FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection☆

H. Nolan[1], R. Whelan[*,1], R.B. Reilly

Trinity Center for Bioengineering, Trinity College Dublin, Ireland

ABSTRACT

Electroencephalogram (EEG) data are typically contaminated with artifacts (e.g., by eye movements). The effect of artifacts can be attenuated by deleting data with amplitudes over a certain value, for example. Independent component analysis (ICA) separates EEG data into neural activity and artifact; once identified, artifactual components can be deleted from the data. Often, artifact rejection algorithms require supervision (e.g., training using canonical artifacts). Many artifact rejection methods are time consuming when applied to high-density EEG data. We describe FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection). Parameters were estimated for various aspects of data (e.g., channel variance) in both the EEG time series and in the independent components of the EEG: outliers were detected and removed. FASTER was tested on both simulated EEG ($n = 47$) and real EEG ($n = 47$) data on 128-, 64-, and 32-scalp electrode arrays. FASTER was compared to supervised artifact detection by experts and to a variant of the Statistical Control for Dense Arrays of Sensors (SCADS) method. FASTER had >90% sensitivity and specificity for detection of contaminated channels, eye movement and EMG artifacts, linear trends and white noise. FASTER generally had >60% sensitivity and specificity for detection of contaminated epochs, vs. 0.15% for SCADS. FASTER also aggregates the ERP across subject datasets, and detects outlier datasets. The variance in the ERP baseline, a measure of noise, was significantly lower for FASTER than either the supervised or SCADS methods. ERP amplitude did not differ significantly between FASTER and the supervised approach.

1. Introduction

The event-related potential (ERP) is computed by aggregating across time-locked electroencephalograms (EEG) epochs. Artifacts – such as eye and muscle movements (measured by electro-oculograms (EOG) and electromyograms (EMG), respectively), and electrode displacement – can be orders of magnitude greater than the ERP, thereby greatly distorting the signal. For example, EEG signals are in the order of tens of μV whereas EOG and EMG signals are in the order of hundreds of μV. In addition to eye and muscle movement artifacts, poor scalp contact for a particular electrode will produce consistently bad data for the duration of the recording. Other artifacts include spurious electrical activity picked up by the EEG amplifier, and current drift.

One simple and computationally inexpensive approach to eye and muscle movement artifact detection and rejection involves deleting portions of the data with artifacts (e.g., EOG data with amplitudes $\pm 75\,\mu V$). However, this can potentially lead to a large loss of data, consequently reducing the quality of the ERP. Electrodes with consistently poor signal quality are typically removed and then recreated using interpolation from the remaining electrodes, which effectively reduces the spatial resolution of the EEG. Therefore, several methods of removing eye and muscle movement artifacts while retaining EEG data have been proposed (Croft and Barry, 2000; Moretti et al., 2003; Schlögl et al., 2007).

There are extant methods for detection of artifacts in high-density EEG data, many of which are applicable only to specific artifact types (e.g., eye movement artifacts). Some methods for artifact detection have a broader scope, however. For example, the Statistical Control of Artifacts in Dense Arrays Studies (SCADS) method, fully described in Junghöfer et al. (2000), used thresholding methods to detect artifacts. In this approach, several editing matrices, containing parameters such as standard deviation (SD), maximum gradient, and maximum amplitude value, are constructed for each channel within each epoch. Thresholds are calculated for each parameter across whole epochs, whole channels, and single channels in single epochs using a non-parametric formula to measure the spread of the distribution. Whole channels, whole epochs, or single channels within single epochs whose
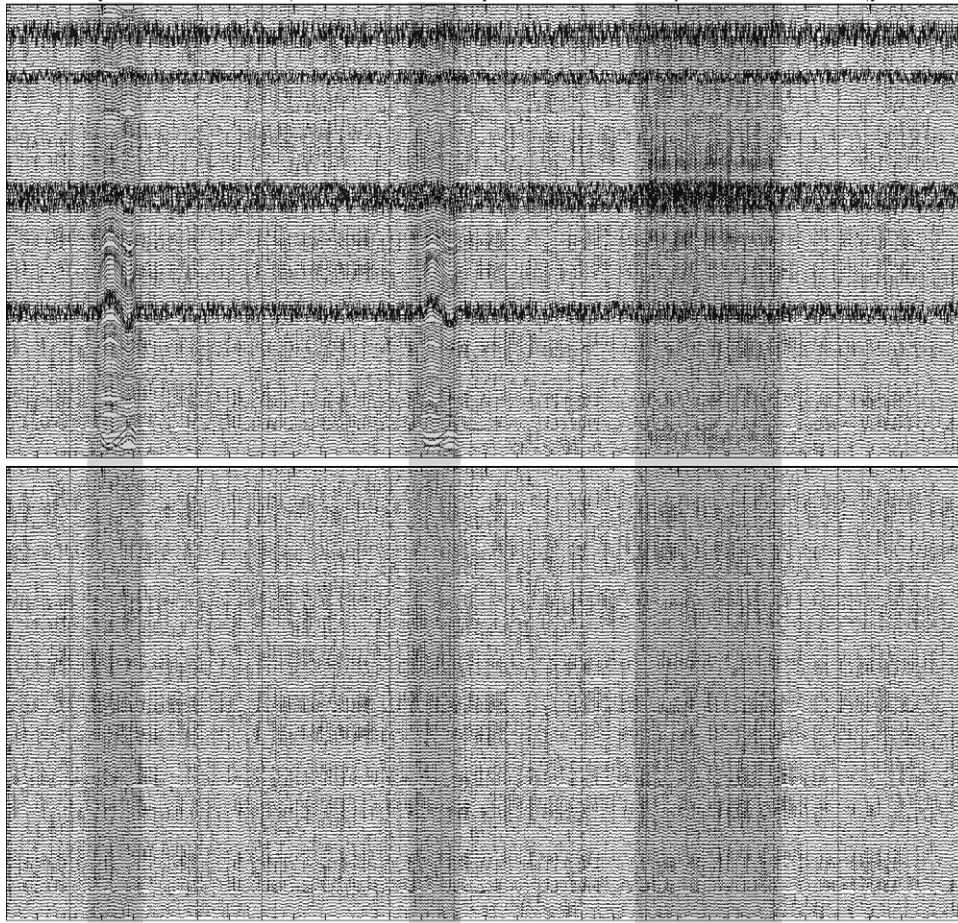
Fig. 1. A section of a dataset with simulated artifacts added (top) and after removal with FASTER (bottom). Two EOG artifacts and one EMG portion added to the data are highlighted. These have been removed in the bottom panel. A number of noise-contaminated channels are also present in the top panel, which have been corrected in the lower panel. This represents a 100% sensitivity (i.e. all artifacts removed) and 100% sensitivity (i.e. no uncontaminated channels or epochs removed) in the three epochs shown.

parameters exceeded the thresholds are removed (epochs) or inter-polated (channels).

