175 ms). As expected, artifacts considerably altered ERP topographies and time courses (Figure 7). Anterior electrodes were particularly affected,

suggesting that ocular artifacts were the major cause of alteration. ADJUST systematically removed the most important artifacts distortions, reconstructing an ERP from the Most Contaminated Epochs that almost overlaps to the one computed from the Least Contaminated Epochs, both in its topography and in its time course (Figure 7).

The amount of distortions caused by the residual artifacts on each ERP topography was quantified by the topography error ε (Equation 13). Topography errors were significantly lower for ADJUST than for uncorrected data for all examined ERPs (Wilcoxon signed rank test (Wilcoxon, 1945): $p = .0097$ for auditory N1, $p = .0019$ for visual P1, and $p = .0019$ for visual N1) (Figure 8). Topography error variability across subjects was also remarkably lower for ADJUST than for uncorrected data (error bars in Figure 8), suggesting that ADJUST typically provides a stable performance across subjects.

To further validate the performance of ADJUST, topography errors obtained using ADJUST were compared to those relative to manual classification. The difference between the topography errors from the two methods was not significant for all ERPs considered (Wilcoxon signed rank test (Wilcoxon, 1945): $p = .56$ for auditory N1, $p = .24$ for visual P1, and $p = .70$ for visual N1) (Figure 8), suggesting that ADJUST performance is equivalent to that of a manual classification by experts.

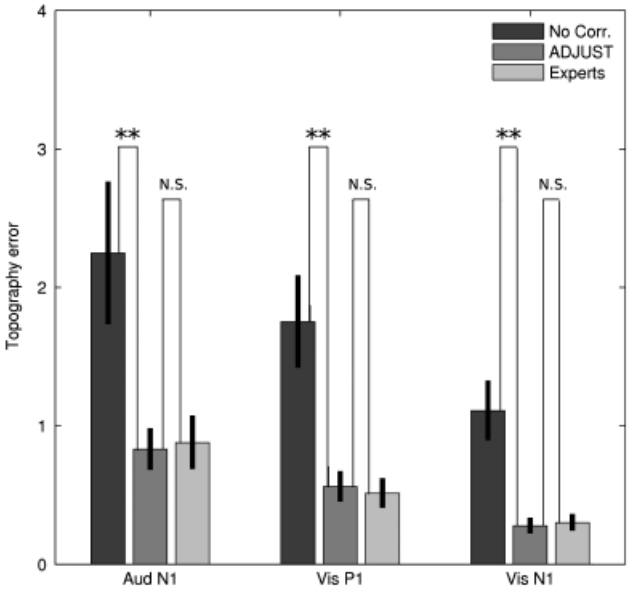

Figure 8. Mean topography error (Equation 13) of all subjects of the validation dataset for the three selected ERPs: Auditory N1, Visual P1, and Visual N1, for data with no correction (dark gray), data after ADJUST correction (mid-gray), and data after correction (light gray) by manual classification. Bars indicate standard deviation of the mean. Symbols over thin black lines indicate the results of a Wilcoxon signed rank test performed between topography errors relative to the conditions linked by the same lines. One asterisk (two asterisks) indicate $p < .05$ ($p < .01$), while N.S. indicates non-significant.

Discussion

In this paper, a completely automatic method for the detection of artifacted ICs from EEG data (ADJUST) has been proposed. The core property of ADJUST is the simultaneous use of multiple spatial and temporal features to detect the artifacted ICs. The key applicative aspect of ADJUST is its completely automatic nature: no trial-and-errors procedures are necessary for tuning parameters, as features are defined *a priori* and the algorithm that computes feature thresholds is completely unsupervised (Bruzzone & Prieto, 2000). The efficiency of ADJUST was demonstrated by a remarkable classification accuracy (95.2% for all artifacts) and by its ability to reconstruct clean ERP topographies from heavily artifacted data. In the following, we discuss these results in the context of the recent literature, and propose future extensions and applications of our method.

Automaticity and Feature Combination

As clearly described in Onton et al. (2006), EEG artifacts can be divided in two classes: non-stereotyped artifacts due to movements of the electrodes on the scalp arising from large muscle movements or external sources, and stereotyped artifacts, mainly due to ocular eye movements and blinks. Artifacts from the first class are problematic for ICA because, since their spatial distribution is extremely variable, they introduce a large number of unique scalp maps, leaving few ICs available for capturing brain sources. Accordingly, ADJUST does not attempt to remove these artifacts, and it relies on a suitable pre-processing for removing them before the ICA decomposition. However, stereotyped artifacts belonging to the second class also display a wide spatial and temporal heterogeneity that is testified by the wide range of measures that have been used to identify them: high-order statistics (Barbati et al., 2004; Delorme et al., 2007), entropy measures (Mantini et al., 2008), and spatial templates (Li et al., 2006; Viola et al., 2009). Here we have proposed an algorithm that identifies ICs due to stereotyped artifacts by computing several measures simultaneously. Its effectiveness is a demonstration that the diversity of EEG artifacts is limited, and can be fully captured by a reduced set of spatial and temporal features, provided that these are used together (see Figure 5). The efficiency of the automatic procedure implemented in ADJUST is based on this simple property: even though the distribution of single feature values does not clearly separate artifact from artifact-free components, this goal is achieved when combining more features together (Figure 4). Remarkably, this result is obtained with a simple thresholding algorithm (Bruzzone & Prieto, 2000) in a completely unsupervised way. This is important because most of the ICA-based artifact removal algorithms proposed in the literature have a supervised seed, either in the form of a training set (Delorme et al., 2007; Li et al., 2006; Mantini et al., 2008) or in the arbitrary tuning of the thresholds separating artifacted from non-artifacted components (Barbati et al., 2004; Joyce et al., 2004; Okada et al., 2007).

