Post-print version (peer reviewed, author accepted manuscript), please cite the published article as follows:

da Cruz, J. R., Chicherov, V., Herzog, M. H., & Figueiredo, P. (2018). An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. Clinical Neurophysiology, 129(7), 1427-1437. DOI: https://doi.org/10.1016/j.clinph.2018.04.600

An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics

Janir Ramos da Cruz[1,2*], Vitaly Chicherov[2], Michael H. Herzog[2], Patrícia Figueiredo[1]

[1] Institute for Systems and Robotics – Lisbon (LARSys) and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Portugal

[2] Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Switzerland

**Corresponding author*

Email: janir.ramos@epfl.ch

Highlights

- A novel automatic pre-processing pipeline for both resting state and evoked EEG data is proposed.
- The proposed automatic pipeline is tested in both clinical and healthy populations.
- The proposed automatic pipeline is as reliable as pre-processing by EEG experts.

Abstract

Objective

With the advent of high-density EEG and studies of large numbers of participants, yielding increasingly greater amounts of data, supervised methods for artifact rejection have become excessively time consuming. Here, we propose a novel automatic pipeline (APP) for pre-processing and artifact rejection of EEG data, which innovates relative to existing methods by not only following state-of-the-art guidelines but also further employing robust statistics.

Methods

APP was tested on event-related potential (ERP) data from healthy participants and schizophrenia patients, and resting-state (RS) data from healthy participants. Its performance was compared with that of existing automatic methods (FASTER for ERP data, TAPEEG and Prep pipeline for RS data) and supervised pre-processing by experts.

Results

APP rejected fewer bad channels and bad epochs than the other methods. In the ERP study, it produced significantly higher amplitudes than FASTER, which were consistent with the supervised scheme. In the RS study, it produced spectral measures that correlated well with the automatic alternatives and the supervised scheme.

Conclusion

APP effectively removed EEG artifacts, performing similarly to the supervised scheme and outperforming existing automatic alternatives.

Significance

The proposed automatic pipeline provides a reliable and efficient tool for pre-processing large datasets of both evoked and resting-state EEG.

Keywords

Electroencephalography, automatic pre-processing, ERP, resting-state

1. Introduction

The electroencephalogram (EEG) is a non-invasive tool for the investigation of human brain function, which has been continuously used for almost one century (Niedermeyer and Lopes da Silva, 2005). However, EEG data are typically contaminated with a number of artifacts. Artifacts are undesired signals that may affect the measurement and change the EEG signal of interest. These artifacts may arise from non-physiological noise sources that originate outside the participant, such as the grounding of the electrodes causing power line noise at 50/60 Hz and at its harmonics, interferences with other electrical devices, or imperfections in electrode settling. Artifacts may also arise from physiological noise sources originating within the participants, such as the ones produced by head, eye, or muscle movements (Urigüen and Garcia-Zapirain, 2015). Head movements may result in spikes and discontinuities due to a rapid change of impedance at one or several electrodes. Reflective eye movements occur frequently and are normally picked up by the frontal electrodes in the frequency range of 1-3 Hz (within the delta wave range). Blinking also contaminates the EEG signal, usually causing a more abrupt change in its amplitude than eye movements. Finally, every movement of the participant generates muscular artifacts that can be found everywhere on the scalp at frequencies higher than 20 Hz (within the beta and gamma waves range).

One simple way to deal with these artifacts is to remove segments of the data that exceed a certain level of artifact contamination, for example, signal amplitudes greater than ± 100 μV. However, this coarse approach may lead to the loss of a great amount of data that could still contain artifact-free information, therefore potentially compromising the subsequent analysis and interpretation of the data. This is true for both evoked-related potentials (ERP) and resting-state (RS) signal fluctuations. Moreover, since participant generated artifacts may overlap in the spectral domain, and on many EEG channels, with the signal of interest, simple spatial and frequency band filtering approaches may be inefficient to remove this kind of artifacts (Tatum et al., 2011). Another method that is commonly used to clean-up EEG data is independent component analysis (ICA; Makeig et al., 1996). Assuming that neuronal signals and noise recorded on the scalp are independent of each other, then the EEG signal can be described by their linear summation. The ICA is used to decompose the EEG data in

statistically independent sources (ICs), so as to separate the neuronal and noise contributions to the signal. The artifactual ICs can then be identified and subsequently subtracted from the EEG data, yielding an artifact-free signal.