Another approach to artifact detection involves the use of independent component analysis (ICA), which is widely available through the EEGLAB software suite (Delorme and Makeig, 2004). ICA is a computational method that separates time series data into statistically independent component (IC) waveforms. ICA outputs a matrix that transforms EEG data to IC data, and its inverse matrix to transform IC data back to EEG data. These matrices give information about an IC's spatial properties, and the data gives information about the IC's temporal activity. Data recorded from scalp electrodes can be considered summations of EEG data and artifact, which are independent of each other: ICA is therefore potentially a useful methodology to separate artifact from EEG signal (Jung et al., 2000; Vorobyov and Cichocki, 2002). There is, however, a need to classify the resulting components as either artifactual or neural (Bian et al., 2006). If detected, artifactual ICs can then be subtracted from the recorded data and the remaining data can be remixed. Several methods for detecting and rejecting artifacts based on ICA have been described previously in the literature. Some of these approaches can be limited by the requirement for the detection process to be trained from predefined artifacts (i.e. supervised), which are not always available, or may not be generalizable (Delorme et al., 2007; Jung et al., 1998; Schlögl et al., 2007; Yandong et al., 2006). For example, eye movement artifacts vary in shape, amplitude and length between subjects. The training procedure is generally carried out manually from visually identified artifacts, and conse-quently full unsupervised automation is not possible with these methods.

In short, methods have been developed for removing various types of artifacts from EEG. However, as many are application dependant, or focused on a single type of artifact, it can be difficult to choose which approach(es) to take. Furthermore, all approaches involve at least some degree of supervision for classification of arti-facts. Given the trend towards ever-denser EEG arrays, such artifact rejection methods are time consuming. We describe here a method called Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER) in which raw data are imported, bad channels removed, epochs extracted, artifacts detected and removed using ICA, subjects' data aggregated, and data sets from subjects with unacceptably artifact-contaminated detected data are removed. This fully automated, unsupervised approach – with raw EEG data as the input and epoched, artifact-attenuated data as the output – would therefore be of use to the many researchers who collect EEG data.

With any new method of processing EEG data, and in partic-ular for a fully automated, unsupervised method, it is essential to quantify the improvement of signal-to-noise and the rate of artifact detection against other established methods. In addition, meth-ods suitable for dense EEG arrays may not be applicable to lower density arrays, and this applicability needs to be tested also. For example, the ability to detect outliers is improved with increasing sample size and the number of independent components that can be estimated robustly is a function of the number of data points.

Therefore, we tested FASTER in a number of scenarios. We compared the FASTER and SCADS methods on 128-, 64-, and 32-scalp electrode arrays simulated EEG data with simulated artifacts. The advantage of testing an artifact rejection method on simulated data is that the sensitivity and specificity of the detection algorithms can be quantified. While simulated data are useful, however, they often contain artifacts with known properties (e.g., EOG amplitude of at least 80 μV). In contrast, real EEG data contain a variety of artifacts whose properties are unknown. Therefore, we compared real 128-channel EEG data from 47 subjects analyzed using the FASTER method, artifacts detected visually by trained individuals, and with SCADS. FASTER was also compared with SCADS on 64- and 32-scalp electrode array subsets of the real data to test the effect of using fewer data points.

2. Methods for EEG data collection/simulation

2.1. Simulated data

Forty-seven sets of simulated data were created. These consisted of 200 epochs of data simulated from P3 dipoles using the BESA Dipole Simulator program (which is found at http://www.besa.de/updates/tools/), which then had artifacts added at random. The procedure for creating artifacts was derived from Delorme et al. (2007). In order to create contaminated channels, white noise was added to a random number of channels (range 0–5). The white noise was of RMS amplitude randomly selected to be 1–10 times that of the channel itself. A random number of the 200 epochs (range 0–15) had a high-amplitude (30–150 μV) low frequency (1–3 Hz) wave added to all channels to simulate an electrode-shift artifact. The position of all artifacts was recorded and the data were then processed. The pre-artifact data and post-processing data were compared for the duration of each artifact to assess the degree of artifact removal. Fig. 1 displays an example of data pre- and post-artifact removal.

2.2. Real data

Forty-seven datasets from healthy controls from a visual oddball paradigm were analyzed (mean age = 37.6 years, 28 males). ERP data were recorded in a soundproofed room using the ActiveTwo Biosemi™ electrode system from 134 electrodes (128-scalp electrodes) organized according the 10-5 system (Oostenveld and Praamstra, 2001) digitized at 512 Hz. The vertical and horizontal electro-oculograms were recorded bilaterally from approximately 3 cm below the eye and from the outer canthi, respectively. An additional two electrodes were placed on the mastoids bilaterally.

Subjects observed blue circles, separated by an inter-stimulus interval of 2 s, presented for 205 trials in a pseudorandom order. Frequent non-target (80%) and infrequent target (20%) circles were 2 cm or 4 cm in diameter, respectively. Subjects were instructed to press a button as quickly as possible following a target stimulus. This type of oddball task typically evokes a positive deflection in the ERP with a latency of approximately 300 ms following the target stimulus. This is called the P3/P300 component. These data were acquired with the understanding and written consent of each subject, the approval of St. Vincent's University Hospital Ethics Committee, and in compliance with national legislation and the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki).

All data were recorded from a Biosemi™ ActiveTwo 128-channel EEG system at 512 Hz. The Biosemi™ system replaces the ground electrodes used in conventional systems with two separate electrodes: Common Mode Sense active electrode and Driven Right Leg passive electrode. These two electrodes form a feedback loop, which drives the average potential of the subject (the Common Mode voltage) as close as possible to the analogue-to-digital reference voltage in the AD-box (the analogue-to-digital reference can be considered the virtual ground of the amplifier). For a detailed description of the referencing and grounding conventions used by the Biosemi™ active electrode system, the interested reader is referred to the following website: http://www.biosemi.com/faq/cms&drl.htm.

3. Methods for artifact detection and removal

3.1. Method overview

We describe here the general approach to artifact detection and removal, in order to give the reader an overview of the procedure. The details of each particular method are described in subsequent sections. The real datasets were converted to EEGLAB format and then referenced to Fz – this was chosen as it was common to the 128-, 64- and 32-channel datasets. The EEG data were filtered offline using equiripple filters between 1 Hz and 95 Hz with a notch filter at 50 Hz (bandwidth 6 Hz) to remove mains interference. Channels were analyzed for artifacts (using either FASTER, SCADS, or supervised rejection), and any contaminated channels were interpolated. Data were then epoched from −500 ms to 1500 ms, and baseline corrected from −200 ms to 0 ms. Epochs were analyzed for artifacts (using either FASTER, SCADS, or supervised rejection), and any contaminated epochs were removed from the dataset. The data were referenced to the average of all scalp electrodes.

For FASTER and supervised processing, ICA was then performed on the dataset, and the resulting ICs were analyzed for artifacts. Contaminated ICs were subtracted from the dataset. After this, for FASTER and SCADS processing, each channel per epoch was analyzed for artifacts, and if found to be contaminated, was interpolated. Finally, for all methods, the ERP of each dataset was taken, baseline corrected from −200 ms to 0 ms (as absolute values of the EEG may have changed following IC removal), and concatenated to make a grand average dataset. For FASTER and supervised methods, each subject's ERP was analyzed for artifacts, and if found to be of poor quality, removed from the average.

