



Research article

A practical guide to the selection of independent components of the electroencephalogram for artifact correction



Maximilien Chaumon^{a,b,*}, Dorothy V.M. Bishop^c, Niko A. Busch^{a,b}

^a Berlin School of Mind and Brain, Luisenstraße 56, 10117 Berlin, Germany

^b Institute of Medical Psychology, Charité University Medicine, Luisenstraße 57, 10117 Berlin, Germany

^c Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

ARTICLE INFO

Article history:

Received 2 July 2014

Received in revised form 18 February 2015

Accepted 19 February 2015

Available online 16 March 2015

Keywords:

EEG

ICA

Artifact

Pre-processing

EEGLAB plugin

ABSTRACT

Background: Electroencephalographic data are easily contaminated by signals of non-neural origin. Independent component analysis (ICA) can help correct EEG data for such artifacts. Artifact independent components (ICs) can be identified by experts via visual inspection. But artifact features are sometimes ambiguous or difficult to notice, and even experts may disagree about how to categorise a particular component. It is therefore important to inform users on artifact properties, and give them the opportunity to intervene.

New Method: Here we first describe artifacts captured by ICA. We review current methods to automatically select artifactual components for rejection, and introduce the SASICA software, implementing several novel selection algorithms as well as two previously described automated methods (ADJUST, Mognon et al. *Psychophysiology* 2011;48(2):229; and FASTER, Nolan et al. *J Neurosci Methods* 2010;48(1):152).

Results: We evaluate these algorithms by comparing selections suggested by SASICA and other methods to manual rejections by experts. The results show that these methods can inform observers to improve rejections. However, no automated method can accurately isolate artifacts without supervision. The comprehensive and interactive plots produced by SASICA therefore constitute a helpful guide for human users for making final decisions.

Conclusions: Rejecting ICs before EEG data analysis unavoidably requires some level of supervision. SASICA offers observers detailed information to guide selection of artifact ICs. Because it uses quantitative parameters and thresholds, it improves objectivity and reproducibility in reporting pre-processing procedures. SASICA is also a didactic tool that allows users to quickly understand what signal features captured by ICs make them likely to reflect artifacts.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The electroencephalogram (EEG) recorded from electrodes placed on the scalp can provide information about underlying brain activity, but attempts to interpret the recorded signal are invariably hindered by the presence of artifacts, i.e. electrical signals of non-neural origin.

One major issue in interpreting scalp EEG is that the signal recorded at each electrode reflects a mixture of several sources of activity of various origin within and outside of the brain. A widely used method that allows one to isolate and subtract independent sources of activity is independent component analysis (ICA). This

method has been introduced to EEG analysis by Makeig et al. (1996), and popularized in the EEGLAB (Delorme and Makeig, 2004), a widely used software package running under MATLAB (The Mathworks). ICA allows isolation of statistically independent sources, called independent components (ICs) as linear combinations of electrodes. Each IC is characterized by a topography (set of inverse weights, describing the projection of the independent source onto the electrode cap), and a time course, which can be thought of as the signal that would have been recorded with an electrode located directly at that source. Because ICs are linear combinations of the original electrode signal, they can be treated in many respects like single electrodes. In particular, they can be subtracted easily from the signal just like one would discard a bad electrode after recording. After removal of a bad electrode, the signal is free of the artifacts that occurred at that electrode. Likewise, after subtraction of an artifactual IC, the remaining signal is free from artifacts that were captured entirely by that IC.

* Corresponding author at: Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Luisenstrasse 56, 10117 Berlin, Germany. Tel.: +49 30 2093 1794.

E-mail address: maximilien.chaumon@gmail.com (M. Chaumon).

This method of component subtraction is widely used to remove artifacts such as eye blinks or muscle activity from EEG recordings (e.g. Delorme et al., 2007; Jung et al., 2000a,b; Mantini et al., 2007; McMenamin et al., 2010; Urrestarazu et al., 2004). Some ICs capture a large amount of non-brain sources that recur in the signal, such as eye and muscle movements, heart beats, high impedance electrodes, or line noise (Jung et al., 2000a). However, although visualisations of IC activations and the effect of their subtraction make a compelling case for the usefulness of this approach in separating artifacts from neural signal, it is usually left to the user to scrutinise the ICA output and judge which ICs capture artifacts. Although selecting artifact ICs should be based on objective criteria, a comprehensive review of the signal features present in classical artifact types is to our knowledge missing in the literature. We will here define and illustrate precisely the features of the most common artifact types, and explain how these features are reflected in various statistical measures that can be computed on ICs, in order to provide investigators with a proper means of deciding which ICs capture artifacts and which ones do not.

The features of artifactual ICs can be visualized using various representations. EEGLAB offers a number of handy visual representations of IC properties that allow a trained observer to accurately identify artifactual ICs, but some features are not immediately obvious from these representations and time-consuming scrutiny and extensive experience is required. A number of automated procedures exist (e.g. Campos Viola et al., 2009; Delorme et al., 2007; Mognon et al., 2011; Nolan et al., 2010; Winkler et al., 2011) that compute objective statistical measures from ICs, and use these measures to automatically decide whether a component is artifactual or not. However, because of the high variability in EEG signals, these methods are inevitably prone to type I and type II errors. Furthermore, although some artifacts are unequivocally considered a nuisance (e.g. badly connected electrode noise), and have to be removed from the signal before analysis, others may be more controversial (e.g. Olbrich et al., 2011), and not every experimenter may want to discard them. We thus promote here an intermediate method, using the objective measures computed by several methods and enhanced EEGLAB visual representations to allow users to decide whether or not individual ICs reflect artifacts and need to be removed from the data or not.

In this paper, we first describe the relevant signal features related to the most common EEG artifacts – ocular artifacts, tonic muscle artifacts, loose electrode connections (high impedance), and exceptional high amplitude events – and show how these features can be mapped onto a number of visually recognizable attributes in visual representations of the signal and on objective statistical features of the signal. Some of these measures have been introduced before in plugins for EEGLAB (ADJUST Mognon et al., 2011; and FASTER Nolan et al., 2010). Second, we introduce the SASICA plugin (Semi-Automated Selection of Independent Components of the electroencephalogram for Artifact correction) for EEGLAB that provides a convenient visualization of all of these measures and allows refining selections manually if needed (Fig. 1). We thereby provide the user with all required information for understanding the reasons why a given component might be removed from the data. Finally, we evaluate all methods against expert classifications for a total of 21 experimental datasets, and illustrate the impact of (in)appropriately identifying and removing artifactual components on signal quality.

2. Methods

2.1. Signals captured by ICA

Several categories of signals are readily isolated by single ICs. Specifically, ICs can capture (1) a source of neural activity, (2)

variations of potential due to blinks, (3) eye movements (saccades), (4) muscle contraction, or (5) line noise or a misconnected (high impedance) electrode, commonly referred to as a “bad channel”.

Importantly, ICA may also fail to separate signals, and many components (often a majority) do not fit a single category. In essence, separating distinct classes of ICs is thus a signal detection problem in which the experimenter needs to avoid two mistakes: missing to-be-detected artifact ICs (type II error) and falsely reporting other non-artifactual ICs (type I error). In the context of artifact correction, the former mistake would imply under-correction while the latter would imply over-correction. Another challenge is to solve this task using objective criteria that can be readily communicated, for example in publications.

All automated methods reviewed here have their own heuristic to identify at least some of these ICs. In the following, we describe all the features of each category of IC, as well as a number of statistical measures designed to reveal these features. We include measures computed by SASICA, CORRMAP (Campos Viola et al., 2009), ADJUST (Mognon et al., 2011), and FASTER (Nolan et al., 2010). We present a summary of all measures offered by these methods in Table 1. We refer the interested reader to the original papers for details on each method.

2.1.1. Neural activity

The success of ICA in EEG analysis is largely due to the plausibility of the solution returned by ICA. Indeed, in most cases, when performed on a full-rank long enough dataset, the topography and time course of at least a handful of components compellingly allow identifying them as capturing selective neural activity. These components are often dipolar, i.e. they are well modeled by one, or sometimes two, dipolar sources (Delorme et al., 2012), and their topography is regular and smooth. Moreover, they often rank amongst the strongest components in the dataset (i.e. those explaining most variance in the signal, and sorted first in EEGLAB), they often contain a peak at physiological frequencies (e.g. alpha, beta, delta or theta), and may show a strong evoked response to sensory stimuli. These properties are listed in Fig. 2A for reference.

The dipolar nature of the components can be measured by first fitting a dipolar source to the component (as implemented in the DIPFIT toolbox distributed with EEGLAB; applied to all components of all datasets tested in this article), and then measuring the residual variance after removing the fitted data. Residual variance is often very low for accurately modeled components (see results, Fig. 2B–D). Therefore, this measure is used routinely within EEGLAB to select neural components for analyses conducted on component time courses. However, it should be noted that some components with low residual variance may be artifactual. For instance blink or saccade components can be very well modeled by dipoles placed in the eyes of the subjects (see Fig. 3A, 5% residual variance of a dipole fit, see Section 2.2.2.5 on residual variance for explanation). Some pure tonic muscle components may also be well modeled by a dipole placed close to the scalp, where muscular activity arises (e.g. Fig. 4B, 9% residual variance). Furthermore, several spatially separated sources of neural activity working in synchrony will not be well modeled by a dipole (e.g. Fig. 2E, 31% residual).

It is often the case that components neatly isolating neural activity rank amongst the first twenty components in a dataset. This feature is an empirical observation that has to our knowledge not been measured so far. In the 8 training datasets used in this article, 50% of the components rated as neural by the experts ranked amongst the 13% largest components. Nevertheless, artifacts can also be of strong amplitude (e.g. blinks), so this feature may not be discriminant for deciding whether a component is neural or artifactual.

ICs capturing neural activity often contain a peak in the Alpha (8–12 Hz, Fig. 2B), Beta (15–30 Hz, Fig. 2C), delta (1–4 Hz), or Theta