Usually, pre-processing of EEG data, including the classification of artifactual ICs, is performed under expert supervision. However, with the advent of both high-density EEG arrays (64-256 channels) and studies of large populations, yielding increasingly greater amounts of data, supervised methods have become excessively time consuming. To cope with this, and to minimize subjectivity, automatic methods have recently been presented (Abreu et al., 2016a, 2016b; Bigdely-Shamlo et al., 2015; Hatz et al., 2015; Nolan et al., 2010). Fully automated statistical thresholding for EEG artifact rejection (FASTER; Nolan et al., 2010), for instance, enables a fully automated pre-processing of ERP data, based on computing z-scores of different signal metrics, and threshold them in order to detect bad channels, bad epochs and artifactual ICs. Tool for automated processing of EEG data (TAPEEG; Hatz et al., 2015) uses a similar approach for the automatic pre-processing of RS EEG data. However, because they are based on z-scores, these approaches are not robust to outliers and as a consequence they tend to have high rejection rates of artifact-free signal. A more promising approach is to use robust statistics instead. For example, the Prep pipeline (Bigdely-Shamlo et al., 2015) provides an automatic pre-processing pipeline including filtering and bad channels identification using the RANSAC (random sample consensus) algorithm. However, in this case the identified bad channels are assumed to be globally bad. Thus, if a channel contains artifactual periods, these are neglected and left in the pre-processed EEG data. Moreover, supervised inspection of pre-processed data for bad epochs is necessary since the Prep pipeline does not provide this feature.

Here, we present APP, a novel Matlab® based fully automatic pipeline for pre-processing and artifact rejection of EEG data (including both ERP and RS data), which is based on state-of-the-art guidelines for EEG pre-processing, ICA decomposition, and robust statistics. APP consists of: 1) high-pass filtering; 2) power line noise removal; 3) re-referencing to a robust estimate of the mean of all channels; 4) removal and interpolation of bad channels; 5) removal of bad epochs; 6) ICA to remove eye-movement, muscular and bad-channel related artifacts; and 7) removal of epoch artifacts. At each step of the pipeline, a number of relevant parameters are estimated from the data and outliers are detected based on a robust data-driven outlier detection scheme.

APP was tested on ERP data from 61 healthy participants and 44 schizophrenia patients performing a visual discrimination task, and on RS data from 68 healthy participants. The inclusion of patient data in the validation of APP is of particular interest since one of the primary applications of EEG is the study of clinical populations. Furthermore, many of these populations, schizophrenia patients in particular, are known to produce more artifacts than healthy volunteers, which is a challenge to automatic pre-processing. We compare APP to three state-of-the-art automatic artifact removal methods, FASTER, TAPEEG, and Prep pipeline, which have shown to be effective at removing a wide range of EEG artifacts. We also compared APP with supervised artifact removal by experts using the CARTOOL software (Brunet et al., 2011).

2. Methods

The proposed pre-processing and artifact removal method APP is first described, including a detailed description of each step. Then, the artifact removal methods FASTER, TAPEEG, and Prep pipeline, as well as the supervised artifact removal by experts, against which APP is compared, are described. Finally, the data acquisition and analysis methods used to validate the proposed method are presented.

2.1. Proposed EEG pre-processing and artifact removal method - APP

The APP pipeline consists of the following steps, which are described in detail in the respective sub-sections below:

1) High-pass filtering, to eliminate signal low-frequency non-stationarity (for example, slow drifts in the mean);
2) Removal of power line noise, with minimum distortion;
3) Robust re-referencing, to a robust estimate of the mean of all channels;
4) Detection, removal and interpolation of bad channels;
5) Detection and removal of bad epochs;
6) Detection and removal of eye movement-, muscular-, and bad channel-related artifacts based on ICA;
7) Detection, removal and interpolation of bad channels in epochs.

At every step, a number of relevant parameters are estimated from the data, and outliers are detected based on a scheme described below in the sub-section Outlier detection. Each step of APP is described in detail below.

2.1.1. High-pass filtering

The changes of the DC value of the EEG over time, called DC drift, can be corrected by high-pass filtering. In APP, in order to counteract the phase delay and to minimize distortion that are introduced by high-pass filtering (Lynn, 1989), we use a zero-phase 3rd order Butterworth filter, run both in forward and reverse directions (Picton et al., 2000). Usually, a cutoff frequency of 0.1 Hz suffices to remove these slow voltage shifts. However, when obtaining recordings from children or patients, which normally have excess body and head movements, which are a common source of sustained shifts in voltage, higher cutoff frequencies are advised (Luck, 2014). In order to make APP more robust to this kind of noise we use a 1 Hz cutoff frequency.