3.2. FASTER

Data artifacts were detected and corrected in five aspects of the EEG data: channels, epochs, ICs, single-channel single-epochs, and aggregated data (i.e. across subjects). Fig. 2 displays a flowchart showing each step of the FASTER method. For each aspect, statistical parameters of the data were calculated. The metric that defined contaminated data was a Z-score of ±3 for that parameter (this is a definition of an outlier). For example, a channel whose variance had a Z-score of 3 would be deemed to be contaminated. Various other methods for thresholding were considered (and rejected), and are discussed in the supplemental methods section.

For the mathematical descriptions of the properties, the following conventions apply:

n = 1, 2, . . ., N indicates a specific channel (N is the number of channels in the dataset)
e = 1, 2, . . ., E indicates a specific epoch (E is the number of epochs in the dataset)
c = 1, 2, . . ., C indicates a specific IC (C is the number of components in the dataset)
x_n indicates the data in channel n
x_{n_e} indicates the data in channel n within epoch e
x_{c_t} indicates the temporal data in component c
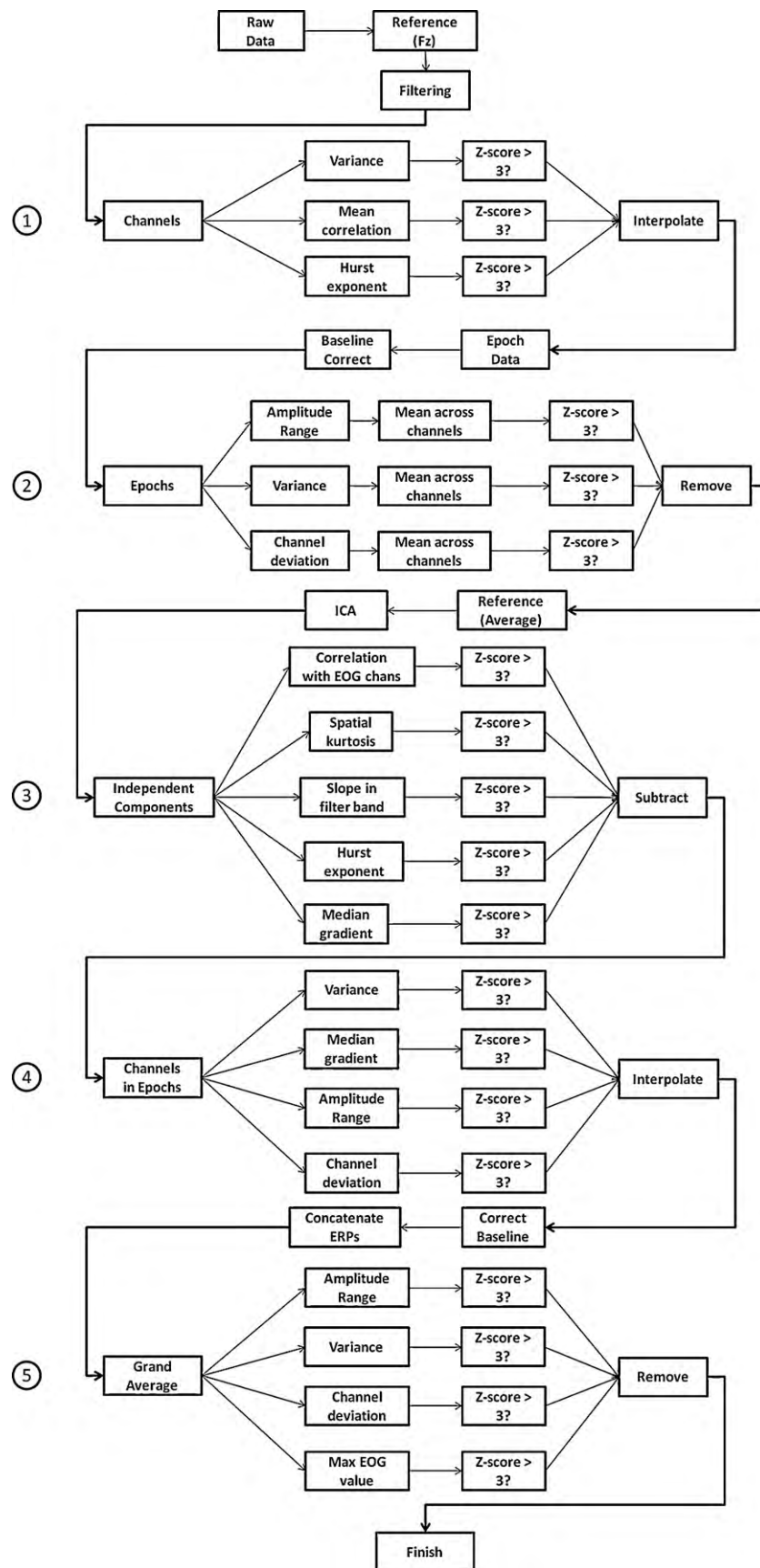x_{c_s} indicates the spatial data in component c

Fig. 2. A flowchart detailing each step of the FASTER method.

s^2x indicates the variance of data

$\langle x \rangle$ indicates taking the mean of data

$\langle x \rangle_N$ indicates taking the mean of data across channels

$\langle x \rangle_E$ indicates taking the mean of data across epochs

$EOG_{1,2,3,4}$ the EOG electrodes

$F(x)$ power spectrum (calculated using MATLAB's *pwelch* function)

3.2.1. Channel artifacts

An electrode or subset of electrodes in an EEG dataset may move during an EEG session, resulting in bad contact with the scalp and therefore a poor quality signal. More rarely, electrodes may also have mechanical faults, for example frayed wiring, which can partially or wholly degrade the signal received. Such electrodes can produce erratic signals. To classify channels as artifactual, three parameters of each channel were calculated:

3.2.1.1. The first parameter was the mean of the channel's correlation coefficients with other channels. Most channels, especially in a high-density system, should correlate highly with neighboring channels. Therefore, a channel with contaminated data will likely have a low correlation with other channels.

Parameter 1: $\frac{\sum_{m=1}^{N} r_{x_n,x_m}}{N}$, the mean correlation coefficient of channel n, where r_{x_n,x_m} is the Pearson correlation coefficient between channels n and m

3.2.1.2. Alternatively, a contaminated channel may correlate quite well with other channels, but have a higher variance (due to additive noise) and therefore the second parameter is the variance of the channel.

Parameter 2: $S_{x_n}^2$, the variance of channel n

3.2.1.3. The third parameter was the *Hurst exponent*. The Hurst exponent is a measure of long-range dependence within a signal. Human phenomena such as EEG have values of $H \approx 0.7$, and signals that deviate from this number are more likely to be artifacts (for more details see Appendix A).

Parameter 3: H_{x_n}, the Hurst exponent of channel n

The above parameters were corrected for reference offset using the method described in (Junghöfer et al., 2000). Channels identified as contaminated were removed and data at this electrode were reconstructed by interpolating from neighboring electrodes using the EEGLAB 7 spherical spline interpolation function.

3.2.2. Epoch artifacts

An epoch in an EEG dataset may at times be contaminated with all-channel noise, typically caused by subject movement, and subsequent physical movement of the electrodes. To detect such epochs, 3 parameters were computed for each channel within the epoch.