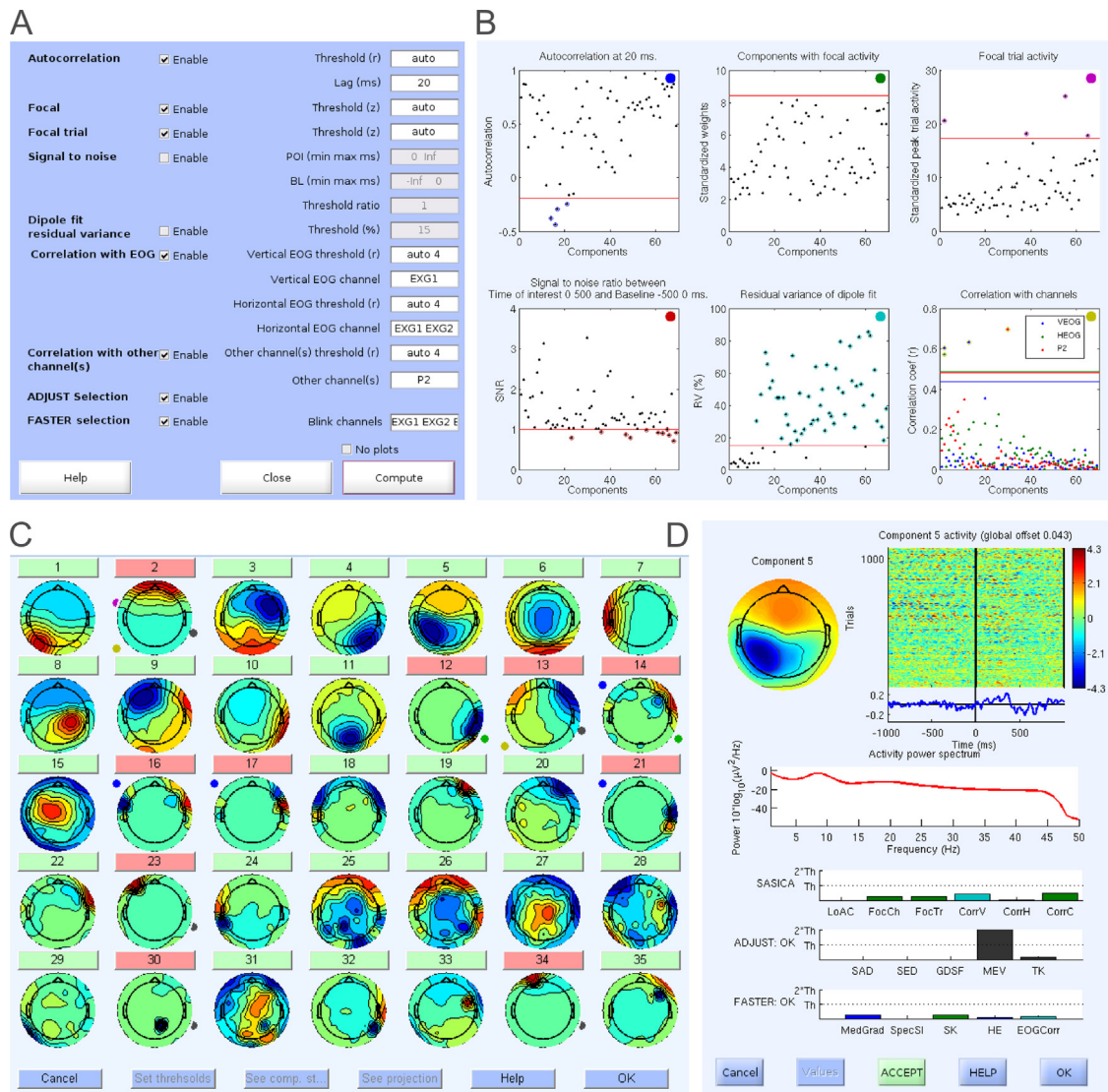


Fig. 1. Graphical user interface. (A) The main graphical interface window allows choosing which methods to use to select components. When pressing the “Compute” button, the plugin computes all enabled methods on the currently loaded EEG dataset and displays the results in separate windows. (B) The first window shows the value of the computed measures (y axis) for all components (x axis). The threshold for selection is shown as a horizontal red line in each panel, and every component that passes the threshold (above or below, according to the measure considered, see text for details) is highlighted in a color specific to the measure being considered, reproduced in the top-right corner of the panel, and used in the window shown in D. A mouse click on any point opens a detailed component properties window. Note that Signal to noise ratio and residual variance are shown in this window here for illustration but were not used to select components in the window shown in (D). (C) In this window, all component topographies are shown, along with a colored button indicating whether a given component is selected for rejection (red) or not (green). Next to the selected components’ topographies, colored dots indicate which computed measure passed threshold. (D) All component properties and measures can be summarized in individual windows, along with the classical EEGLAB plots (topography, single trial time course and power spectrum). All measures are scaled so that larger bars mean that a measure is more likely to select the component for rejection (see Section 2.2.2 for details).

(~5 Hz) frequency range. This is particularly true of components whose topography loads mostly on posterior, middle, or frontal sensors for Alpha, Beta, and Theta frequencies, respectively. This feature may nevertheless also not be diagnostic for neural activity in and of itself because some neural components may be devoid of a prominent peak in these physiological bands (e.g. Fig. 2D).

Finally another feature of neural ICs is a tendency to show strong evoked responses to sensory stimuli (Fig. 2B–D). However, this feature is not diagnostic alone either, because not all tasks involve sensory stimulation, and not all components showing an evoked response can be classified as reflecting pure neural activity (e.g. ambiguous components may capture some evoked activity, Fig. 4F). In some situations, artifacts can also occur in an event related manner (e.g. electrical artifact due to button press, Fig. 5C). Nevertheless, a measure of the ratio in power between prestimulus and poststimulus activity may help in identifying neural components. A

measure of this type is implemented in SASICA (see Sections 2.2.2.4 and 3.9.1).

In sum, although there is ample information in the signal that a trained observer could use to identify most ICs capturing neural activity, this information is scattered across multiple features of the signal, which individually do not unequivocally allow identification. This reason, and because some ICs returned by ICA are inherently ambiguous make it important to identify artifactual ICs.

2.1.2. Blink components

Components capturing blink activity are the easiest components to identify. Their topography is essentially flat (i.e. inverse weights close to zero) at all but a few frontal and all EOG electrodes. Activity is usually very large during blinks and the components rank amongst the first dozen components because blinks generate artifacts of extreme amplitude. Time courses show abrupt

Table 1
Measures computed by the three automated tools evaluated here. Abbreviations refer to those used in figures and throughout the paper.

Tool	Artifact type	Measure	Abbreviation
SASICA	Blinks/vertical eye movements	Correlation with vertical EOG electrodes	CorrV
	Horizontal eye movements	Correlation with horizontal EOG electrodes	CorrH
	Muscle	Low autocorrelation of time-course	LoAC or AutoCorr
	Bad channel	Focal channel topography	FocCh
	Rare event	Focal trial activity	FocTr
	Non dipolar component	Residual variance	ResVar
	Bad channel	Correlation with Bad channel	CorrCh
	Eye blinks/saccades	Correlation with EOG electrodes	EOGcorr
FASTER	“Pop-Off”	Spatial Kurtosis	SK
	White noise	Slope of the power spectrum	SpecSI
	White noise	Hurst exponent	HE
	White noise	Median slope of time-course	MedGrad
ADJUST	Eye blinks	Temporal Kurtosis	TK
	Eye blinks	Spatial average difference	SAD
	Eye blinks	Spatial variance difference	SVD
	Vertical Eye Movements	Maximum epoch variance	MEV
	Horizontal Eye Movements	Spatial eye difference	SED
	Generic Discontinuities	Generic discontinuity spatial feature	GDSF

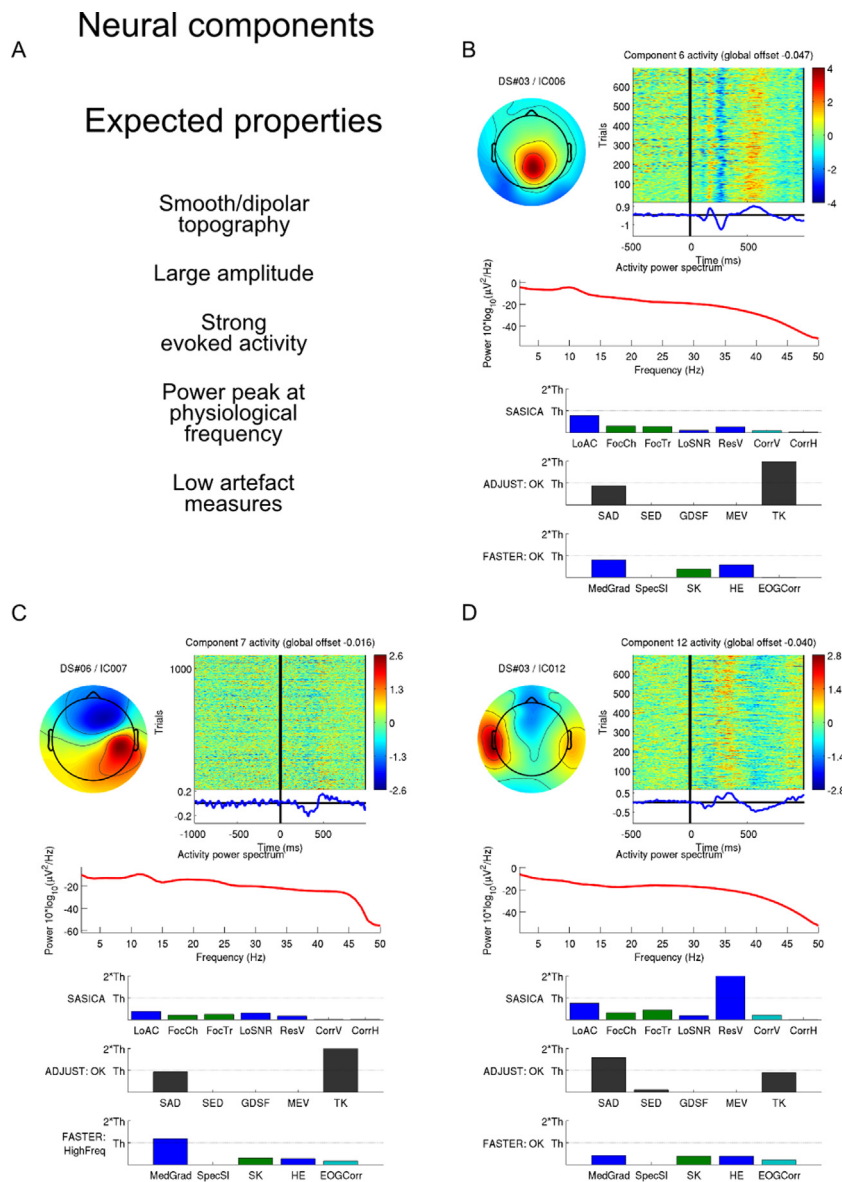
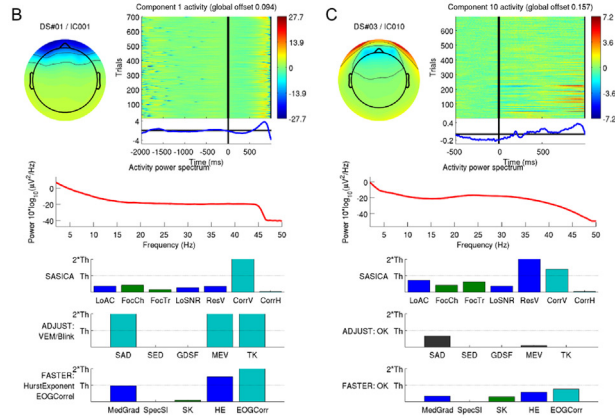


Fig. 2. Neural components. Three example neural components and their properties, as shown by SASICA. (A) Properties to pay particular attention to in order to determine if a component captures Neural activity. None of these properties should be met for a component to be considered as isolating Neural activity. (B and C) Two exemplar neural components, showing all of the properties listed in (A). (D) Neural component with non-dipolar topography, where the Residual Variance (ResV) measure passed threshold. Abbreviations for all measures are listed in Table 1.

Blink components

A Expected properties

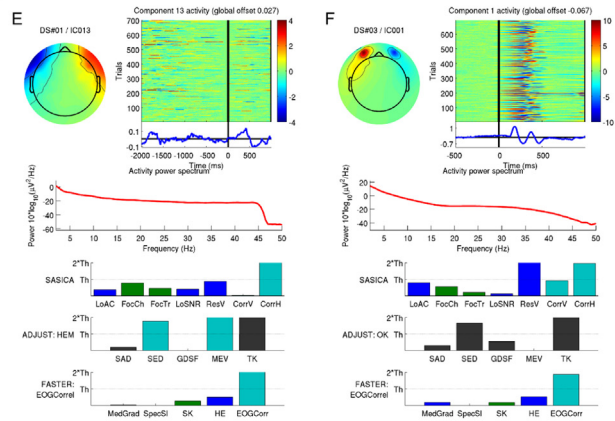
- Frontal topography
- Large amplitude
- Opposite polarity below the eyes
- No peak at physiological frequencies
- High correlation with vertical EOGs
- High eye movement related measures



Horizontal eye movement components

D Expected properties

- Opposite sign bilateral frontal topography
- Step-like events
- Opposite polarity around the eyes
- No peak at physiological frequencies
- High correlation with vertical/horizontal EOGs
- High eye movement related measures



Non-artifact components may be mistaken for ocular components

G Expected properties

- Inverse weight at posterior channels
- Noisy time course
- No opposite polarity around the eyes
- Weak correlation with EOGs