2.1.2. Power line noise removal

The removal of the power line interference (50/60 Hz noise and its harmonics) is usually accomplished using notch filters. However, these usually lead to several artifacts such as significant signal distortions around the notch filter frequencies, as well as phase distortions. In order to avoid such artifacts, in APP we apply the CleanLine plugin of EEGLAB which uses a multi-taper regression method with a Thompson F-statistics for identifying sinusoidal power line noise from ICs (Delorme and Makeig, 2004; Mullen, 2012). In Figure 1, an illustrative example of the EEG power spectrum before and after applying the CleanLine method, as desired the power line interference and its harmonics are greatly attenuated without distortion at the neighbor frequencies.

Figure 1 – Insert approximately here

2.1.3. Robust re-referencing

In most EEG amplifiers, the full common-mode rejection ratio can only be achieved after re-referencing. Ideally, the average signal across mastoid and earlobes electrodes would be used as the reference, since they are close to the EEG electrodes but record less brain activity. However, many researchers do not record from mastoids nor earlobes. Instead, the Cz and FCz electrodes are typically used as references, since they

do not introduce lateralization bias in the data. Nevertheless, in several experiments (e.g., when the response errors are of interest), these two electrodes are a poor choice of reference since the main effect is expect around those locations. A good compromise, when the number of electrodes is large enough (> 32 electrodes), is to use the mean of all electrodes as the reference (Nunez et al., 1997).

The main disadvantage of using the channel average as the reference is that it is not resistant to outliers, namely those associated with bad channels and eye blinks. In the presence of such artifacts, the noise introduced in the reference will spread to the other scalp electrodes. In order to minimize this effect, in APP, we use the biweight estimates of the mean as an approximation of the true mean of the channels, since it offers high resistance to outliers and low sampling variability (Hoaglin et al., 1983, 1985). The biweight estimation is accomplished in two steps: first, the median and median absolute deviation (MAD) are used to assign a zero weight to extreme values; and second, a weighted mean is calculated by assigning decreasing weights nonlinearly to zero as going away from the center of the distribution (For more details, see Appendix).

To demonstrate the behavior of the biweight estimate of the mean of the channels, we depict two illustrative cases in Figure 2, one without bad channels and another with one bad channel. In the first case, the biweight mean behaves similarly to the ordinary mean, with the two means having a high correlation. We also see that the biweight mean is less affected by the eye blinks (as captured by the spike like behavior in the waveform, which affects mainly the frontal electrodes). However, as we introduce the bad channel the biweight mean remains nearly unchanged while the mean of the electrodes is severely contaminated, which results in a decrease in the correlation between the two.

It should be noted that, although we could have performed re-referencing after steps 4 and 5 to avoid distribution of noise across all channels, early re-referencing is important to increase the common-rejection ratio and consequently the SNR, which improves the performance of the metrics used to detect bad channels (correlation and dispersion criteria) as well as bad epochs (mean global field power and mean deviation from channel biweight estimate of the channel means). Although such early re-referencing can cause problems, namely by distributing noise to all channels, we demonstrated that using the biweight estimate of the mean circumvents this problem.

Figure 2 – Insert approximately here

2.1.4. Detection, removal and interpolation of bad channels

One of the most common artifacts in EEG recordings is the presence of bad channels, which often result from poor contact between the scalp and the electrode. In APP, the bad channel detection is accomplished in two stages, based on: 1) temporal features; and 2) ICA decomposition, described below in the sub-section Detection and removal of artifacts based on ICA. The bad channel detection based on temporal features follows two criteria:

1) *Correlation criterion*: Due to volume conduction, the EEG signal recorded from each channel is highly correlated with the EEG signals from neighboring channels. Therefore, we compute the Pearson correlation between each EEG channel and all other channels, and take the mean of the 4 highest correlation coefficients. Under the high correlation assumption, bad channels are classified as the ones whose mean correlation departs significantly from the distribution of the mean correlation across channels.
2) *Dispersion criterion*: As a result of the poor contact, bad channels are more prone to additive Gaussian noise, exhibiting data dispersions spread over a larger range compared with good channels. To quantify the dispersion of the data in each channel, we use the biweight estimate of the standard deviation (see Appendix). The rationale of using such robust measure is owing to the frontal electrodes being more susceptible to eye movement artifacts, which inflate their ordinary standard deviation. Under the high dispersion assumption, bad channels are classified as the ones whose biweight estimate of the standard deviation departs significantly from the respective distribution across channels.