3.2.2.1. The movement of electrodes on the scalp results in a change in impedance between the scalp and the electrodes, which consequently affects the electrode voltage offsets. This offset change contaminates epochs, and can be identified by its high amplitude. To detect this contamination, the first parameter computed was the amplitude range of the epoch.

Parameter 4: $\left\langle \max(x_{n_e}) - \min(x_{n_e}) \right\rangle_N$, the amplitude range in epoch e

3.2.2.2. Shifting electrodes may also produce less extreme movements that may not have sufficient amplitude range to exceed the threshold of a Z-score equal to 3, but still contaminate an epoch. This type of artifact may be reflected in a high deviation of that epoch's average value from the average values across all channels. The second parameter computed was the deviation from each channel's average value.

Parameter 5: $\left\langle \langle x_{n_e} \rangle - \langle x_n \rangle \right\rangle_N$, the deviation from the channel average in epoch e

3.2.2.3. Subject movement also produces EMG interference. A high variance will reflect such activity, and so the third parameter calculated was the variance.

Parameter 6: $\left\langle S_{x_{n_e}}^2 \right\rangle_N$, the variance in epoch e

3.2.3. IC artifacts

The *Infomax* (Bell and Sejnowski, 1995) algorithm was employed to perform the ICA decomposition. The number of data points needed to find C stable components from ICA is typically kC^2 for each data channel, where k is a multiplier. The k value was set to 25, as recommended in (Onton et al., 2006). For example, our real data were of length $512 \times 205 = 104960$ points, given 128-scalp channels, 4 EOG channels and 2 mastoid channels, the maximum possible number of ICs would be 134. However, this would not have met the $k = 25$ criterion, which would necessitate $25 \times (134)^2 = 448900$ data points. Therefore, the value of C was reduced to C_{pca} by performing Principal Component Analysis (PCA) on the EEG data and keeping only the first C_{pca} principal components (Shlens, 2005). C_{pca} was calculated as:

$$C_{pca} = \text{floor}\left(\sqrt{\frac{L}{k}} \right),$$

where L is the length of the EEG dataset (in samples), and floor indicates rounding down to the nearest integer.

This reduces the rank of the data and so a smaller number of ICs are computed. Interpolation of channels also reduces the rank of the data, and so C_{pca} was corrected to account for this:

$$C_{pca} = \min\left(\text{floor}\left(\sqrt{\frac{L}{k}} \right), N - N_{interpolated} \right),$$

where N was the original number of channels and $N_{interpolated}$ was the number of channels interpolated in the channel interpolation step.

ICA often produces ICs which consist entirely of artifactual data. These can then be subtracted from the dataset, leaving EEG data without the artifact. To classify ICs, five parameters were computed.

3.2.3.1. To identify artifacts caused by eye blinks (vertical EOG, VEOG) or saccades (horizontal EOG, HEOG), the correlation coefficients of each IC time series with the four recorded EOG (two VEOG and two HEOG) data channels was calculated, and the maximum absolute value was taken as the first parameter. The absolute was taken to account for possible differences in polarity between the EOG channel data and the IC time series.

Parameter 7: $\max(\left| r_{x_{c_t}, x_{EOG_{1,2,3,4}}} \right|)$, the maximum of the absolute correlation coefficient between component c time-course and EOG channels

3.2.3.2. Another common type of artifact singled out by ICA is a short, high-amplitude, single-electrode offset, often termed a "pop-off". An IC consisting of a pop-off has spatial data which shows activity in a single-channel and none otherwise. This is reflected in a high kurtosis value in the spatial data, as kurtosis measures the peakedness of data. The second parameter computed was the kurtosis of the spatial data.

Parameter 8: $\frac{\mu_4}{\mu_2^2} - 3$, where μ_i gives the ith central moment of the spatial data, is the equation for kurtosis of the spatial information in component c

3.2.3.3. There is typically white noise in the acquired data due to hardware properties. White noise has a close-to-flat frequency power spectrum, as opposed to EEG components which have a $1/f$

power spectrum distribution. Residual white noise may remain after filtering, albeit with a very low contribution. Independent components consisting of white noise were identified by calculating the slope of the spectrum over the low-pass filter band as the third parameter.

Parameter 9: $\left\langle \frac{dF(x_{c_t})}{df} \right\rangle \Big| f_{LP1} < f < f_{LP2}$, the mean slope of the power spectrum of the component c time-course, between the band edges of the low-pass filter band

3.2.3.4. The fourth parameter estimated was the Hurst exponent. Parameter 10: $H_{x_{c_t}}$, the Hurst exponent of component c time-course

3.2.3.5. The fifth parameter was the median gradient value, which is above threshold if the IC contains considerable high-frequency content, was also calculated for each IC time series. Parameter 11: $median\left(\frac{d(x_{c_t})}{dt} \right)$, the median slope of the component c time-course

3.2.4. Single-channel, single-epoch artifacts

Following the previous three steps, a high percentage of artifacts will have been removed. Some small transient artifacts may remain on single channels, within single epochs – for example, short bursts of white noise due to transient electrical faults, or electrodes that lost contact during a recording and were not sufficiently noisy to be detected as bad channels. Such artifacts were corrected by interpolating single channels within single epochs, using spherical splines. To detect the artifacts, four parameters were computed for each channel within each epoch.

3.2.4.1. The first parameter was the variance, to detect single channels in single epochs with additive noise.

Parameter 12: $S^2_{x_{n_e}}$, the variance of channel n in epoch e

3.2.4.2. The second was the median gradient, to detect other high-frequency activity. Parameter 13: $median\left(\frac{d(x_{n_e})}{dt} \right)$, the median slope of the channel n in epoch e

3.2.4.3. The third was the amplitude range of the channel, to detect pop-offs. Parameter 14: $\max(x_{n_e}) - \min(x_{n_e})$, the amplitude range of channel n in epoch e

3.2.4.4. Fourth, in order to detect electrical drift, the deviation of the mean amplitude in the epoch for each channel from the whole-channel mean amplitude was calculated. Parameter 15: $\left\langle x_{n_e} \right\rangle - \left\langle x_n \right\rangle$, the deviation from the channel average of channel n in epoch e

3.2.5. Contaminated datasets

After each file had been processed, a grand average dataset was created (i.e. all subjects' data were aggregated) so that each epoch was the ERP of a processed file. In a typical EEG study there are often subjects whose data are contaminated by artifacts to the extent that their data are not a true reflection of neural processes, and therefore distort the grand average data. These subjects' data are often removed entirely from the grand average. In order to identify these subjects, the epoch artifact detection method was repeated for the grand average:

3.2.5.1. Parameter 16: $\left\langle \max(x_{n_e}) - \min(x_{n_e}) \right\rangle_N$, the amplitude range in epoch e.

3.2.5.2. Parameter 17: $\left\langle S^2_{x_{n_e}} \right\rangle_N$, the variance in epoch e.

3.2.5.3. Parameter 18: $\left\langle \left\langle x_{n_e} \right\rangle - \left\langle x_n \right\rangle \right\rangle_N$, the deviation from the channel average in epoch e.

An additional parameter – the maximum absolute value of the EOG channels in the ERP – was computed for each epoch in order to determine whether eye movement artifacts remained.