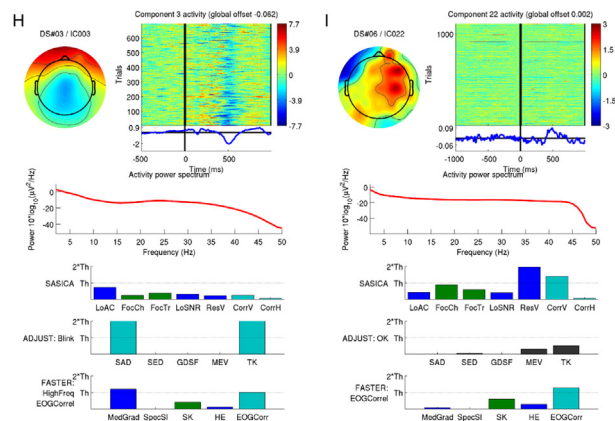


Fig. 3. Ocular components. Exemplar ocular components and their properties, as shown by SASICA. (A) Properties to pay particular attention to in order to determine if a component captures blink activity. (B and C) Two exemplar blink components, where measures designed to identify ocular components (cyan bars) passed threshold, and showing all properties listed in (A). (D) Properties to pay particular attention to in order to determine if a component captures horizontal eye movements. (E and F) Two exemplar horizontal eye movement components, where measures designed to identify ocular components (cyan bars) passed threshold, and showing all of the properties listed in (D). In panels (B), (C), (E) and (F), two situations are depicted, in which EOG electrodes are rendered on the topographical maps (C and F), or not (B and E). (G) Properties that can be found in non-artifact components that may be mistaken for ocular components. (H) Component mistaken for a blink component due to large inverse weights at frontal electrodes (see Section 3.3 and Fig. 8 for reasons why this is not an ocular component). (I) Component whose high correlation with horizontal EOGs induced erroneous selection by SASICA and FASTER (see Section 3.4 for reasons why this is not an ocular component). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

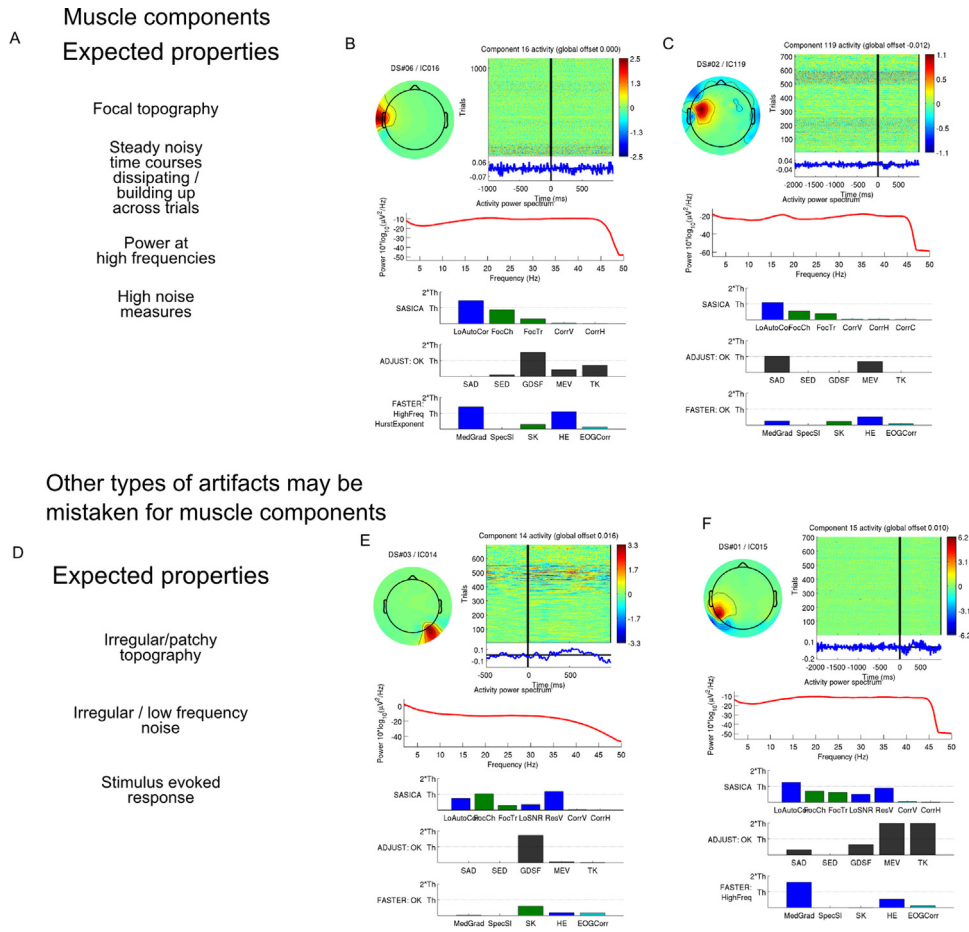


Fig. 4. Muscle components. Example muscle components and their properties, as shown by SASICA. (A) Properties to pay particular attention to in order to determine if a component captures muscle activity. (B and C) Two exemplar muscle components, where some measures designed to identify noisy components (blue) passed threshold, and showing all of the properties listed in (A). (D) Properties that can be found in non-muscle components mistaken for muscle components. (E) Example of component that in spite of a focal topography, fails to qualify as muscle component because it does not show the expected steady noise activity characteristic of muscle components. (F) Some components reflect mixtures of signals. This component captures at the same time a noisy and very focal activity pattern with low autocorrelation and median gradient measure, characteristic of muscle activity, but also some very high brief events and evoked activity that led the experts to categorize it as capturing rare events.

high amplitude variations in otherwise comparatively close to zero amplitudes, and power spectra show no power peak at physiological frequencies. Correlation with EOG electrodes, if available, is high. These properties are listed in Fig. 3A for reference.

Fig. 3B and C illustrates blink components showing all the features listed in Fig. 3A. In Fig. 3B, infra-ocular EOG electrodes, although present in the dataset are not rendered on the topography (i.e. their spatial coordinates are not registered in the dataset). In Fig. 3C on the other hand, EOG electrodes are rendered on topographical maps (i.e. their spatial coordinates are registered in the dataset). The topography reveals an abrupt polarity reversal at the frontmost locations where the two EOG electrodes are rendered.

Blink components can be identified automatically by template matching with a stereotypical activity pattern, as implemented in CORRMAP. ADJUST combines several spatial and temporal features. Spatial Average Difference (SAD) and Spatial Variance Difference (SVD) capture components with strong differences in signal between anterior and posterior channels, and Temporal Kurtosis (TK) indicates the occurrence of rare high amplitude events (i.e. heavy tailed distribution). Another straightforward measure is to correlate the time course of the ICs with EOG electrodes (as implemented in FASTER and SASICA). Each of these measures taken separately may not enable detection of all blink components, but their supervised combination can bring a trained observer close to perfect detection (see Results).

2.1.3. Saccade components

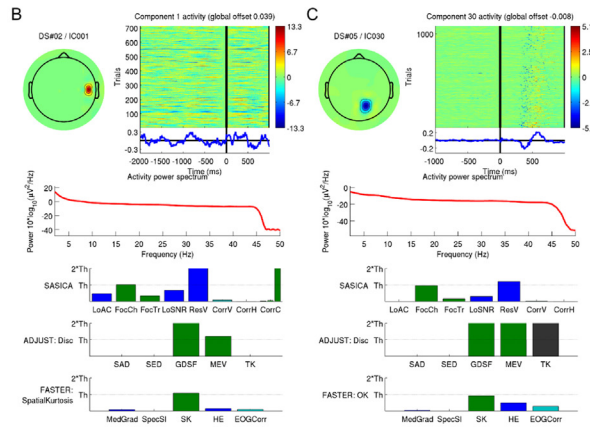
Components capturing horizontal saccade activity load maximally onto anterior electrodes, but with opposite polarity on both sides. Vertical saccades load maximally onto anterior sites, with a topography similar to that of blink components. Time courses show abrupt step-like variations and power spectra show no power peak at physiological frequencies. Correlation with EOG electrodes, if available, is high. These properties are listed in Fig. 3D for reference.

Fig. 3E and F illustrates saccade components showing all the features listed in Fig. 3D. Similarly to blink components, the displayed topography of saccade components depends on whether or not EOG electrodes are included and rendered in topographical plots. Fig. 3E shows the case where EOG electrodes are not rendered in the topographical maps, and Fig. 3F shows a case where four electrodes placed under and on both sides of the eyes (next to the lateral canthi) are rendered.

These components are identified automatically by template matching with a stereotypical activity pattern (implemented in CORRMAP). ADJUST detects vertical and horizontal eye movements separately. For vertical eye movements, it combines a Maximum Epoch Variance (MEV) measure that captures components with strong within epoch variability with the same SAD measure as for blink components. For horizontal eye movements, it uses MEV along with a Spatial Eye Difference (SED) measure to capture strong differences between two lateral regions of the EEG cap. FASTER and

Bad Channel components

- A**
- Expected properties
- Focal (one channel) topography
 - Noisy time course
 - High correlation with marked bad channel
 - High spatial / intertrial noise measures



Ambiguous mixture components

- D**
- Expected properties
- More spread-out topography
 - Stimulus evoked response
 - Transient noise activity

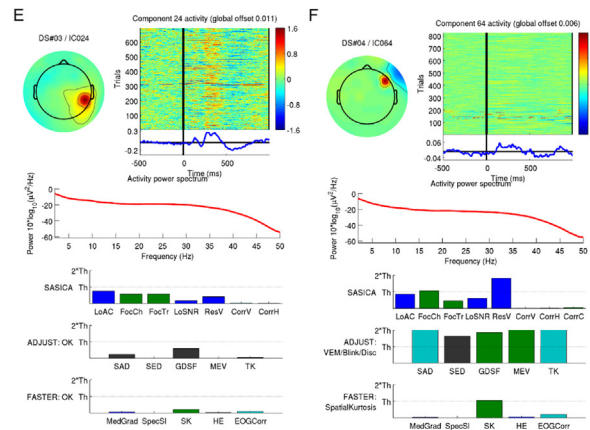


Fig. 5. *Bad channel components.* Example bad channel components and their properties, as shown by SASICA. (A) Properties to pay particular attention to in order to determine if a component captures activity of a bad channel. (B and C) Two exemplar bad channel components, where measures designed to identify isolated noise and discontinuities (green) passed threshold, and showing all of the properties listed in (A). (D) Properties that can be found in ambiguous components mistaken for bad channel components. (E) Example component with a smooth (although focal) topography, and a clear evoked response, mistaken by some users for a bad channel component. (F) Example component illustrating the overlap between the Bad channel and the Rare events component categories. This component captures activity of one bad channel that occurred mostly during a few trials, leading to ambiguous classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SASICA use correlation of component time courses with EOG electrodes (EOGCorr, CorrV and CorrH) to detect eye movements and blink components.

2.1.4. Muscle components

Tonic muscle activity arising from neck, jaw, and face muscles produces a stereotypical activity at electrodes at the edge of the electrode cap. Although subjects are usually asked to sit still and relax, uncontrollable postural activity, as well as muscular activity due for instance to yawning or swallowing may occur and be captured in the EEG. Components capturing muscle activity are usually very focal, encompassing a local group of electrodes (sometimes with opposite polarity) on the edge of the electrode cap. Time courses show a steady noise activity, often remarkable because they do not vary with task events (i.e. no ERP is visible), but rather across trials. Postural muscles may indeed relax when the subject finds a comfortable posture (Fig. 4B), or appear temporarily during the experiment (Fig. 4C). The power spectrum of these components often shows strong power at high frequencies (>20 Hz). These properties are listed in Fig. 4A for reference.

Muscle components can be detected automatically because their time course reveals noise patterns and their topography is focused

on electrodes around the edge of the electrode cap. This can be detected by measuring the high time-point by time-point variability, captured by the low autocorrelation (LoAC) measure of SASICA, or by high Median Gradient (MedGrad) value, or low Hurst Exponent (HE) computed by FASTER. ADJUST and CORRMAP do not attempt to detect muscle components specifically.

2.1.5. Bad channel

When a bad channel shows strong amplitudes, uncorrelated with other channels, it is readily isolated by ICA in a single component. Such bad channel components have a focal topography, restricted to the bad channel, and their time course reflects the noisy nature of the recording. They may also show a very high level of correlation with marked bad channels. These properties are listed in Fig. 5A for reference.

Fig. 5B and C illustrates two exemplar components showing all the features listed in Fig. 5A. In addition, as noted before, the component illustrated in Fig. 5C shows a strong event related response that corresponded to an artifact generated by response button press.