A threshold on the number of bad channels is set to 5%, as suggested by (Picton et al., 2000). If the number of bad channels detected is above the threshold, the dataset is flagged and the pre-processing is stopped. If the number of bad channels is below the threshold, bad channels are removed. APP uses the spherical spline method to interpolate the missing data based on the neighboring channels (Perrin et al., 1989). It has been shown by Fletcher and colleagues, in an extensive study of interpolation errors in scalp topographic mapping that the spline class algorithms minimize interpolation errors (Fletcher et al., 1996).

2.1.5. Detection and removal of bad epochs

ICA decomposition can readily isolate in a few ICs typical EEG artifacts such as eye movements, bad channels and muscle tension, since these have stable scalp projections (as described below in sub-section Detection and removal of artifacts based on ICA). However, other kinds of artifacts affecting many or all electrodes simultaneously, such as head movements, are reflected in multiple (maybe hundreds of) ICs, with only a few ICs remaining to capture neuronal activity (Onton et al., 2006). Therefore, before submitting the EEG data to ICA, APP prunes the EEG data for epochs that deviate from the typical epoch ranges. In case of RS analysis, the data is split into epochs of 2 s, while for ERP analysis the data are epoched according to the respective event trials. To detect the bad epochs, APP computes two parameters:

1) *Mean global field power*: Electrode movements can inflate the impedance between electrodes and the scalp. This causes an increase in the electrode voltage offset resulting in high amplitudes across the scalp. To quantify the overall scalp EEG strength for each epoch, APP uses the mean of the global field power (GFP), which is defined as the standard deviation of the EEG signal across all electrodes at a given time point (Lehmann and Skrandies, 1980):

$$\boldsymbol{GFP(t)} = \sqrt{\frac{\sum_{i=1}^{n}\left(\boldsymbol{x_i(t)} - \bar{\boldsymbol{x}}\boldsymbol{(t)}\right)^2}{\boldsymbol{n}}} \qquad \textbf{Eq. 1}$$

where, $\boldsymbol{x}_i(t)$ is the signal at electrode i at time point t, $\bar{\boldsymbol{x}}(t)$ is the average signal across all electrodes at time t, and n is the number of electrodes.

2) *Mean deviation from channel biweight estimate of the channel means*: In certain cases, the electrodes movements do not produce high enough amplitudes to be detected using the mean GFP, although they still severely contaminate the respective epochs. In order to account for these, APP quantifies the mean deviation from the biweight estimate of the channel mean (MDCM) as:

$$\boldsymbol{MDCM(k)} = \langle\left|\bar{\boldsymbol{x}}_{\boldsymbol{i_k}}\right| - \bar{\boldsymbol{x}}_{\boldsymbol{biweight_i}}\rangle_{\boldsymbol{n}} \qquad \textbf{Eq. 2}$$

where, $\langle ... \rangle_n$ is the mean operation across all n channels, $\left|\bar{\boldsymbol{x}}_{i_k}\right|$is the absolute value of the mean data in channel i within epoch k and $\bar{\boldsymbol{x}}_{biweight_i}$ is the biweight estimate of the means of the data in channel i.

In the case of visual stimulation using short presentation durations (e.g. 30 ms), we further apply a threshold of ±100 μV to the vertical electrooculogram (EOG) signal within an interval of 100 ms before and after stimulus onset, in order to determine if the participant saw the stimulus or not. If the vertical EOG signal is above the threshold, the participant most likely blinked during stimulus presentation; in this case, the epoch is discarded. Moreover, in case the reaction times (RTs) are of interest, APP provides the option to discard responses that are too fast or too slow, according to thresholds specified by the user.

2.1.6. Detection and removal of artifacts based on ICA

EEG signals can be assumed as linear mixtures of electrical potentials originating from multiple brain sources propagating instantaneously to the scalp, as well as artifacts (Sarvas, 1987). Formally, this can be described as:

$$\boldsymbol{X} = \boldsymbol{WS} \qquad \textbf{Eq. 3}$$

where, $\boldsymbol{X} = [\boldsymbol{x}(1), \dots, \boldsymbol{x}(N)] = [\boldsymbol{x}_i(\mathrm{j})]_{n \times N}$ are the EEG signals recorded from the scalp, where each row is one channel, n is the number of channels and N is the number of time samples; $\boldsymbol{S} = [\boldsymbol{s}(1), \dots, \boldsymbol{s}(N)] = [\boldsymbol{s}_i(\mathrm{j})]_{m \times N}$ is a matrix of unknown sources, in which the rows represent brain sources and artifacts, and $\boldsymbol{W}$ is a $n \times m$ unknown mixing matrix.