3.2.5.4. Property 19: $\max(x_{EOG_{1,2,3,4e}})$, the maximum value in the EOG channels in epoch e.

Thresholds were calculated for each parameter, and any epoch (i.e. subject) that surpassed that threshold was considered contaminated and removed from the grand average file.

3.3. SCADS

The SCADS method was implemented following the description in (Junghöfer et al., 2000): data were converted to a single reference (Fz, as this was the reference chosen for data using the FASTER method). Editing matrices were composed, consisting of the maximum, standard deviation, and maximum gradient of each channel in each epoch. Limits were computed using the non-parametric formula to detect contaminated channels and channels in epochs. Contaminated channels were interpolated using spherical splines. Data were converted to average reference. Editing matrices were recomposed. At this stage, Junghöfer recommends visual identification of remaining artifacts in channels within epochs. To make this a fully automated process and thus a valid comparison with the FASTER method, this visual identification was substituted with the 3 Z-score threshold used by the FASTER method. Epochs with 10 or more contaminated channels were deemed bad, as suggested by Junghöfer. They were removed from the dataset. Remaining contaminated channels in epochs were interpolated using spherical splines.

3.4. Supervised, expert artifact identification

Contaminated epochs, contaminated channels and artifactual ICs were identified by seven trained individuals, all of whom had extensive EEG processing experience and who were currently engaged in high-density EEG research. The raw data were filtered and epoched, and the participants were then instructed to list the epochs and channels they believed were artifact-contaminated. This was done by visual identification, using EEGLAB software, which allows users to scroll through EEG data. They were also instructed to list any datasets they believed to be contaminated to the extent that artifact removal would not improve the quality of the data. Such datasets were not used to calculate the average. In the remaining datasets, the listed artifacts were removed, and ICA was performed on the data. Participants were then instructed to list any ICs which they believe were artifact-contaminated, which were subtracted from the data. Each participant analyzed between 4 and 9 datasets each.

4. Artifact removal quantification

The simulated data results were calculated from 47 files. We attempted to quantify the performance of each method by using metrics of *sensitivity* and *specificity*. The first, sensitivity, was the percentage of true artifacts that were detected. For example, if there were 100 artifact-contaminated epochs in a dataset, and 50 of these contaminated epochs were detected then the sensitivity would be 50%. The second, specificity, is the percentage of artifact-free EEG data that was wrongly classified as contaminated by artifacts. For example, if there were 100 artifact-free epochs and 5 of these were wrongly classified as contaminated then the specificity would be 95%. The formulae for artifact detection were calculated using the following formulae: sensitivity = detected artifacts/total contaminated data; specificity = 1 − (detected false artifacts/total uncontaminated data). Fig. 3 displays an example of the effect of different levels of sensitivity.

In order to quantify the performance of the artifact removal method on real data, a metric for measuring artifact contribution to EEG was defined as the variance in the 500 ms baseline prior to the
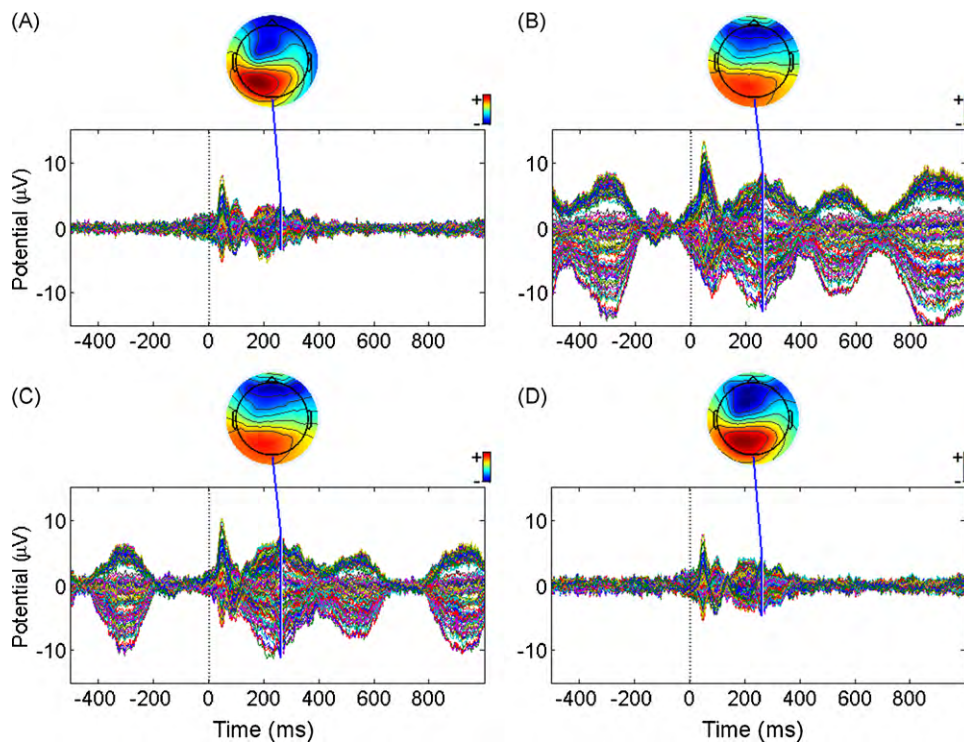
Fig. 3. The ERPs calculated from simulated data with a different amount of artifacts present. (A) The ERP from the dataset before artifact addition. (B) The ERP after addition of all artifact types (see text for details). (C) The ERP at 50% sensitivity. (D) The ERP after complete removal of all artifacts (i.e. a 100% sensitivity).

stimulus presentation (Handy et al., 2003). This variance should be lower in datasets with fewer artifacts. This metric has the advantage of making no assumptions about the signal content. The disadvantage is that a signal with lower amplitude has lower variance, and therefore any method that removed all data would have perfect performance. Therefore, P3 amplitude was also reported, in order to demonstrate that the EEG signal was not being attenuated. The percentage of epochs removed was also recorded, as a measure of data retention. These measures were compared for each processing method.

Type 1 error probability was set at .05, $p_{corrected}$ was set using Bonferroni correction for multiple comparisons adjusted for correlation among dependent variables (see Appendix A).

5. Results

5.1. Simulated data

Due to the non-normal distribution of the detection rates, Wilcoxon Signed Ranks tests were conducted. Table 1 displays the results of the simulated data analysis.

5.1.1. Channel artifacts

Wilcoxon Signed Rank tests showed that there was a significantly higher channel detection sensitivity for the SCADS method compared to FASTER using 128 channels ($Z = 2.701$, $p < 0.01$). This significance was not present using 64 or 32 channels. Analysis of the channel detection specificity of each method, shows that FASTER had a significantly higher specificity than SCADS using 128-, 64- and 32- channel datasets ($Z = 5.375$, $Z = 5.517$, $Z = 5.907$, respectively, $p < 0.001$).

5.1.2. Epoch artifacts

Wilcoxon Signed Rank tests showed that there was a significant improvement in the performance of the FASTER method over SCADS for epoch detection sensitivity using 128-, 64- and 32- channels ($Z = 5.381$, $Z = 5.408$, $Z = 5.315$, respectively, $p < 0.001$). SCADS had a significantly higher specificity than FASTER using 128-, 64- and 32- channel datasets ($Z = 5.304$, $Z = 5.304$, $Z = 5.304$, respectively, $p < 0.001$).