Bad channels can be detected automatically because of their Focal Channel topography (FocCh) in SASICA, their high Spatial Kurtosis (SK) in FASTER, or by the Generic Discontinuity Spatial

Feature (GDSF) and MEV in ADJUST. Also, components capturing an isolated bad channel correlate by definition very highly with data recorded at that channel, so correlation of ICs with the bad channel in question allows identifying these ICs in SASICA (CorrC, Fig. 5B).

2.1.6. Rare events

In some cases, a few high-amplitude events may also be isolated with ICA. If the events in question occur at only one electrode, the associated IC topography is focal and the IC in question thus also qualifies for the bad channel category described above (Fig. 6B). If they occur at many electrodes (Fig. 6C), for instance if the subject moves or touches the electrode cap, the topography is less predictable. These properties are listed in Fig. 6A for reference. Note that if a component captures a unique event, a sensible strategy to remove the corresponding artifact from the data could be to reject the affected trial(s) and to compute the ICA anew.

Rare events are detected in SASICA with the Focal Trials (FocTr) measure, in ADJUST with GDSF and MEV, and in FASTER with SK.

2.1.7. Ambiguous components

Finally, it is important to keep in mind that not all ICs may be neatly and unequivocally classified as neural or artifactual. Rather, some components reflect an ambiguous mixture of signals, and should be handled with care. We do not recommend systematically rejecting such components, since part of their signal may be of neural origin. Exemplar mixture components that need special attention are illustrated in the last panel of Figs. 3–5. We will comment further on these components in Section 3.

2.2. SASICA

SASICA computes a number of measures on IC topographies and time courses and marks components for rejection. The plugin can be downloaded from <https://github.com/dnacombio/SASICA> and installed following the instructions on the EEGLAB website (<http://sccn.ucsd.edu/wiki/EEGLAB.Plugins>).

2.2.1. Graphical user interface

The graphical user interface is shown in Fig. 1. Each measure can be enabled or disabled, and thresholds may be adapted according to the requirements of a particular dataset or study (Fig. 1A). After computation, each measure or method is characterized by a unique color that is used to mark ICs identified as artifactual in other figures (Fig. 1B and C). A comprehensive display of component properties (Figs. 1D and 2–6) is invoked by clicking any component in these figures.

2.2.2. Measures computed from components

Below we introduce the 6 measures that SASICA computes from components and the rationale behind their use for selecting artifact components.

For all measures we use the following conventions:

- $n = 1, 2, \dots, N$ a specific channel (N being the number of channels in the dataset)
- $c = 1, 2, \dots, C$ a specific component (C being the number of components in the dataset)
- $k = 1, 2, \dots, K$ a specific trial (K being the number of trials in the dataset)
- $t = 1, 2, \dots, T$ a specific time sample (T being the number of time points in each trial)
- $x(n, k, t)$ or $x(c, k, t)$: EEG data at channel n or component c , on trial k and at time t
- $W_c(n)$: inverse weight of channel n in component c

- $Z_J(x)$: the z-score of x along dimension J (channels, time or trials), i.e. mean subtracted and divided by standard deviation across J elements of x

For each measure presented below, a threshold for rejection has to be set. SASICA allows setting either an absolute threshold entered by the user, or an adaptive threshold that selects components whose value on a given measure is beyond a number of standard deviations away from the average of all components for the current dataset. For most measures, we set the default threshold to 2 standard deviations. This adaptive threshold helps accounting for the fact that variable ranges of measures occur across different datasets, as illustrated in Fig. 7. Indeed, subtle differences in preprocessing and the amount of available data in a given experiment lead to large differences in measures computed on ICs. For channel correlation measures (vertical and horizontal EOG, and designated bad channels), we recommend using a more conservative threshold of 4 standard deviations from the mean. Indeed, the vast majority of components is very weakly correlated with specific channels, and most components correlate with r values close to zero. This is shown in Fig. 7 in the two top left panels, “CorrV” and “CorrH” for SASICA-computed correlation with vertical and horizontal EOG channels and bottom right panel “EOGCorr” for the FASTER-computed maximal correlation with any EOG channel. Two standard deviations away from average in these measures is still a fairly low value (around 0.2–0.25 in the training datasets), for which correlation is still rather unspecific to ocular artifacts.

Please note that in the plots in Figs. 1–6 (and in general in property plots produced by SASICA), all measures are scaled so that higher bars mean that a component is more likely to be selected by a given measure. For the autocorrelation, and signal to noise ratio measures, the scales are inverted because lower values mean that a given component is more likely to be rejected. Higher bars for these two measures thus mean lower values of Autocorrelation or SNR, hence the LoAC and LoSNR abbreviations used in figures.

2.2.2.1. Autocorrelation. Components reflecting brain activity are usually strongly autocorrelated. This means that the level of signal in a component at any time point usually correlates with the signal of this same component a few ms before. To the contrary, noisy components like muscle components tend to show low autocorrelation. This measure computes the autocorrelation of each component at the specified lag in ms and suggests rejection if the autocorrelation value is below the specified threshold. Interestingly, the same measure is also used in a recently developed alternative method to ICA for artifact correction using canonical correlation analysis to detect muscle components and correct for muscle artifacts (Clercq et al., 2006; Vos et al., 2010).

Autocorrelation is defined at lag l for component c by:

$$A_c = \sum_{t=l}^T x_c(t) \times x_c(t-l)$$

Default lag is set to 20 ms, which showed the best match with expert classifications in the training datasets. This measure serves a similar goal as the Hurst Exponent (HE) measure of FASTER (see Section 2.3).

2.2.2.2. Focal topography. Components reflecting brain activity rarely affect only one electrode. Components that load mostly onto one electrode are thus likely to reflect artifacts (bad channel or rare events), rather than brain activity. We measure focality of the topographies by computing the z-score of the ICA inverse weights across channels. Components that have a channel with its maximum absolute weight above threshold are considered focal.

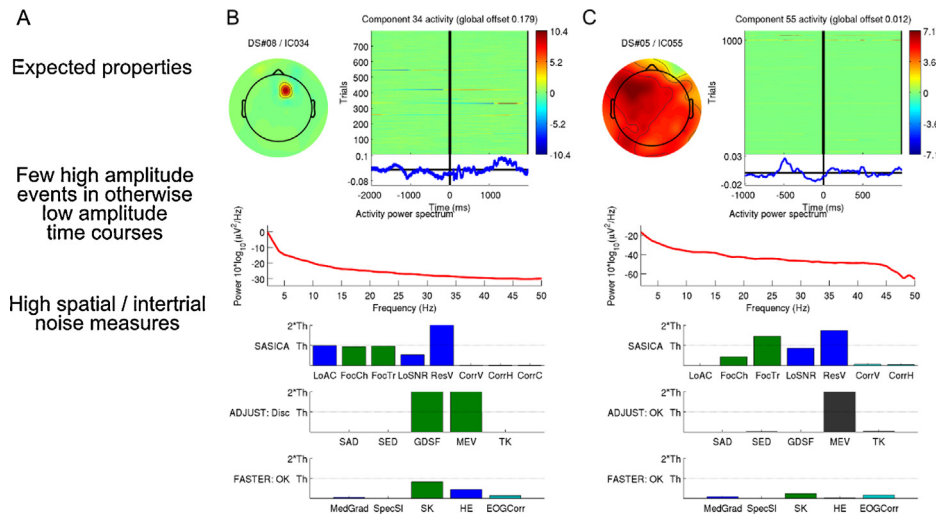


Fig. 6. Rare events. Example components capturing rare events and their properties, as shown by SASICA. (A) Properties to pay particular attention to in order to determine if a component captures activity of a rare event. (B) Example rare event component, where the event occurred at one electrode. This component thus qualifies for the Bad channel component category as well. (C) Example rare event component, where the event occurred at many electrodes.

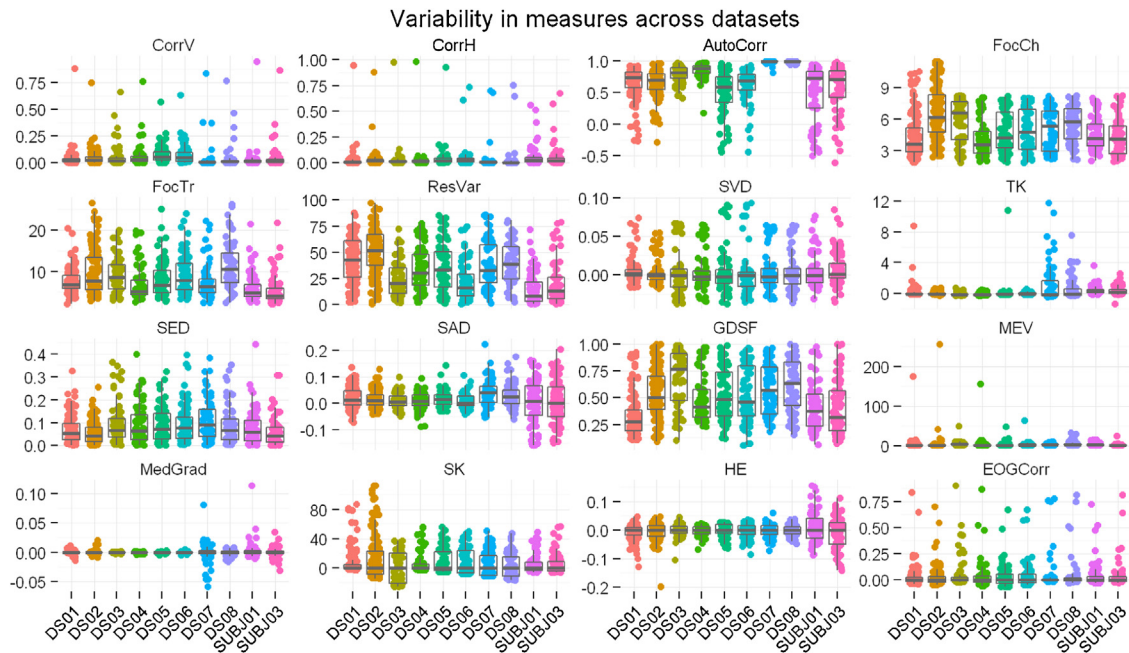


Fig. 7. Measure ranges vary with recording and preprocessing settings.

Each panel shows one measure for individual components of the 8 training datasets, and two exemplar test datasets. Each dot represents one component. Variability within datasets is represented with overlaid boxplots. Variability across datasets is most striking across different experiments. Note that we show SUBJ 1 and SUBJ 3 in the testing dataset in order to show variability of horizontal EOG (CorrH) measure, which SUBJ 2 does not have because the electrode is absent from that dataset. Abbreviations in panel titles correspond to each measure and are listed in Table 1.

Focal measure of component c :

$$F_c = \max_n \left(Z \left(W_c(n) \right) \right)$$

where $\max_n(\cdot)$ denotes maximum across channels. This measure serves a similar goal as the Spatial Kurtosis (SK) measure of FASTER, and the Generic Discontinuity Spatial Feature (GDSF) of ADJUST, (see Section 2.3).

2.2.2.3. Focal trial activity. Artifacts can occur with extremely large amplitude but on rare occasions. Thus, the same strategy as for the focal topography measure above can be applied to detect rare events, but instead of computing the z-score of ICA inverse weights

across channels, we compute the z-score of the range of ICs' activity across trials (the range is defined here as the amplitude difference between the maximum and minimum points for a given trial). Components that have trials above threshold are considered to reflect focal trial activity.

Focal trial activity of component c :

$$FT_c = \max_k \left(Z \left(\max_t (x(c, k, t)) - \min_t (x(c, k, t)) \right) \right)$$

where $\max_k(\cdot)$ and $\max_t(\cdot)$ denotes maximum across trials and time points, respectively. This measure serves a similar goal as the

Maximum Epoch Variance (MEV), and the Temporal Kurtosis measures of ADJUST (see Section 2.3).

2.2.2.4. Correlation with channels. Channels strongly contaminated by artifacts can often be identified early, either by design (EOG, electromyogram or electrocardiogram channels) or during recording and preprocessing of the data (channels with strong electrical artifacts due to misconnection or line noise). Often, these artifacts are captured in a single component of the ICA solution and this component is highly correlated with the channel in question. We measure the correlation of specific channels (EOGs or any other channel) with all components and set a threshold for rejection. Using correlation as a way to detect components contaminated with EOG activity has been used previously (e.g. Joyce et al., 2004; Okada et al., 2007) and a measure of maximal correlation with EOGs is implemented in FASTER (see below).