ICA methods estimate the mixing matrix $\boldsymbol{W}$ from the EEG by maximizing the statistical independence of the sources from each other (Hyvärinen and Oja, 2000). In this work, the second order blind source identification (SOBI) method is used, which is based on the simultaneous diagonalization of inter-signal correlation matrices over time (Belouchrani et al., 1997). The SOBI method was chosen based on its previously reported merits in terms of ability to separate the EEG neuronal sources from artifacts, insensitivity to the duration of the data segments and ability to preserve more brain activity than other ICA algorithms (Daly et al., 2013; Romero et al., 2008; Tang et al., 2005).

APP identifies ICs related with three major kinds of artifacts, as follows:

1) *Eye movements or blinks*: ICs exhibiting high temporal correlation with the EOG signals are selected. In order to improve the signal-to-noise ratio (SNR) of the EOG signals, the signal of the right EOG electrode is subtracted from the

left one (HEOG) and the signal of lower EOG electrode is subtracted from the upper one (VEOG). It is worth noting that these ICs are usually ranked among the first few components because of their extreme amplitude and the fact that their topographies are essentially flat except for a few frontal electrodes (Chaumon et al., 2015).

2) *Muscle related artifacts*: Muscle activity from the jaw, facial muscles and neck movements often contaminate the EEG recordings. Muscular activity related components are easily spotted due to their high time-point-to-time-point variability. In APP, these ICs are detected by their low autocorrelation as a consequence of their noisy time course (Chaumon et al., 2015). The autocorrelation is defined at lag l (default set to 20 ms) for component $\boldsymbol{x_c}$ by:

$$\boldsymbol{A_c} = \sum_{\boldsymbol{t=l}}^{\boldsymbol{T}} \boldsymbol{x_c(t)} \times \boldsymbol{x_c(t-l)} \qquad \textbf{Eq. 4}$$

3) *Bad channels*: In order to cope with borderline cases that cannot be detected in the first step (described in Detection, removal and interpolation of bad channels), we include an extra step based on ICA in order to detect focal topographies typically associated with bad channels, we use the Generic Discontinuity Spatial Feature (GDSF) implemented in ADJUST (Mognon et al., 2011), which has been reported to perform better in identification of bad channels than existing alternatives (Chaumon et al., 2015), and it is measured by:

$$\boldsymbol{GDSF_c} = \boldsymbol{max}\left(\left|\boldsymbol{w_{c_i}} - \langle \boldsymbol{e^{-\|r_m - r_i\|} w_{c_i}} \rangle_{\boldsymbol{m}}\right|\right)_{\boldsymbol{n}} \qquad \textbf{Eq. 5}$$

where w_{c_i} is the topography weight of channel i in IC c, $\|r_m - r_i\|$ is the distance between channel r_m and r_i, $\langle ... \rangle_m$ is the average of all channels $m \neq i$ and $max(...)_n$ is the maximum over all channels; and n is the number of channels.

After the identification of the artifactual ICs, the EEG signal is reconstructed based on the remaining ICs, effectively removing the artifacts from the reconstructed signal.

2.1.7. Detection, removal and interpolation of bad channels in epochs

After the ICA step, most of the artifacts are already removed. However, some transient artifacts may still affect single channels in specific segments/epochs of the EEG data (e.g., electrodes that became faulty or lost connection in the middle of the recording). APP detects epochs affected by such transient artifacts based on two criteria, which are evaluated for each channel within each epoch:

1) *Dispersion criterion*: The biweight estimate of the temporal standard deviation within each channel is used to quantify the dispersion (similarly to sub-section Detection, removal and interpolation).
2) *High-frequency criterion*: The mean of the difference between two consecutive EEG signal time points of each channel within each epoch is used to quantify high-frequency activity.

Next, the bad channels within each epoch are interpolated using spherical splines. In the case of RS data, after interpolation of the bad channels within each epoch, epochs are concatenated with a 500 ms inverse Hanning window at the intersections.