5.1.3. Eye and muscle movement artifacts

A Wilcoxon Signed Rank test showed significant improvement in EOG removal percentage using the FASTER method compared

Table 1
The percentage of artifacts removed for each method and array size.

Method	No. channels	Percentage of artifacts removed						
		Channels Sensitivity, specificity (%)	Epochs Sensitivity, specificity (%)	EOG (%)	EMG (%)	Discontinuities (%)	Linear trends (%)	White noise (%)
FASTER	128	94.47, 98.96	60.24, 97.53	98.99	99.91	82.97	93.04	92.93
	64	97.02, 98.48	61.83, 97.54	99.07	99.19	94.68	97.02	95.30
	32	5.88, 96.81	58.64, 97.49	97.64	94.17	93.61	95.59	91.29
SCADS	128	100, 96.42	0.15, 99.99	11.69	99.73	79.43	91.19	98.13
	64	100, 94.10	0.15, 99.99	14.03	98.27	87.43	93.00	96.59
	32	5.88, 84.50	0.15, 99.99	18.31	95.36	83.33	93.22	96.53

Table 2
The median baseline variance and mean amplitude and the % of Staepochs removed for each method and array size.

Method	No. channels	Baseline variance (μV)	P3 amplitude (μV)	Epochs removed (%)
FASTER	128	0.193	7.417	3.2
	64	0.207	7.746	3.1
	32	0.208	8.387	3.3
SCADS	128	0.252	6.990	2.6
	64	0.241	6.931	2.4
	32	0.217	7.732	2.4
Supervised	128	0.321	6.885	3.3
Raw	128	0.575	14.943	–
	64	0.745	26.481	–
	32	0.538	12.263	–
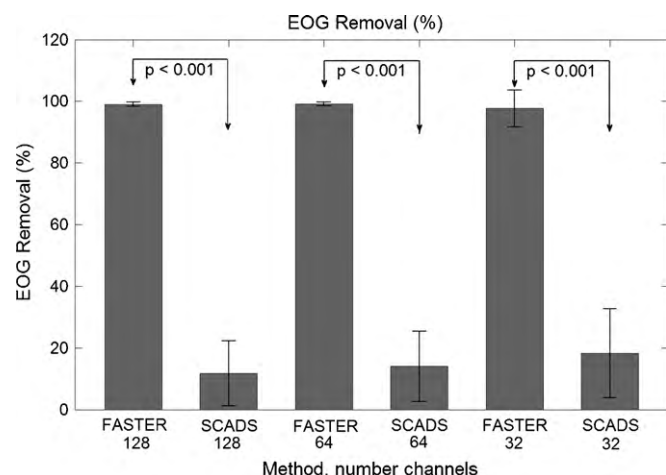
Fig. 4. A comparison of EOG removal percentages between FASTER and SCADS. Error bars represent standard error of the mean.

Fig. 5. The baseline variance of different array sizes between FASTER, SCADS and supervised. Values shown are median and error bars display the interquartile range.

with the SCADS methods for datasets of 128, 64 and 32 channels ($Z = 5.905$, $Z = 5.841$, $Z = 5.754$, respectively, $p < 0.001$). Bonferroni correction for 3 comparisons was used. A Wilcoxon Signed Rank test showed no significant difference in removal percentage between FASTER and SCADS methods for simulated EMG artifacts. Fig. 4 displays the percentage of eye movement artifacts that were removed for each method.

5.1.4. Single-channel, single-epoch artifacts

For linear trends, a Wilcoxon Signed Rank test showed a significant improvement in removal percentage using the FASTER method compared with the SCADS methods for the 64-channel array only ($Z = 3.330$, $p < 0.005$). For white noise, a Wilcoxon Signed Rank test showed a significantly lower removal percentage using the FASTER method compared with the SCADS method for datasets of 128 and 32 channels ($Z = 2.691$, $p < 0.05$, $Z = 3.02$, $p < 0.01$, respectively). For discontinuities, a Wilcoxon Signed Rank test showed no significant difference in removal percentage between FASTER and SCADS methods.
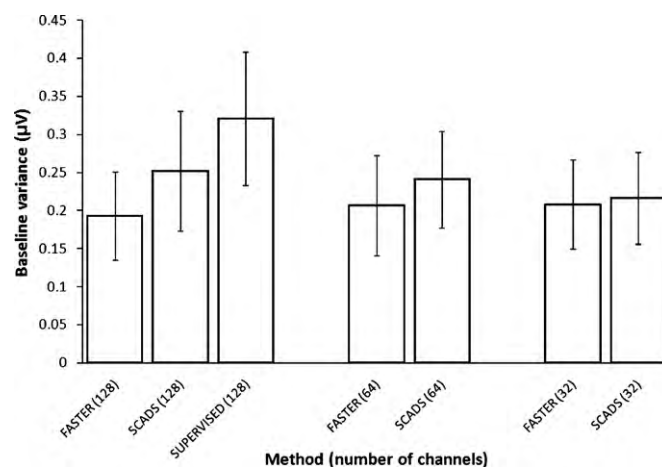
5.2. Real data

Table 2 displays the results from the analysis of the real data.

5.2.1. Baseline variance

To compare baseline variances across methods, the median value was taken across channels for each dataset. These data are displayed in Fig. 5.

A Shapiro–Wilk test of normality showed that the baseline variances were non-normally distributed. Therefore, non-parametric

statistics were employed. A Friedman test comparing baseline variances of FASTER, SCADS and supervised for 128 channels was significant ($p < .001$). Follow up Wilcoxon tests showed that the baseline variances for FASTER were significantly lower than either SCADS or supervised ($Z = 4.85$, $p < .001$; $Z = 3.55$, $p < .001$, respectively). The baseline variance for SCADS was significantly lower than for the supervised method ($Z = 4.31$, $p < .001$). There were no significant differences in baseline variance between FASTER and SCADS when 64 channels were analyzed, nor were there differences between FASTER and SCADS when 32 channels were analyzed.

In order to compare the effect of array size within methods, we compared baseline variances of the common channels for each array size. Alpha was set at $p_{corrected} = .0418$. The FASTER method showed no significant differences in baseline variances depending on the number of channels processed. The SCADS method had a lower baseline variance using 64 channels vs. 128 channels ($Z = 3.77$, $p < .0418$), when using 32 channels vs. 128 channels ($Z = 3.42$, $p < .0418$) and using 32 channels vs. 64 channels ($Z = 2.13$, $p < .0418$).