SASICA can take three types of channels to compute correlation with ICs. It has two fields for horizontal and vertical EOG, as well as one field for other “bad channels”. If two channels are entered for either vertical or horizontal EOGs, the difference of EOG channels is automatically computed. This is meant to increase the signal to noise ratio of the EOG signal before correlating it with ICs (see Section 3).

2.2.2.5. Additional measures. We provide here additional measures that could be used flexibly to select components, either for rejection, or for further processing. We do not recommend using these measures to routinely reject artifact ICs.

2.2.2.5.1. Weak signal to noise ratio. In event related potential (ERP) studies, it may be useful to select components with high activity in a specific time window compared to a baseline time window. We provide a measure to do this. In this measure, we take the standard deviation across trials of activity (z-scored over the whole time period) averaged in a period of interest (classically after a stimulus) and in a baseline period (classically before the stimulus) and compute the ratio of these two values. If activity increases after stimulus, this ratio is expected to raise, and if it passes a threshold set by the user, the component will be selected. This measure is well suited to identify components showing weak or strong evoked activity according to needs.

2.2.2.5.2. High residual variance of dipole model. Neural sources as isolated with ICA are often well modeled by a single dipole or a pair of dipoles (Delorme et al., 2012). A simple measure of the goodness of fit of dipoles is the residual variance, i.e. proportion of variance that remains in the data after subtracting the data modeled by the dipole. Dipole fitting is performed using the DIPFIT2 plugin, provided by default with EEGLAB. A threshold of 15% is set by default in EEGLAB, which we use by default in SASICA if this measure is selected. Components with more residual variance than threshold will be selected.

2.3. Other automated selection tools

There are several other plugins for EEGLAB that compute statistical properties of ICs and classify components for rejection. We include some of them in SASICA in order to improve classification performance.

FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection; Nolan et al., 2010) is a complete suite of automatic preprocessing routines that performs the entire preprocessing pipeline, from filtering to grand average (i.e. combining several subjects in one average dataset). It allows controlling what steps to undertake and setting thresholds and parameters for every operation it performs. For component rejection, as already mentioned, it uses correlation with EOG channels (EOGCorr), Spatial Kurtosis (SK), Power Spectrum Slope (SpecSl), Hurst Exponent (HE), and

the Median Gradient (MedGrad) of component time-courses. By default, any component whose value on one specific measure is beyond 3 standard deviations from the average is selected for rejection. We have implemented the FASTER ICA artifact selection routines in SASICA to allow users to make use of these measures and classifications in addition to SASICA's own measures.

ADJUST (Automatic EEG artifact Detection based on the Joint Use of Spatial and Temporal features; Mognon et al., 2011) uses elaborate detection algorithms based on temporal and spatial filters to identify components reflecting eye movement artifacts (Blinks, Horizontal and Vertical Eye Movements) and Generic Discontinuities. Noteworthy, it combines explicitly spatial and temporal features to classify components into these four categories. For instance, blink components are detected when a component has a high temporal kurtosis (TK), larger absolute mean inverse weights at frontal electrodes than at posterior electrodes (Spatial Average Difference, SAD), the same sign on left and right portions of the electrode cap, and higher signal variance at frontal than at posterior scalp regions (Spatial Variance Difference, SVD). Other decision algorithms are detailed in the original paper (Mognon et al., 2011). It offers a convenient way to examine the results using EEGLAB's native visualization. We have implemented ADJUST's algorithms within SASICA to allow users to make use of these measures and classifications in addition to SASICA's own measures. Note that because the decision algorithm of ADJUST combines multiple features to classify components, sometimes a given component may show a high value of one feature, but not others required to classify it as artifactual (e.g. the three components shown in Fig. 2). In SASICA, the bars showing computed feature values are displayed in color only when these measures pass threshold and collectively trigger rejection by ADJUST (e.g. Fig. 3B, C, E and H).

MARA (Multiple Artifact Rejection Algorithm; Winkler et al., 2011) is a plugin that combines several measurements in an automated way using a machine learning approach to classify components as artifacts. The measures it uses are Current Density Norm (measure of the smoothness of the topographies), Range Within Pattern (range of amplitudes in the inverse weight matrix), Mean Local Skewness, power at 8–13 Hz, a parameter of the Fit of the Power Spectrum with a 1/F function, as well as the error or this fit (see original paper for details). These features are combined in an automated way using a support vector machine algorithm.

CORRMAP is a tool specifically designed to detect ocular and cardiac artifacts by using the level of correlation between IC maps and a template map chosen by the user. It conveniently uses the STUDY feature of EEGLAB (Delorme et al., 2011), which allows processing a whole group of datasets at once. This tool only allows selection of components based on a user-defined template and thus is only suited for artifacts matching it (typically eye and heartbeat artifact components).

2.4. Validation

We validated our approach by first evaluating how reliably experimenters familiar with ICA would classify components in each of the five categories (blinks, saccades, muscle, isolated bad channel or rare events). To do this, we compared the ratings provided by experimenters, with consensual ratings obtained from expert ICA experimenters. Second, we implemented all methods in SASICA and evaluated them against the same consensual expert classifications in a “training” set of datasets. Finally, we used a different “test” set of 13 datasets to validate SASICA's algorithms (using default settings determined on the training datasets) against new data which had not been used to develop the toolbox. This approach allowed us to develop a selection method robust to many different preprocessing settings, and to

test performance on a more homogenous set of data from a single study.

2.4.1. Datasets

Technical specifications, experimental settings, and subjects' demographics are shown in Table 2. For all datasets, a standard preprocessing procedure was used: (1) visual inspection of the raw signal to exclude bad portions of data, (2) re-referencing, down-sampling and filtering as mentioned in Table 2, (3) epoching, (4) prestimulus baseline removal, (5) ICA using the extended infomax algorithm (from EEGLAB).

2.4.2. Task and measurements

Five experimenters familiar with ICA (hereafter referred to as users) reviewed all ICs for each of the eight training datasets using the manual EEGLAB tool (Tools > Select components by map). This tool presents the topographical maps of all components in a large figure (similar to Fig. 1C) and the user is invited to examine each component to decide whether or not it should be rejected. Clicking on a button above each component pops up a window showing component properties (topography, single trial time courses, and power spectrum). For the present task, we asked users to give "reasons" for their decision to discard a given component. Reasons could be any of 'Blink', 'Saccade', 'Muscle', 'Isolated channel', 'Few trials' (to identify rare events), or 'Other'. Experimenters were given instructions describing each type of artifact as in Sections 2.1.2–2.1.7.

Three expert users (including authors MC and NB) examined the eight test datasets using the same manual EEGLAB tools. In addition to the above mentioned reasons, these expert users were also given the possibility to classify components as "Neural", when they unequivocally matched a neural pattern, as described in Section 2.1.1. After the experts rated all training datasets independently, they sat together and revised their ratings until they reached consensus. Components for which no consensus could be reached were classified as "Other". These classifications are used as a reference to evaluate the responses of the five experimenters and as ground truth for the hit rate and false alarm rate measures described below. The experts have extensive practice with ICA and complied strictly with the criteria described in Section 2.1 for their classifications.

Finally, two experts (including author MC) examined and consensually classified the 920 components of the 13 test datasets using the same procedure. These classifications are used here to further evaluate all automated methods using a larger body of data.

2.4.3. Agreement measures

We measured classification agreement between the five experimenters and the consensual ratings of the three experts for the training datasets for each artifact category and classification method separately. While considering a given method and a given category, components were considered hits, misses, or false alarms, taking the experts' rejections in that category as ground truth. For instance, for evaluating SASICA's focal topography measure for the detection of artifacts from the muscle category, a given component was considered a hit if the experts classified it as "muscle", and the focal topography measure for that component was above threshold. We measured in this way the overall agreement between users and the experts in the test datasets, and between each individual measure of the three automated tools tested and the experts. We report accordingly hit rate and false alarm rate for all measures. We also computed standard signal detection measures (sensitivity *d'* and criterion *c*; Macmillan and Creelman, 2004) to distinguish sensitivity from bias in selections.

We also used Krippendorff's Alpha to measure inter-rater reliability on the training datasets. Krippendorff's Alpha is preferred over more popular measurements such as percentage of

Table 2 Technical specifications and subject demographics for all datasets used. * for the test datasets, the high-pass filter was an analog filter applied during acquisition.

No.	Data		Subject								
	Sampling rate (Hz)	Filtering (band pass, in Hz)	No. of EEG electrodes (+external)	Ref.	No. of trials	No. of total data points	Bad channels	Task	Age	Gen.	Hand.
1	256	1–45	128 (+6)	Avg.	702	72,244,224	NA	Visual detection of faint stimuli	37	F	R
2	256	1–45	128 (+6)	Avg.	714	73,479,168	C3 C4 B24	Saccade task	26	F	R
3	512	1–45	64 (+12)	Avg.	696	41,693,184	NA		19	F	R
4	512	1–45	64 (+12)	Avg.	825	49,420,800	C1		24	F	R
5	512	1–45	64 (+5)	Avg.	1083	76,520,448	NA	Visual detection of faint stimuli	25	F	R
6	512	1–45	64 (+5)	Avg.	1111	78,498,816	NA		31	F	R
7	512	–	64 (+8)	Pz	800	117,964,800	NA	Visual detection of faint stimuli, response by saccade	27	M	R
8	512	–	64 (+8)	Pz	800	117,964,800	Fz	Working memory task	26	M	R
9–21	250	0.01–50*	64 (+7)	R. mastoid	660 ± 84	35067057 ± 4e6	NA		24 ± 4	NA	NA

agreement or Cohen's Kappa, because it accounts for chance agreement and for disagreement in ratings. It is a general measure that includes several other reliability measures as special cases (Hayes and Krippendorff, 2007). It is equal to 0 in case of complete unreliability (random classifications here), and to 1 for perfect reliability (all raters fully agree). We used Krippendorff's Alpha to measure agreement between experts (individual classifications before consensus) and between users. We also computed the 95% confidence intervals of Krippendorff's Alpha using 500 bootstrap resamples.

Finally, we measured the ratio of the variance of the EEG signal captured in correctly rejected components to the variance of all components selected for rejections by the experts (i.e. the variance explained by hits as defined above divided by the variance explained by hits and misses), and the ratio of falsely rejected variance to the variance of all components classified either as "Neural" or as "Other" by the experts (i.e. the variance explained by false alarms divided by the variance of false alarms and correct rejections), to further evaluate performance of all methods. These measures indicate how much of the variance rejected by experts was correctly rejected by automated methods and users and how much variance kept by experts was wrongly rejected by automated methods.

3. Results

In the following, we examine agreement in rejections by experimenters, then each of the reviewed automatic methods in comparison to the consensual artifact classifications by the experts on the training datasets. We also report how automated methods compared to expert classifications on the test datasets. Table 3 summarizes the results.

Overall, blink and saccade IC classifications are the most agreed upon, followed by muscle IC classifications with somewhat lower accuracy. Bad channels and rare event IC classification is less reliable with automated methods, but we notice that human observers also tend to disagree most on these types of artifact ICs. The overall amount of variance correctly and wrongly rejected by all methods is shown in Table 3, revealing that a non-negligible portion of variance gets misclassified by all automated methods. A further breakdown of rejections along with the amount of signal variance involved by artifact category and automated measure for training and test datasets is shown in Supplementary Fig. 1.

In the following, we examine classifications in each artifact category separately.

3.1. Reviewing duration, reliability of experts and users

The four users and three experts took in total between 1h30 and 4h48 to review all 8 training datasets. The two experts who reviewed the thirteen test datasets took 1h30 in total.

Krippendorff's Alpha showed a similar level of agreement between experts (Alpha confidence interval [0.42–0.51] on initial classifications, before consensus was reached), and between regular users (Alpha confidence interval [0.43–0.50]).