2.1.8. Outlier detection

For each of the parameters computed in the previous steps, APP uses a general outlier detection and removal scheme based on the distribution of the data. First, the Shapiro-Wilk test is performed test for normality; if the data is Leptokurtic (kurtosis > 3), the Shapiro-Francia test is used instead, since it has been shown to outperform the Shapiro-Wilk test for high peaked data (Royston, 1993). If the data distribution is fairly normal distributed, the scheme uses a modified z-score, which is based on the median and the MAD, instead of the mean and standard deviation, respectively, in order to be robust to outliers (Iglewicz and Hoaglin, 1993). Following the recommendation by Iglewicz and Hoaglin, absolute modified z-scores larger than 3.5 are defined as outliers (as opposed to the more commonly used threshold of 3.0). If the data is not normal, then an adjusted boxplot is used to find outliers (Hubert and Vandervieren, 2008), which includes the medcouple, a robust measure of skewness, in the determination of the whiskers. This provides a more accurate representation of the data and avoids many data points to be considered as outliers as in the common boxplot. Data points outside the whiskers are defined as outliers.

2.2. Alternative EEG pre-processing and artifact removal methods

2.2.1. FASTER

For the ERP data analysis, APP is compared with the state-of-the-art artifact removal method "Fully Automated Statistical Thresholding for EEG artifact Rejection" (FASTER; Nolan et al., 2010), as implemented in the EEGLAB toolbox (Delorme and Makeig, 2004). FASTER performs the entire pre-processing from filtering to participants' grand average, calculating multiple parameters at several steps, and using a threshold of 3 standard deviations from the average to detect outliers.

2.2.2. TAPEEG

For the RS data analysis, APP was compared with a state-of-the-art artifact removal method "Tool for Automated Processing of EEG data" (TAPEEG; Hatz et al., 2015), using our own implementation built onto routines from FASTER and FieldTrip (Oostenveld et al., 2010). Similarly to FASTER, TAPEEG also uses a stepwise procedure to calculate multiple parameters and a threshold of 3 standard deviations from the average as well as hard thresholds to detect outliers.

2.2.3. Prep pipeline

For the RS data analysis, APP was also compared with the Prep pipeline, another state-of-the-art standardized pre-processing method for large-scale EEG analysis (Bigdely-Shamlo et al., 2015). Prep pipeline is an EEG pre-processing pipeline that filters EEG data, re-references the data, and removes bad channels. Since the Prep pipeline does not prune the EEG data for bad epochs of the recording, visual inspection, by experts, was conducted to remove bad segments of the data in this case. The segments removed in this way do not correspond directly to the data epochs removed by APP or TAPEEG, as they are allowed to have variable durations. The plug-in of the algorithm for EEGLAB toolbox was used for comparison.

2.2.4. Supervised artifact identification

Finally, for both ERP and RS data analysis, APP was compared against supervised pre-processing and artifact identification by three researchers with experience in analyzing high-density EEG using CARTOOL (Brunet et al., 2011). Pre-processing of the raw EEG data included DC correction, band-pass filtered between 1 and 40 Hz, 50 Hz noise removal using notch filters. The data was then visually inspected for bad channels and bad epochs; in the case of RS data, data segments of variable duration were removed, which do not exactly match the epochs used in APP or TAPEEG. Bad

channels were interpolated using 3D splines, and the data was re-referenced to the average reference.

2.3. Methods for EEG data acquisition and analysis

2.3.1. EEG recording apparatus

The EEG was recorded using a BioSemi Active 2 system (Biosemi) with 64 Ag-AgCl sintered active electrodes positioned in a cap according to the 10-20 system, referenced to the common mode sense (CMS) electrode. The EOG was recorded with electrodes positioned about 1 cm above and below the right eye and 1 cm lateral to the outer canthi. The recording sampling rate was 2048 Hz.

2.3.2. Participants and procedure

An ERP dataset was collected from 61 healthy participants and 44 schizophrenia patients (Table 1) performing a vernier offset discrimination. The vernier stimulus consisted of 2 vertical lines of 10' (arc minutes) separated by a gap of 1', with a fixed horizontal offset of about 1.2', and it was presented for 30 ms. Participants were requested to report the perceived offset direction by pushing a right or left button and to guess when they were not sure. A total of 160 trials were presented, with an inter-trial pause varying randomly between 1000 and 1500 ms. This type of visual stimuli evokes a strong negative ERP (N1 component) around 200 ms after the stimulus onset in healthy participants, while schizophrenia patients tend to have reduced amplitudes (Plomp et al., 2013).

One RS dataset was collected from 68 healthy participants, Table 1, performing a 5 minutes eyes-closed EEG recording. No specific instructions were given to the participants besides avoiding head movements.

All participants had good visual acuity of at least 0.80, as measured by the Freiburg Visual Acuity Test using both eyes (Bach, 1996). Participants gave informed consent before the experiments. All procedures complied with the Declaration of Helsinki and were approved by the local ethics committee.