5.2.2. P3 amplitude

Shapiro–Wilk tests of normality showed that the amplitude data were normally distributed, and therefore parametric statistics were used. FASTER produced a P3 with a significantly higher amplitude compared to supervised processing, with alpha set at $p_{corrected} = 0.0446$ ($t(38) = 2.201$, $p_{corrected} < 0.0446$). Paired t-tests (alpha set at $p_{corrected} = 0.0411$) showed a significantly higher P3 amplitude for FASTER vs. SCADS using 64 electrodes ($t(42) = 2.773$, $p_{corrected} < 0.0411$), and significantly higher P3 latencies (alpha set at $p_{corrected} = 0.0371$) for FASTER for 128, 64 and 32 electrodes ($t(42) = 2.803$, $t(42) = 2.642$, $t(43) = 3.684$, respectively,
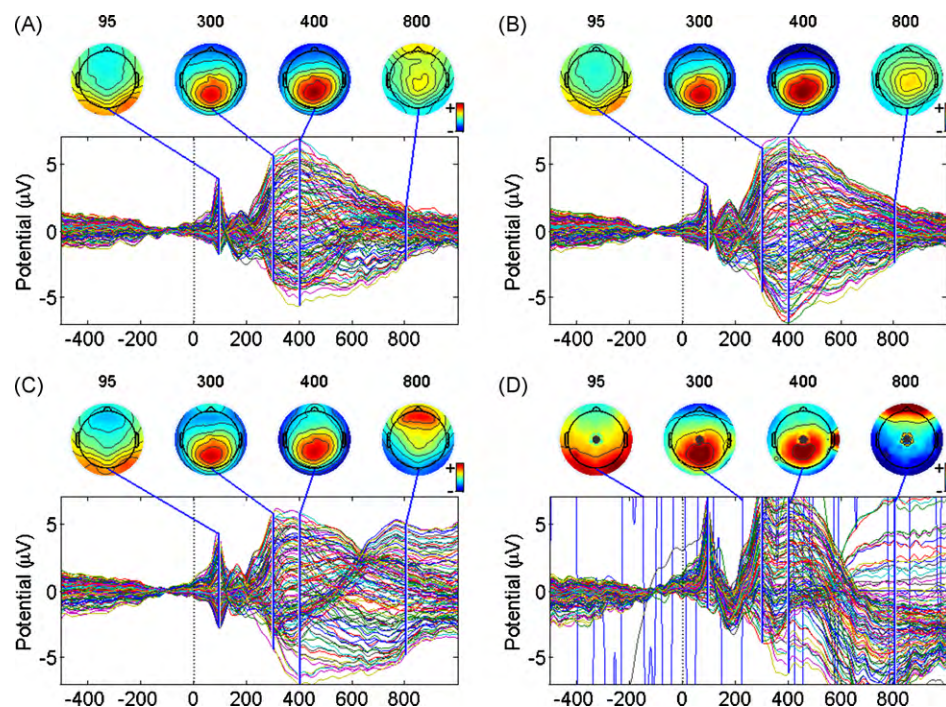
Fig. 6. Butterfly plots of ERPs after various processing methods. (A) Supervised processing. (B) FASTER processing. (C) SCADS processing. (D) Filter only. The late frontal component present in C is due to uncorrected EOG artifacts. A 30 Hz low-pass filter was applied for display only.

$p_{corrected}$ <.0371). Fig. 6 displays butterfly plots of ERPs after various processing methods and Fig. 7 displays the ERP to the target stimulus at Pz.

5.2.3. Contaminated subject removal

From a total of 47 subjects, FASTER identified 4 subjects as contaminated in the 128-channel dataset, 4 in the 64-channel dataset, and 2 in the 32-channel dataset. In order to test the effect of identifying contaminated subjects in smaller sample sizes, sub-sets of the data (ranging from 10 to 47 subjects, with random inclusion of subjects) were generated. The composition of each sub-set was shuffled 100 times and the number of contaminated subjects was recorded for each sample size. Fig. 8 displays these data.

6. Discussion

The aim of this study was to quantify the utility of FASTER – a fully automated statistical thresholding method for EEG artifact rejection, which also incorporates ICA. In order to quantify the per-

formance of FASTER simulated data were analyzed using FASTER and a variant of SCADS, which is a similar statistical method of artifact detection. Furthermore, real data were analyzed using FASTER, SCADS, and by supervised detection. FASTER was also tested across different numbers of scalp electrodes (128, 64 and 32). The results of the analysis of the simulated data showed that FASTER had generally high sensitivity and specificity for detection of artifacts. FASTER had >90% sensitivity and specificity for detection of contaminated channels, eye and muscle movement artifacts, linear trends and white noise. SCADS had a significantly higher sensitivity to channel artifacts than FASTER using 128 channels. However, specificity was significantly higher for FASTER vs. SCADS across array sizes. FASTER generally had >60% sensitivity and specificity for detection of contaminated epochs, versus 0.15% for SCADS. ERP amplitude was significantly high for FASTER versus the supervised approach. FASTER was effective in attenuating artifacts in 128-, 64-, and 32-scalp electrode arrays.

One assumption of FASTER is that uncontaminated EEG parameter distributions should be distributed normally, or approximately normally. This assumption may be violated if low numbers of
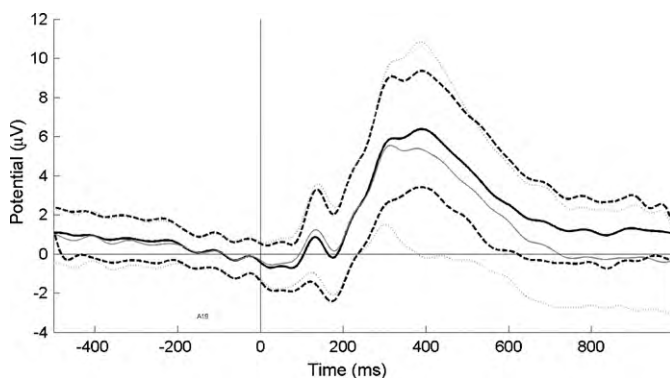

Fig. 7. ERP to the target stimulus at Pz. The black line is after FASTER processing, the gray line is after SCADS processing. Dashed lines represent standard deviation. A 15 Hz low-pass filter was applied for display only.
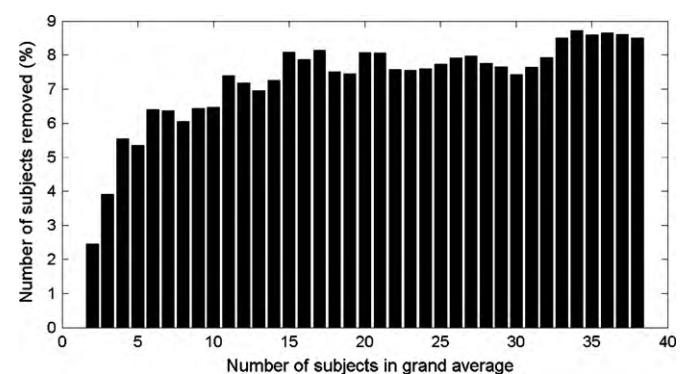

Fig. 8. The number of subjects removed by FASTER as a function of number of subjects present.

data points are employed. The results from the real and simulated datasets show that FASTER worked effectively on datasets with 128-, 64- and 32-scalp electrodes. However, the sensitivity for contaminated channels decreased dramatically, from 94.47% for 128-scalp electrodes and 97.02% for 64-scalp electrodes to 5.88% for 32-scalp electrodes. This may be a result of using simulated data (where the artifact distribution is normal), as the number of contaminated channels was lower in the 32-electrode datasets, and a result of using statistical thresholding, as the threshold calculation was based on a smaller sample size and consequently variance among channels was increased. This issue also occurred with the SCADS method. The performance on the detection of other artifacts, such as eye and muscle movement artifacts, was largely unaffected by the number of channels. The SCADS method had a 100% sensitivity to channel artifacts for 128- and 64-electrode datasets, but its specificity dropped as the number of channels decreased, from 96.42% for 128-scalp electrodes to 84.91% for 32-scalp electrodes. This resulted in an average of 4 uncontaminated channels being interpolated for 32-electrode sets.