3.2. Neural components

Components classified as Neural by the experts were rarely selected for rejection by users (note that the users were not offered the possibility to classify components as of Neural origin, but amongst all the components classified as Neural by the experts, only a few were selected as coming from one of the artifact categories by a minority of users). Users falsely classified a few neural components that were highly similar to blink components as such

(see Section 3.3), and some users also classified some particularly focal neural components as isolated channels (Fig. 5E).

SASICA misclassified one neural component in the training datasets because of a particularly low autocorrelation. This component was part of a dataset where no high-pass filter was applied (see Table 2), and where all components were highly autocorrelated (dataset nos. 7 and 8, Fig. 7) due to slow drifts in the data. To avoid such misclassifications, a rapid look at the summary figure displayed by SASICA (similar to Fig. 1B), showing all SASICA measures, would reveal a particularly high level of autocorrelation in this dataset, which should alert an observant experimenter to further examine this measure. It is then up to the experimenter to decide if such a highly saturated measure is useful in this situation (Autocorrelation for this dataset is $0.99 \pm .02$). ADJUST and FASTER mistook two components of neural origin for blinks (see Section 3.3). In addition, FASTER mistook two neural components (e.g. Fig. 2C) for artifacts because the MedGrad measure reached exceptional values that passed threshold.

In the test datasets, a number of components with large inverse weights on frontal channels were misclassified as blink components, either by ADJUST or FASTER. Two neural components were also misclassified by SASICA for capturing a focal topography, and for low autocorrelation.

3.3. Blinks

Users rejected blink components very consistently. They correctly identified 89% of the blink components labeled by the experts (Table 3). There were only a few such components in each dataset. Components on Fig. 3 B and C were correctly identified by users. Automated methods identified most blink components (Table 3), but missed a few and mistook a few neural components for blinks (see below). Properties that may lead to the misidentification of non-artifactual components for ocular components are listed in Fig. 3G for reference.

Interestingly, two components in training dataset nos. 3 and 4 (illustrated in Fig. 8, for dataset no. 4, and Fig. 3H, for dataset no. 3) were mistaken for eye blinks due to their topography by some users and also mistaken for artifacts by ADJUST and FASTER. Indeed, the topography for these components is close to typical blink topographies (Fig. 3B and C). However, because two EOG electrodes were placed under the eyes in these datasets, actual eye blink topographies look very different in these datasets. In the actual blink component found for these datasets, there was an abrupt polarity reversal at the most anterior sites (Figs. 8 and 3B). As shown in Fig. 8, the falsely identified components actually capture strong evoked activity around 300 to 700 ms at frontal sites. We illustrate the effect of subtracting this non-blink component on the evoked potential at a frontal channel (Fpz) in dataset no. 4 in Fig. 8. Mistakenly subtracting the wrong component in this case may lead to the complete removal of an important part of the ERP.

ADJUST misdetected these components because it is designed to detect blink components by matching component topographies to a template with large difference in inverse weights between anterior and posterior sites (SAD measure). Although ADJUST additionally requires that the amplitude variance be larger at frontal than at posterior sites (SVD measure), this control measure did not prevent misclassification of this component. Nevertheless, ADJUST's performance at detecting blink components was remarkably good and eventually surpassed both SASICA and FASTER with the test datasets, where only two EOG electrodes were used. FASTER also computes the maximal correlation of component time courses with all EOG electrodes, and considers any component whose maximal correlation is above threshold as artifactual. The falsely detected components in dataset nos. 3 and 4 correlated relatively strongly with one or more EOG channels and were thus misclassified by

Table 3

Performance of all users and automated methods in comparison with expert consensus classifications. HR and FAR refer to average Hit and False Alarm Rates across users for each measure, with respect to expert classifications. The “Neural” row has only a false alarm cell because users and automated methods do not explicitly classify a component as “Neural”. Thus only incorrect selection of a neural component as artefactual can be counted here. Overall performance is measured with respect to classification in any of the artifact categories (i.e. not “Other” or “Neural”). The “Var(Overall)” row corresponds to the amount of variance accounted for by the corresponding components. Performance above 20% is highlighted in light gray, and above 50% in dark gray.

Training datasets	Human Users										SASICA										ADJUST										FASTER									
	Blink		Saccades		Muscle		Isolated chin		Eye trials		Channel (EOG) correlation		Auto-correlation		Focal Component		Focal Trails		Horizontal eye movement		Vertical eye movement		Blinks		Generic Discontinuity		Median Gradient		Spatial Kurtosis		Hung Exposure		EOG correlation							
	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR						
Experts	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR						
Blinks	89.3	0.4	5.4	2.4	0.0	12.3	0.0	12.3	0.0	7.1	100.0	1.8	0.0	5.6	0.0	2.5	14.3	5.3	0.0	1.2	85.7	0.3	71.4	1.6	0.0	7.5	0.0	2.7	0.0	1.3	14.3	0.9	71.4	2.4						
Saccades	8.1	1.1	80.1	0.4	0.0	12.4	0.7	12.5	2.2	7.1	58.8	1.4	0.0	5.7	0.0	2.6	5.9	5.4	23.5	0.6	0.0	1.2	17.6	2.0	0.0	7.7	0.0	2.7	0.0	1.4	11.8	0.8	58.8	1.7						
Muscle	0.0	1.4	0.0	2.6	78.9	6.4	5.3	12.8	0.9	7.5	0.0	3.0	42.6	2.4	3.7	2.4	0.0	5.9	0.0	1.3	0.0	1.3	0.0	2.5	9.3	7.3	16.7	1.4	0.0	1.4	3.7	0.8	0.0	3.3						
Bad Channel	0.0	1.4	0.0	2.6	5.8	12.7	68.1	7.0	9.7	6.7	0.0	3.0	1.7	5.9	19.0	1.0	6.9	5.3	1.7	1.1	1.7	1.1	3.4	2.2	48.3	3.7	0.0	2.9	12.1	0.3	0.0	1.1	0.0	3.4						
Bad Trials	0.0	1.4	0.0	2.6	4.5	12.8	23.2	11.2	34.2	4.6	1.8	2.9	0.0	6.1	3.6	2.4	37.5	2.6	3.6	1.0	1.8	1.1	1.8	2.4	16.1	6.7	0.0	2.9	3.6	1.1	1.8	1.0	1.8	3.2						
Neural	1.1		0.5		1.3		0.7		0.1		0.0		1.0		0.0		0.0		0.0		0.0		2.0		0.0		2.9		0.0		0.0		2.9							
Overall	HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR					
	65.4		13.3		21.9		12.6		15.6		8.6		12.5		4.9		64.6		14.2		75.6		19.0		67.9		19.0		62.6		14.2		75.6		19.0					
Var(Overall)	64.6		13.4		84.9		2.9		27.1		5.0		78.7		8.8		67.9		13.4		84.9		2.9		27.1		5.0		78.7		8.8		67.9		13.4					
	d'		c		d'		c		d'		c		d'		c		d'		c		d'		c		d'		c		d'		c		d'		c					
Sensitivity and criterion	1.51		-0.36		0.37		-0.96		0.36		-1.19		0.51		-1.40		0.63		-0.97		0.71		-1.03		0.57		-1.45		0.63		-0.97		0.71		-1.03					

Test datasets	SASICA				ADJUST				FASTER															
	Channel (EOG) correlation		Auto-correlation		Focal Component		Eye Trials		Horizontal eye movement		Vertical eye movement		Blinks		Generic Discontinuity		Median Gradient		Spatial Kurtosis		Hung Exposure		EOG correlation	
	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR
Experts	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR
Blinks	88.2	1.0	0.0	4.0	0.0	4.3	5.9	5.6	0.0	1.9	88.2	4.4	94.1	2.7	29.4	7.4	23.5	2.0	0.0	2.5	0.0	0.3	64.7	1.8
Saccades	22.2	2.4	0.0	4.0	0.0	4.3	33.3	5.4	88.9	1.0	11.1	5.9	11.1	4.3	11.1	7.8	0.0	2.4	0.0	2.5	0.0	0.3	33.3	2.4
Muscle	0.0	3.2	16.5	1.1	13.5	2.1	2.4	6.4	1.2	2.0	11.8	4.7	4.1	4.4	21.8	4.7	4.7	1.9	5.3	1.9	0.0	0.4	0.0	3.3
Bad Channel	14.3	2.4	7.1	3.9	57.1	3.4	7.1	5.6	0.0	1.9	50.0	5.3	14.3	4.2	64.3	7.0	0.0	2.4	42.9	1.9	0.0	0.3	14.3	2.5
Bad Trials	8.3	2.5	4.2	3.9	12.5	4.0	54.2	4.4	8.3	1.7	8.3	5.9	0.0	4.5	20.8	7.5	0.0	2.5	8.3	2.3	8.3	0.1	8.3	2.6
Neural	0.2		0.2		0.2		0.7		0.0		0.7		1.7		0.0		1.5		1.0		0.0		0.7	
Overall	HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR		HR		FAR	
	25.6		10.0		24.9		8.2		12.1		4.1		67.9		19.0		62.6		14.2		75.6		19.0	
Var(Overall)	67.9		19.0		62.6		14.2		75.6		19.0		67.9		19.0		62.6		14.2		75.6		19.0	
	d'		c		d'		c		d'		c		d'		c		d'		c		d'		c	
Sensitivity and criterion	0.63		-0.97		0.71		-1.03		0.57		-1.45		0.63		-0.97		0.71		-1.03		0.57		-1.45	

FASTER. Overall, these misclassifications illustrate the difficulty of creating a fully automated method, and emphasize the need for careful examination of component properties.

SASICA uses correlation with EOG electrodes and, if two electrodes are entered, computes the difference between these electrodes before the correlation with components' activities. This considerably increases the power to detect eye movements accurately. In the examples above, EOG electrodes were placed above and below the eyes and were both submitted to SASICA. The difference between these two electrodes correlated with neither of the two falsely identified components. At the default correlation threshold used (4 standard deviations away from mean), SASICA missed no blink component and misclassified one non-ocular component (classified as capturing activity of a bad trial by experts) due to high correlation with one EOG channel.

Confirming the above results in the test datasets, similar misclassifications of blink components also occurred. ADJUST and FASTER misclassified seven components for the same reason described above: components with strong inverse weight at frontal electrodes but a clear neural topography, power spectrum and time course were mistaken for blink or vertical eye movement components. SASICA on the other hand missed 3 blink components. This can be explained by the fact that only two EOG electrodes were present in these datasets (one below the right eye and one next to the lateral canthus of the left eye). Sensitivity could probably be improved by using bipolar montages around the eyes, which are exploited by SASICA's difference operation mentioned above. It is thus recommended to use several EOG channels around the eyes and combine them in SASICA to detect eye movement artifacts most accurately.

In situations where EOG electrodes cannot be used (e.g. sleep studies, or long term EEG recordings from epileptic patients), a method that does not rely on EOGs is preferable. In this case, the ADJUST eye movement detector that does not resort to EOG electrodes is recommended.

3.4. Saccades

Users correctly identified saccade components 80% of the time in the training datasets. Since in some experiments from which these datasets are drawn, subjects were asked to maintain fixation, there were sometimes very few (if any) such components in each dataset. Properties that may lead to the misidentification of non artificial components for ocular components are listed in Fig. 3G for reference.

Automated methods mistook some ambiguous components for saccades in the training datasets. For instance, SASICA and FASTER, using correlation with EOG electrodes, mistook the component illustrated in Fig. 3I for an ocular component. This component does not isolate eye movement activity specifically and should not be discarded. Its topography includes non-negligible inverse weight at central channels, which are unlikely to reflect ocular artifacts. Furthermore, this component comes from a dataset where EOG electrodes were registered and rendered on the topographies. If this component was related to eye movement artifacts, it would thus show a typical polarity reversal, as shown in Fig. 3E, for instance. Subtracting this component may thus remove signal not directly related to the ocular artifact.

In the test datasets there was a total of only 9 saccade components identified by the experts. SASICA could detect two, and FASTER three of these based on correlation with EOG electrodes, while ADJUST could detect eight of them without using correlation with EOGs. Like for blink components, the lack of several EOG electrodes around the eyes in these datasets severely impaired performance of detection methods based on correlation with EOGs.