ADJUST Potential Extensions

The architecture of ADJUST (Figure 1) is intrinsically flexible: extensions to other types of artifacts may be implemented by building new detectors for those artifacts and adding them in parallel to the original ones. One type of artifact that may be included in the future is the one generated by muscle contractions, which may be identified by a spectral feature (Barbati et al., 2004; Delorme et al., 2007; Joyce et al., 2004). Another natural extension is to add features based on correlation with ECG or EOG signals (Joyce et al., 2004; Okada et al., 2007). In particular, EOG signals would improve the identification of blink ICs as they typically display a polarity flip across the eyes (Talsma & Woldorff, 2005), which would be easily captured by an *ad-hoc* feature. More generally, ADJUST architecture might

be used in the future for the classification of neural components from continuous EEG data, for example, by integrating it with algorithms of IC clustering as the one implemented in the EEGLAB software. This approach is potentially promising for studies focused on the relation between event-related and ongoing activity (Buiatti, 2008). Contrary to the case of artifacts, features expressing data regularity and stationarity might be chosen.

Potentially, ADJUST can be adapted to MEG data. Spatial filters for spatial features computation can be easily imported since they are based on scalp areas, which are identified by polar coordinates; sensors involved are automatically detected by inspecting channels coordinates.

Due to its ease of application for its automatic nature, absence of constraints on the experimental paradigm, and flexibility to new extensions, we believe that ADJUST is suitable for routine automatic artifact removal in research and clinical settings. Additional tests on populations prone to artifacts (like clinical data, or data on children) will help to further improve the method.

REFERENCES

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, *18*, 14–30.

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P. M., & Tecchio, F. (2004). Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clinical Neurophysiology*, *115*, 1220–1232.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1004–1034.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, *45*, 434–444.

Bruzzone, L., & Prieto, D. F. (2000). Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *38*, 1171–1182.

Buiatti, M. (2008). The correlated nature of large-scale neural activity unveiled by the resting brain. *Rivista Di Biologia-Biology Forum*, *101*, 353–373.

Croft, R., & Barry, R. (2000). Removal of ocular artifact from the EEG: A review. *Clinical Neurophysiology*, *30*, 5–19.

Croft, R., & Barry, R. (2002). Issues relating to the subtraction phase in EOG artifact correction of the EEG. *International Journal of Psychophysiology*, *44*, 187–195.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Delorme, A., Sejnowski, T. J., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*, 1443–1449.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111.

Gasser, T., Sroka, L., & Mocks, J. (1985). The transfer of EOG activity into the EEG for eyes open and closed. *Electroencephalography & Clinical Neurophysiology*, *61*, 181–193.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology*, *55*, 468–484.

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, *45*, 1199–1211.

Hine, J., & Debener, S. (2007). Late auditory evoked potentials asymmetry revisited. *Clinical Neurophysiology*, *118*, 1274–1285.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: John Wiley & Sons.

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, *41*, 313–325.

Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds), *Advances in Neural Information Processing Systems*, *10*, 894–900.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*, 163–178.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*, 1745–1758.

Kenemans, J. L., Molenaar, P. C. M., Verbaten, M. N., & Slangen, J. L. (1991). Removal of the ocular artifact from the EEG: A comparison of time and frequency domain methods with simulated and real data. *Psychophysiology*, *28*, 114–121.

Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and superGaussian sources. *Proceedings of the 4th Joint Symposium of Neural Computation*, *7*, 132–139.

Li, Y., Ma, Z., Lu, W., & Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiological Measurement*, *27*, 425–436.

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., & Ilmoniemi, R. J. (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *Journal of Neuroscience*, *21*, 1370–1377.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*, *16*, 47–62.

Mantini, D., Franciotti, R., Romani, G. L., & Pizzella, V. (2008). Improving MEG source localizations: An automated method for complete artifact removal based on independent component analysis. *NeuroImage*, *40*, 160–173.

Niedermeyer, E., & da Silva, F. H. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields* (Fifth edition). Hagerstown, MD: Lippincott Williams & Wilkins.

Okada, Y., Jung, J., & Kobayashi, T. (2007). An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements. *Physiological Measurements*, *28*, 1523–1532.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, *30*, 808–822.

Oster, P. J., & Stern, J. A. (1980). Measurement of eye movement electrooculography. In I. Matin & P. H. Venables (Eds.). *Techniques in Psychophysiology*, 275–309.

Peters, J. F. (1967). Surface electrical fields generated by eye movement and eye blink potentials over the scalp. *Journal of EEG Technology*, *7*, 27–40.

Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *Journal of Neuroscience*, *28*, 7585–7598.

Talsma, D., & Woldorff, M. G. (2005). Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 115–148). Cambridge, MA: MIT Press.

Vanhatalo, S., Palva, J. M., Holmes, M. D., Miller, J. W., Voipio, J., & Kaila, K. (2004). Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. *Proceedings of the National Academy of Sciences USA*, *101*, 5053–5057.

Verleger, R., Gasser, T., & Mocks, J. (1982). Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity. *Psychophysiology*, *19*, 472–480.

Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, *47*, 589–593.

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, *120*, 868–877.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80–83.

Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1983). The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, *16*, 127–147.

(Received March 30, 2009; Accepted March 29, 2010)