We did not test FASTER on datasets with fewer than 32 channels. The number of components returned from an ICA decomposition is equal to or less than the number of channels used: hence, component independence may decrease as the number of electrodes decreases. It is possible that FASTER may be effective at removing artifacts at lower electrode densities, but if so used, consideration must be taken that components may also contain both artifact and EEG data, and so some EEG data may be lost. Similarly, the performance of the channel detection is unknown.

A novel feature of FASTER is that subjects with consistently poor data are rejected from the grand average on objective criteria. This method may be optimal for studies with large numbers of subjects (as outliers are more easily identified). However, FASTER can detect contaminated subjects in samples with 12 or more subjects.

FASTER generally outperformed modified SCADS on our data (both real and simulated). We analyzed our data using SCADS in order to provide a comparison between statistical thresholding methods. It should be noted, however, that SCADS was not designed with unsupervised full automation as a goal. It is likely that SCADS would have performed better had some element of supervision been involved. A key feature of FASTER was the application of a statistical thresholding method to ICA data, and the use of ICA to subtract the EOG contribution to the EEG data. It seems that the combination of statistical thresholding and ICA may result in an effective method for artifact detection.

The present set of benchmarking tests were conducted with data from healthy subjects on an oddball task (designed to evoke a P3 ERP). It remains to be seen how FASTER will generalize to data from psychiatric or neurologic patients, older adults, or from children. In addition, future work could examine FASTER with data from other ERP paradigms, such as mismatch negativity (MMN), which typically have smaller amplitudes than P3. Paradigms such as MMN tend to have shorter epoch lengths, which may also influence the ability of FASTER to detect artifacts. Initial data from our laboratory suggest that FASTER can effectively analyze MMN data from patients with Parkinson's disease.

The computational requirements of the FASTER method are generally low, with the exception of the ICA decomposition, which can take some time. In this study, we used the Infomax ICA algorithm, processed under 64-bit MATLAB, and the entire processing protocol took up to an hour per dataset on a 64-bit dual-core machine running *Linux Ubuntu*, of which approximately 40 min was the ICA decomposition time. Using a different ICA algorithm, for example the FastICA algorithm, and/or a compiled binary implementation of the method would increase the speed.

A key benefit of FASTER is that its use does not require the user to have any knowledge of signal processing. The user can simply select the folder containing the raw data and FASTER will output an artifact-attenuated, grand-averaged dataset. Alternatively, researchers who prefer to process EEG data using their favored approach may choose to use FASTER to conduct a first pass of their data. Furthermore, given the diversity of signal-processing methods applied to EEG data (e.g., different criteria for artifact rejection), the use of FASTER may help standardize the approach to EEG analysis. It is worth noting that the baseline variance of supervised processed data was itself quite variable (nearly twice as high as FASTER), indicating that even researchers experienced in EEG analysis can vary in their degree of artifact rejection. Our intention is to integrate FASTER into the EEGLAB (Delorme and Makeig, 2004) processing software as a plugin, with the source code for FASTER freely available. EEGLAB is a popular, free, software tool that is already used by many researchers. The default settings in this plugin will correspond to the values employed in the present study. However, the user will be able to manipulate any of the settings described in this manuscript via a graphical user interface. This plugin will include the option to display the baseline variance of each dataset. Further work on FASTER could include the addition of artifacts for removal of further artifact types, e.g., MRI artifacts.

The aim of this paper was to demonstrate a fully automated, unsupervised method for EEG artifact detection and rejection. The data from this study show that FASTER can reliably detect artifacts commonly found in EEG data and may even be superior to supervised detection for some metrics. It is hoped that FASTER will be of use to the EEG research community.

Appendix A.

- The Hurst exponent is a measure of long-range dependence within a signal, with a range of 0–1. It can be thought of as the tendency of a signal to promote trends or not. Discussions of its estimation can be found in Bardet et al. (2003) and Mielniczuk and Wojdy (2007). Human phenomena such as EEG have values of $H \approx 0.7$, and so it can be used as a measure for detecting non-biological EEG data (Vorobyov and Cichocki, 2002; Bian et al., 2006). To detect channels with a large amount of non-biological noise, Hurst exponents were estimated using MATLAB's *wfmbesti* function (the discrete second order derivative estimate was used).
- The equation used for Bonferroni correction with correlated dependent variables was:

$$\alpha_{\text{new}} = \alpha_{\text{old}} \left(\bar{r} + \frac{1 - \bar{r}}{N} \right)$$

where α_{new} was the new effective significance value; α_{old} was the original significance value (e.g. 0.05); $\bar{r}$ was the mean Pearson correlation coefficient among all dependent variables; N was the number of comparisons made.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jneumeth.2010.07.015.

References

Bardet J-M, Lang G, Oppenheim G, Philippe A, Stoev S, Taqqu MS. Semi-parametric estimation of the long-range dependence parameter: a survey. In: Birkhäuser, editor. Theory and applications of long-range dependence; 2003. p. 557–77.

Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comp 1995;7:1129–59.

Bian N-Y, Wang B, Cao Y, Zhang L. Automatic removal of artifacts from EEG data using ICA and exponential analysis. Adv Neural Networks – ISNN 2006;2006:719–26.

Croft RJ, Barry RJ. Removal of ocular artifact from the EEG: a review. Clin Neurophys 2000;30:5–19.

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

Delorme A, Sejnowski TJ, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 2007;34:1443–9.

Handy TC, Grafton ST, Shroff NM, Ketay S, Gazzaniga MS. Graspable objects grab attention when the potential for action is recognized. Nat Neurosci 2003;6:421–7.

Jung T-P, Humphries C, Lee T-W, Makeig S, McKeown MJ, Iragui V, et al. Extended ICA removes artifacts from electroencephalographic recordings. In: Proc 1997 conf on Adv. in neural info. proc. systems 10. Denver, CO, United States: MIT Press; 1998.

Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 2000;37:163–78.

Junghöfer M, Elbert T, Tucker DM, Rockstroh B. Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology 2000;37:523–32.

Mielniczuk J, Wojdy P. Estimation of Hurst exponent revisited. Comput Stat Data Anal 2007;51:4510–25.

Moretti DV, Babiloni F, Carducci F, Cincotti F, Remondini E, Rossini PM, et al. Computerized processing of EEG-EOG-EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials. Int J Psychophysiol 2003;47:199–216.

Onton J, Makeig S, Christa N, Wolfgang K. Information-based modeling of event-related brain dynamics. Prog Brain Res 2006:99–120.

Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophys 2001;112:713–9.

Schlögl A, Keinrath C, Zimmermann D, Scherer R, Leeb R, Pfurtscheller G. A fully automated correction method of EOG artifacts in EEG recordings. Clin Neurophys 2007;118:98–104.

Shlens J. A tutorial on principal component analysis; 2005.

Vorobyov S, Cichocki A. Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. Biol Cybern 2002;86:293–303.

Yandong L, Zhongwei M, Wenkai L, Yanda L. Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. Physiol Meas 2006:425.