Another example of misclassification and its consequences on the ERP is shown in Fig. 9. The data comes from dataset no. 7 of the training set, recorded during a task where subjects had to perform large saccades on almost every trial. In this case (and for dataset no. 8 from the same experiment), due to this unusual task that entails strong recurring artifacts on every trial, the returned ICA

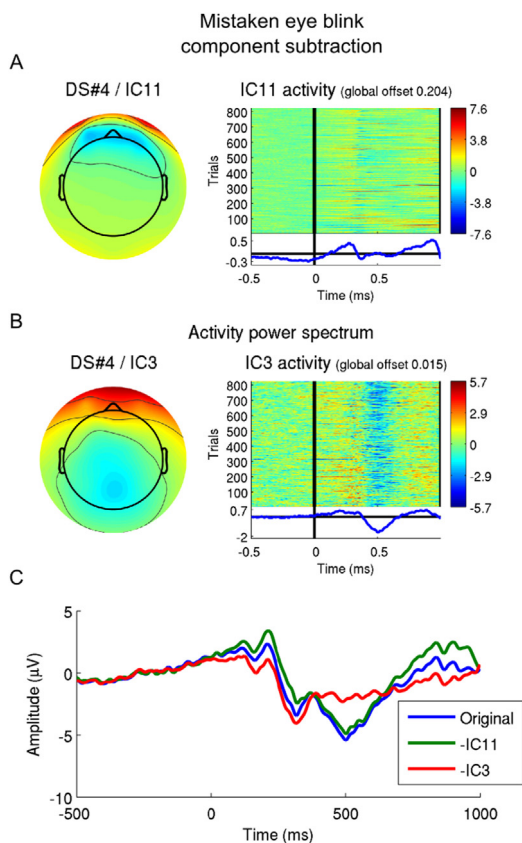


Fig. 8. Mistaken eye blink component subtraction. Results of removing components wrongly identified as blink components in Fig. 3. We take as an example a representative dataset (no. 4). (A) Blink component. (B) Neural component whose topography resembles that of a classical blink component with strong weight at frontal channels. (C) Time course of the event related potential across all trials at electrode Fpz in this dataset before any component subtraction (blue), after subtraction of the blink component in (A) (green), and after subtraction of the non-blink component in (B) (red). The latter operation wipes out entirely a strong evoked activity that is unrelated to blinks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

solution contains several saccade components of similar topography that seem to capture saccades in opposite directions occurring on almost every trial (Fig. 9A and B). ADJUST did not detect any of these saccade components. This is due to the fact that the MEV measure, required to classify a component as capturing eye movements, failed to reach rejection threshold for these components. FASTER and SASICA identified the saccadic components in Fig. 9A and B, but failed to identify the one in Fig. 9D, whose correlation with EOGs was below threshold. This example highlights again how fully automated methods may lead to inappropriate selections, and the need to carefully examine rejections, in particular when unusually strong activities are expected to occur (like in this saccade task).

3.5. Muscle

Users identified 79% of components categorized as muscle components by the experts in the training datasets. Fig. 4B and C shows two typical muscle components identified by users and experts. Fig. 4E and F shows components that were wrongly categorized as pure muscle components by some users. Users often mistook components with isolated channels in the periphery of the cap for muscle components, even if these did not show the steady noise activity pattern characteristic of muscle activity. Some properties that may lead to the misidentification of non-artifactual components as ocular components are listed in Fig. 4D for reference.

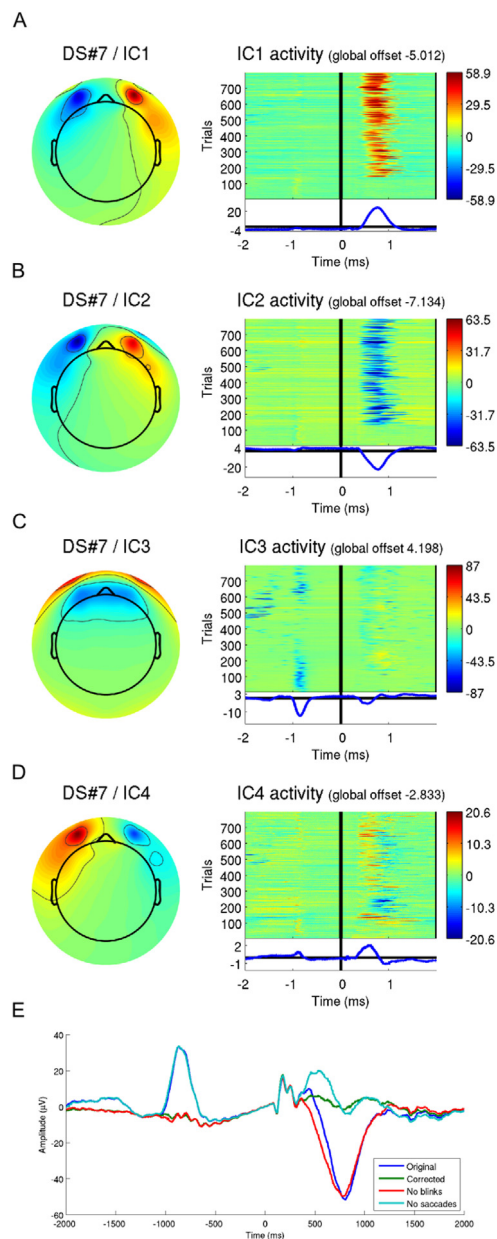


Fig. 9. Complete removal of saccade and blink components. In some cases, artifacts may not have all of the properties that automated methods expect. In the case of dataset no. 7 here, where subjects had to perform a large saccade on almost every trial, several components were clearly identified by the expert observers as saccadic artifacts (IC1, 2 and 4), but some automated methods failed to identify some of these as artifactual components (see text for reasons). The time courses at the bottom show the ERP at electrode Fpz before and after removal of the blink (IC3, correctly identified by all methods) and saccade (IC1, 2 and 4) components. In this case, only the expert users identified all four components.

In SASICA, the autocorrelation measure captured 43% of components classified as coming from muscle activity by the experts in the training datasets. ADJUST does not attempt to select muscle components. FASTER's median gradient measure that is meant to detect noisy components captured 17% of the muscle components selected by the experts in these datasets (see Table 3).

Fig. 4E and F shows examples of mismatches between users, automated methods, and experts in the training datasets. Fig. 4E shows a component which in spite of a clearly focal topography, does not show the usual noisy time courses characteristic of tonic muscular activity and thus fails to qualify for being a muscle component. Fig. 4F shows an example component that the experts did not

classify as a muscle component but whose time course had a very low autocorrelation and high median gradient. This component was considered by the experts as capturing a few high amplitude events (due to a few high amplitude points in an otherwise relatively low amplitude ERP image), and thus did not qualify for being purely muscle-related component.

In the test datasets, the Autocorrelation measure achieved only 16% sensitivity and none of FASTER's measures identified more than 5% of the muscle components. Surprisingly, ADJUST's generic discontinuity detector achieved 22% of detection of these components (see Table 3).

3.6. Bad channels

Users categorized only 68% of the components identified by the experts as coming from a bad channel. Fig. 5 B and C illustrates components capturing bad channels that were detected by users and at least one of the automated methods tested. Fig. 5E illustrates components that some users mistook for a bad channel, although it showed a clear evoked response and its topography actually spread across more than one channel. Fig. 5F illustrates the overlap between rare events and bad channels, since this bad channel, being active for only a short period of time, was classified as reflecting activity of a few trials by the experts.

All automated methods tested showed poor results with bad channels. All detected less than half of the components identified by the experts in the training datasets (see Table 3). It should be noted that classification of these components was also less consensual amongst users. When bad channels are known in advance, SASICA allows the experimenter to search for components with particularly high correlation with known bad channels. The example component of Fig. 5B and C could be detected by SASICA's focal component measure and the component in Fig. 5B was also identified due to its high correlation with a known bad channel in the dataset.

Interestingly, up to 64% of bad channels in the test datasets were correctly identified by ADJUST's Generic Discontinuity detector (and surprisingly 50% by the vertical eye movement detector). With these datasets, FASTER's Spatial Kurtosis and SASICA's Focal Component measures performed also better than with the training datasets (an increase of more than 30% correct detections in both cases, see Table 3). Generally, this high variability in automated measures performance highlights the unreliability of all methods to detect bad channels.

3.7. Rare events

Users identified components capturing activity of a few trials with least accuracy and identified only 34% of the components identified by the expert in this category. Again, the crosstalk of this category with the bad channel category described above is evident by the fact that users also categorized as bad channels some components that the experts classified as capturing activity from a few trials. Automated methods performed generally poorly with these components, with SASICA's focal trial measure identifying 37% of the components labeled as few trials by the expert and all other methods identifying at most 16% of these components.

In the test datasets, automated methods performed slightly better, SASICA with 54% correct detections with its Focal Trial measure, and ADJUST's Generic discontinuity detector performing at 21% correct detections, FASTER not reaching more than 8% detections.

3.8. Other

The "Other" category was used by the experts whenever a component did not capture just one type of artifact, but rather

a mixture. It is important to note that these components should generally not be rejected. They reflect mixtures of signals, some of which undoubtedly are of neural origin, since many of these ambiguous components show various forms of event related responses.

3.9. Additional measures

Additional measures may be used in specific situations. We provide these measures here for completeness and recommend using them only in specific situations where e.g. positive selection of interesting components is wished, rather for systematically discarding artifact components.

3.9.1. Signal to noise ratio

Signal to noise ratio identifies components with little or no evoked activity. In some situations, it may be useful to select only components that contain clear activity evoked by some stimulus. In these cases, users may want to only keep components in which the ratio between pre- and post-stimulus activity is large enough. This option is inactive by default in SASICA because we think it should be used only in some specific cases, such as performing a component specific analysis on components showing strong event related responses. To illustrate the method, we computed signal to noise ratio in all datasets using 500 ms post- vs. 500 ms pre-stimulus activity as period of interest and baseline, respectively, and show the values in all figures. All tasks for all datasets used here had a stimulus event at time 0.

3.9.2. Residual variance of dipole fit

Some neural sources are well modeled by a single dipole or a pair of dipoles symmetrically located in the two hemispheres. It may thus be useful to select only ICs that are well approximated by a dipolar source. EEGLAB allows estimating the location of the IC sources by means of dipole fitting. When performing analyses in component space, EEGLAB suggests working only with ICs whose residual variance after subtraction of the best fitting dipolar model is below a certain threshold. We added this feature in the plugin to help select ICs based on their dipolar fit and perform analyses at the component level.

However, we do not recommend cleaning data with this measure before performing channel based analysis. Indeed, as we have already mentioned, not all components returned by ICA reflect pure neural or artifact activity and usually a lot of them actually capture a mixture of signals that are statistically independent from each other but do not necessarily map onto a pure neural or artifactual process. At least three reasons exist for this. First, physiological artifacts that occur concomitantly with neural processes (e.g. blinks or saccades) should at least in principle be captured together with these neural processes in individual components. Second, it is often the case that clearly artifactual components show some evoked responses to experimental events (see e.g. muscle artifact in Fig. 4C). Finally, spread-out populations of neurons that respond in synchrony are not well modeled by dipoles although they do capture potentially important neural activity.

It is thus a risky bet to consider that all the neural activity one wants to study at the channel level is well isolated in components that are accurately modeled by a single dipole and not at all by any other type of component, and we strongly advise experimenters to think carefully before using this method to reject all components passing the default threshold of 15% residual variance. This feature is retained in SASICA to facilitate characterization of components using this method in future research.

3.10. Overall rejection performance

Table 3 shows overall performance of all methods at rejecting artifact components while keeping non-artifact components (classified as Neural or Other by the experts). Sensitivity is far from perfect for both the training and test datasets, showing that all methods are prone to error. When turning to individual measures, however, automated methods can be very good at selecting specific types of artifacts, and all methods show very low levels of false selection of neural components (“Neural” row in Table 3). This is important because it shows that these methods, although they are not perfect at detecting artifacts, generally rarely mistake neural components for artifacts (but see Figs. 8 and 9).