Table 1 - Insert approximately here

2.3.3. EEG data analysis

The ERP data were subjected to APP, FASTER or supervised pre-processing and artifact rejection. Epochs ranging from -100 to +400 ms around the stimulus onset were used for data segmentation. We then computed the average ERP and the respective GFP time series, for each participant, and subsequently obtained the grand-average GFP for each group (controls and patients), for each pre-processing and artifact rejection method.

In the RS study, the data were subjected to pre-processing and artifact rejection using APP, TAPEEG, Prep pipeline or the supervised method. Subsequently, the pre-processed data were split into epochs of 4 s in each participant. For each epoch, the relative amplitude of different frequency bands (delta: 1-4 Hz, theta: 4-8 Hz, low alpha: 8-10 Hz, high alpha: 10-13 Hz and beta: 13-30 Hz) was calculated as in (Wan et al., 2016), and the mean relative band amplitude across epochs was determined for each channel. In order to account for the different scalp distributions of the frequency bands, the relative amplitude of each band was averaged across the channels belonging to one of 5 scalp regions (frontal, left tempo-parietal, right tempo-parietal, central and parieto-occipital; for the definition of the regions see Supplementary Figure S1). Furthermore, the peak alpha frequency (PAF) and the individual alpha band (IAB; PAF – 2 Hz to PAF + 2 Hz) were calculated and averaged across the parieto-occipital channels. In summary, we obtain 5 (frequency bands) x 5 (scalp regions) + 2 (IAB and PAF) = 27 parameters per participant, which were submitted to statistical analysis.

3. Results

The results obtained by applying the proposed data pre-processing and artifact removal pipeline APP, as well as its alternative pipelines, are presented here, first for the ERP data and then for the RS data.

3.1. ERP data analysis

3.1.1. GFP analysis

The grand-average GFP time series obtained for the control and patient groups, using each data pre-processing and artifact rejection method, are shown in Figure 3. A GFP N1 peak around 200 ms after stimulus onset can be observed. The average GFP N1 amplitudes at the peak latencies for each *Group* (Controls and Patients), as a function of the pre-processing *Method* (APP, FASTER and Supervised) are shown in Figure 4. A 2-way repeated measures ANOVA with a Greenhouse-Geisser correction on the peak

N1 amplitudes showed significant main effects of *Group* ($F(1, 103) = 38.74$, $P < 0.001$) and *Method* ($F(1.699, 174.96) = 23.367$, $P < 0.001$), as well as a significant interaction effect between them ($F(1.699, 174.96) = 9.898$, $P < 0.001$). This interaction indicates that differences between groups depend on the pre-processing method used. The largest difference between patients and healthy participants' mean peak N1 amplitudes was obtained using APP (1.59 ± 0.83 µV vs 3.21 ± 1.55 µV), followed by the supervised scheme (1.68 ± 0.72 µV vs 3.19 ± 1.51 µV), and FASTER (1.53 ± 0.65 µV vs 2.71 ± 1.24 µV). Moreover, post hoc tests using Bonferroni-Holm correction on the factor *Method* revealed that APP and the supervised scheme lead to similar N1 amplitudes for all participants (2.53 ± 1.53 µV and 2.56 ± 1.46 µV, respectively, $p = 1$). However, the mean N1 amplitudes for all participants after FASTER pre-processing had been reduced to 2.21 ± 1.19 µV, which was statistically significant different to APP ($p < 0.001$) and the supervised scheme ($p < 0.001$).

Figure 3 – Insert approximately here

Figure 4 – Insert approximately here

3.1.2. Channels interpolated

The percentage of channels interpolated using each pre-processing method, for both groups, is presented in Figure 5(a). A 2-way repeated measures ANOVA with a Greenhouse-Geisser correction showed a significant main effect of *Method* ($F(1.67, 172.38) = 153.126$, $P < 0.001$) but no significant effects of *Group* ($F(1, 103) = 1.44$, $P = 0.233$) nor a significant interaction ($F(1.67, 172.38) = 2.79$, $P = 0.074$). Post hoc tests using Bonferroni-Holm correction on the factor *Method* revealed that APP interpolated a similar amount of channels to the supervised scheme (1.05 ± 1.00 % and 1.02 ± 0.66 %, respectively, $p = 0.925$). However, FASTER interpolated around of 3.29 ± 1.45 % channels, which was statistically significant more than APP ($p < 0.001$) and Supervised ($p < 0.001$).