Table 3 also shows that the amount of variance correctly and wrongly attributed to artifacts by users and automated methods varies across methods. Users classified only 64.6% of artifact variance as such, and mistook 13.4% of “Neural” and “Other” variance for artifacts. Automated methods also missed an important portion of artifact variance, but due to a more conservative criterion (more negative c), were more sparing of non-artifact variance. Performance of automated methods on the test datasets was overall poorer than on the training datasets. A detailed breakdown of variance repartition across rejection categories is shown in Supplementary Fig. 1.

4. Discussion

In this paper, we have reviewed five types of typical EEG artifacts isolated by ICA and several measures currently offered by automated methods to detect these types of artifacts. We introduce the SASICA plugin for the EEGLAB toolbox to help users select artifactual ICs. We evaluated several methods and algorithms for detecting in particular blinks, saccades, muscle noise, bad channels, and rare events in 8 training datasets using different setups from our lab, as well as a 13 dataset test study. Overall, all methods (manual, or automatic) were not perfectly consistent, showing that there are inherent limitations to the precision of artifact selection using ICA. Using fully automated methods, there is to date no way to guarantee that no neural component will be rejected. Important misclassifications occurred with all automated methods that had a large impact on ERPs (Figs. 8 and 9). We thus emphasize here the importance of first defining precisely the goals of artifact correction (blink, muscle, bad channel removal), prior to attempting artifact correction with ICA. Second, we recommend using SASICA and the appropriate measures to select potentially artifactual components. With the help of a rapid and convenient overview of diagnostic measures, SASICA allows users to make informed and efficient decisions based on objective criteria. Taking this approach of using statistical measures not to make decisions, but to guide the experimenter's decision is in our view essential to efficiently remove EEG artifacts using ICA while avoiding mistakenly removing neural signals from the data.

SASICA provides suggestions for ICs to reject along with information about the decisive criteria for each suggestion. The global view of all components offered by SASICA (Fig. 1B and C), which allows rapid reviewing of all components is an advantage over manual rejections and other automated methods. Furthermore, the new measures introduced with this tool help refine selection in some critical situations. Finally, SASICA can also be easily used to select components for other purposes than rejection via additional measurements (e.g. signal to noise ratio and residual variance of dipole fits). A command line interface allows e.g. easy selection of components with given levels of correlation with specific channels, or selection of components that have a particularly strong evoked response.

A striking aspect of the results of our evaluation procedure is that perfect agreement was never achieved across methods. Even between human experts, consensus could only be reached after discussion. This reflects the difficulty of defining reliable rejection criteria a priori. Although all users had experience with ICA analysis, and all were aware that the goal of this evaluation was to reject artifactual components, they still spontaneously disagreed on many occasions. This limitation puts an upper limit to the level of agreement that could be reached by automated methods. No matter what automated method is used, there is always a certain level of disagreement with human users. The main challenge experimenters are facing is to avoid by all means rejecting important neural data. In this context it is important to note that it is very hard for experimenters not to rely spontaneously on implicit rules to classify the more ambiguous components. These different attitudes are strongly influenced by personal experience as an experimenter and the goals of a given experiment.

The physiological processes underlying certain artifacts have functional consequences that have to be taken into account when analyzing data. For instance, the ocular artifact generated during the execution of saccades are associated with a series of brain potentials that experimenters may not want to remove from their data (Dandekar et al., 2012; Gaarder et al., 1964). We therefore recommend awareness and reporting of the constraints imposed by a specific experimental paradigm, and a good understanding and overseeing of any automatic algorithm used to automatically select components for rejection. For instance, high frequency gamma-band oscillations (>40 Hz) have power at frequencies that overlap with those of muscle artifacts. Components capturing muscle activity could thus also capture at least some neural gamma activity. An experimenter used to analyzing high frequency gamma oscillations may thus want to preserve his or her preprocessing strategy to interfere with their ability to capture gamma sources, and prefer to avoid removing any component with power at high frequencies (Chaumon et al., 2009). Another experimenter using Brain Computer Interface may on the contrary want to automatically subtract muscle artifacts, to avoid entering this data in the classification algorithms but still attempt to keep concomitant EEG activity, and therefore favor an ICA subtraction method (Fatourechhi et al., 2007). While it is laudable to attempt to minimize these experimenter specific biases, they cannot be easily discarded and it is in our opinion more important to be aware of them and attempt to correct them by having clear goals and criteria set up front, rather than just considering them a bad practice and ignoring their existence.

ICA is by far not the only method available to correct for artifacts. A multitude of other approaches allow potentially preserving neural signal while discarding artifacts. These methods include low-pass filtering, regression methods, principal component analysis (Ille et al., 2002; McMenamin et al., 2009; Wallstrom et al., 2004), signal space projection (Uusitalo and Ilmoniemi, 1997), or canonical correlation analysis (Clercq et al., 2006). But ICA is a powerful method that allows correcting for several types of artifacts at once. It is used to remove artifacts in EEG, but also in magnetoencephalographic (MEG) data (Barbati et al., 2004; Parra et al., 2005), as well as the strong artifacts created in the EEG by concomitant functional magnetic resonance imaging (Mantini et al., 2007). It has proven useful in correcting ocular (Jung et al., 2000b), cardiac (Campos Viola et al., 2009), muscle (Crespo-García et al., 2008; with some discussion, in McMenamin et al., 2010; Olbrich et al., 2011), or ictal artifacts (Urrestarazu et al., 2004). We have thus focused this review on ICA because of its widespread use, and because, as we have shown, it is not devoid of pitfalls. It is therefore important to inform experimenters how to properly classify artifactual ICs.

Acknowledgements

M.C. is supported by the Deutsche Forschungsgemeinschaft (DFG) (CH 1246-1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jneumeth.2015.02.025>.

References

- Barbati G, Porcaro C, Zappasodi F, Rossini PM, Tecchio F. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clin Neurophysiol* 2004;115(5):1220–32, <http://dx.doi.org/10.1016/j.clinph.2003.12.015>.
- Campos Viola F, Thorne J, Edmonds B, Schneider T, Eichele T, Debener S. Semi-automatic identification of independent components representing EEG artifact. *Clin Neurophysiol* 2009;120(5):868–77, <http://dx.doi.org/10.1016/j.clinph.2009.01.015>.
- Chaumon M, Schwartz D, Tallon-Baudry C. Unconscious learning versus visual perception: dissociable roles for gamma oscillations revealed in MEG. *J Cogn Neurosci* 2009;21(12):2287–99, <http://dx.doi.org/10.1162/jocn.2008.21155>.
- Clercq WD, Vergult A, Vanrumste B, Van Paesschen W, Van Huffel S. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Trans Biomed Eng* 2006;53(12):2583–7, <http://dx.doi.org/10.1109/TBME.2006.879459>.
- Crespo-García M, Atienza M, Cantero JL. Muscle artifact removal from human sleep EEG by using Independent component analysis. *Ann Biomed Eng* 2008;36(3):467–75, <http://dx.doi.org/10.1007/s10439-008-9442-y>.
- Dandekar S, Privitera C, Carney T, Klein SA. Neural saccadic response estimation during natural viewing. *J Neurophysiol* 2012;107(6):1776–90, <http://dx.doi.org/10.1152/jn.00237.2011>.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134(1):9–21, <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>.
- Delorme A, Mullen T, Kothe C, Acar ZA, Bigdely-Shamlo N, Vankov A, et al. EEGLAB, SIFT, NIFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput Intell Neurosci* 2011;2011(10–10):10, <http://dx.doi.org/10.1155/2011/130714>.
- Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S. Independent EEG sources are dipolar. *PLoS ONE* 2012;7(2):e30135, <http://dx.doi.org/10.1371/journal.pone.0030135>.
- Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage* 2007;34(4):1443–9, <http://dx.doi.org/10.1016/j.neuroimage.2006.11.004>.
- Fatourechi M, Bashashati A, Ward RK, Birch GE. EMG and EOG artifacts in brain computer interface systems: a survey. *Clin Neurophysiol* 2007;118(3):480–94, <http://dx.doi.org/10.1016/j.clinph.2006.10.019>.
- Gaarder K, Krauskopf J, Graf V, Kropfl W, Armington JC. Averaged brain activity following saccadic eye movement. *Science* 1964;146(3650):1481–3, <http://dx.doi.org/10.1126/science.146.3650.1481>.
- Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Measures* 2007;1(1):77–89.
- Ille N, Berg P, Scherg M. Artifact correction of the ongoing eeg using spatial filters based on artifact and brain signal topographies. [Miscellaneous Article]. *Journal of Clin Neurophysiol* 2002;19(March (2)):113–24.
- Joyce CA, Gorodnitsky IF, Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology* 2004;41(2):313–25, <http://dx.doi.org/10.1111/j.1469-8986.2003.00141.x>.
- Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 2000a;37(02):163–78.
- Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 2000b;111(10):1745–58, [http://dx.doi.org/10.1016/S1388-2457\(00\)00386-2](http://dx.doi.org/10.1016/S1388-2457(00)00386-2).
- Macmillan NA, Creelman CD. *Detection theory: a user's guide*. second ed. Mahwah, NJ: Psychology Press; 2004.
- Makeig S, Bell AJ, Jung T-P, Sejnowski T. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, vol. 8. Cambridge, MA: MIT Press; 1996. p. 145–51, Retrieved from (<http://sccc.ucsd.edu/~scott/pdf/NIPS95b.pdf>).
- Mantini D, Perucci MG, Cugini S, Ferretti A, Romani GL, Del Gratta C. Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. *NeuroImage* 2007;34(2):598–607, <http://dx.doi.org/10.1016/j.neuroimage.2006.09.037>.
- McMenamin BW, Shackman AJ, Maxwell JS, Bachhuber DRW, Koppenhaver AM, Greischar LL, et al. Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage* 2010;49(3):2416–32, <http://dx.doi.org/10.1016/j.neuroimage.2009.10.010>.
- McMenamin BW, Shackman AJ, Maxwell JS, Greischar LL, Davidson RJ. Validation of regression-based myogenic correction techniques for scalp and source-localized EEG. *Psychophysiology* 2009;46(3):578–92, <http://dx.doi.org/10.1111/j.1469-8986.2009.00787.x>.
- Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 2011;48(2):229–40, <http://dx.doi.org/10.1111/j.1469-8986.2010.01061.x>.
- Nolan H, Whelan R, Reilly RB. FASTER: fully automated statistical thresholding for EEG artifact rejection. *J Neurosci Methods* 2010;192(1):152–62, doi:10.1016/j.jneumeth.2010.07.015.
- Okada Y, Jung J, Kobayashi T. An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements. *Physiol Measure* 2007;28(12):1523, <http://dx.doi.org/10.1088/0967-3334/28/12/006>.
- Olbrich S, Jödicke J, Sander C, Himmerich H, Hegerl U. ICA-based muscle artefact correction of EEG data: What is muscle and what is brain?: Comment on McMenamin et al. *NeuroImage* 2011;54(1):1–3, <http://dx.doi.org/10.1016/j.neuroimage.2010.04.256>.
- Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. *NeuroImage* 2005;28(2):326–41, <http://dx.doi.org/10.1016/j.neuroimage.2005.05.032>.
- Urrestarazu E, Iriarte J, Alegre M, Valencia M, Viteri C, Artieda J. Independent component analysis removing artifacts in ictal recordings. *Epilepsia* 2004;45(9):1071–8, <http://dx.doi.org/10.1111/j.0013-9580.2004.12104.x>.
- Uusitalo MA, Ilmoniemi RJ. Signal-space projection method for separating MEG or EEG into components. *Med Biol Eng Comput* 1997;35(2):135–40, <http://dx.doi.org/10.1007/BF02534144>.
- Vos DM, Riès S, Vanderperren K, Vanrumste B, Alario F-X, Huffel VS, et al. Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 2010;8(2):135–50, <http://dx.doi.org/10.1007/s12021-010-9071-0>.
- Wallstrom GL, Kass RE, Miller A, Cohn JF, Fox NA. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *Int J Psychophysiol* 2004;53(2):105–19, <http://dx.doi.org/10.1016/j.ijpsycho.2004.03.007>.
- Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behav Brain Funct* 2011;7(1):30, <http://dx.doi.org/10.1186/1744-9081-7-30>.