3.1.3. Trials removed

The percentage of trials rejected using each pre-processing method, for both groups, is presented in Figure 5(b). A 2-way repeated measures ANOVA with a Greenhouse-Geisser correction showed significant main effects of *Group* ($F(1, 103) = 17.85$, $P <$

0.001) and *Method* ($F(1.378, 141.93) = 32.94$, $P < 0.001$), but no statistically significant interaction ($F(1.378, 141.93) = 0.548$, $P = 0.515$). Patients in general had more trials rejected than healthy participants (4.78 ± 1.57 % and 3.97 ± 1.64 %, respectively). Post hoc tests using Bonferroni-Holm correction on the factor *Method* revealed that APP rejected significantly fewer trials than FASTER for all participants (3.62 ± 2.00 % and 4.17 ± 0.85 %, respectively, $p = 0.007$). Moreover, the supervised scheme rejected 5.16 ± 1.53 % trials, which was significantly more than APP ($p < 0.001$) and FASTER ($p < 0.001$).

3.1.4. Independent components removed

The percentage of ICs removed using each pre-processing method (APP and FASTER), for both groups, is presented in Figure 5(c). A 2-way repeated measures ANOVA yielded non-significant effects of *Method* ($F(1, 103) = 0.303$, $P = 0.583$) and *Group* ($F(1, 103) = 0.918$, $P = 0.340$) as well as non-significant interaction ($F(1, 103) = 0.078$, $P = 0.781$). On average, APP removed 6.65 ± 3.03 % of ICs, while FASTER removed 6.89 ± 3.44 %.

3.1.5. Channels interpolated per trial

The percentage of channels interpolated per trial by each pre-processing method (APP and FASTER), for both groups, is presented in Figure 5(d). A 2-way repeated measures ANOVA showed a significant effect of *Method* ($F(1, 103) = 167.51$, $P < 0.001$) but no significant effect of *Group* ($F(1, 103) = 1.167$, $P = 0.282$) nor a significant interaction ($F(1, 103) = 3.022$, $P = 0.085$). APP interpolated fewer channels per trial than FASTER (1.89 ± 0.82 % and 2.65 ± 0.52 %, respectively).

Figure 5 – Insert approximately here

3.2. RS data analysis

3.2.1. Frequency bands

A 3-way repeated measures ANOVA with a Greenhouse-Geisser correction was conducted on the main factors: *Method* (Supervised, APP, TAPEEG and Prep pipeline), *Region* (frontal, left tempo-parietal, right tempo-parietal, central and parieto occipital), and *Band* (Delta, Theta, Low Alpha, High Alpha and Beta). The results showed a non-significant main effect of *Method* ($F(2.920, 195.609) = 0.357$, $P = 0.779$), but

significant main effects of *Region* ($F(3.806, 255.025) = 22.488$, $P < 0.001$) and *Band* ($F(2.065, 138.327) = 1448.719$, $P < 0.001$), as well as a significant 3-way interaction ($F(17.110, 1146.356) = 1.790$, $P = 0.024$). The ANOVA yielded also a significant *Band* x *Region* interaction ($F(7.495, 502.152) = 11.239$, $P < 0.001$), and non-significant interactions *Band* x *Method* ($F(5.881, 394.053) = 1.010$, $P = 0.418$) and *Method* x *Region* ($F(9.785, 655.585) = 0.916$, $P = 0.516$). Moreover, two 1-way repeated measures ANOVA's with a Greenhouse-Geisser correction were conducted to compare the effect of *Method* on PAF and IAB of the parieto-occipital region. For PAF, there was no significant effect of *Method* ($F(2.816, 188.657) = 1.859$, $P = 0.142$). Similarly, there was no significant of *Method* on IAB ($F(2.616, 175.282) = 0.943$, $P = 0.411$). In general, all the pre-processing methods yielded similar EEG relative amplitudes in each frequency band for all the 5 scalp regions, as well as similar PAF and IAB for the parieto-occipital region (for summary statistics, see Supplementary Table S1).

The Pearson correlation coefficients between the EEG relative amplitude in each frequency band, as well as the PAF and IAB, obtained using the automatic pipelines (APP, TAPEEG or Prep pipeline) and the ones obtained using the supervised scheme, are shown in Table 2. On average, APP, TAPEEG, and Prep pipeline were found to correlate well with the supervised scheme ($r = 0.922 \pm 0.033$, $r = 0.923 \pm 0.046$, and $r = 0.928 \pm 0.035$, respectively).

Table 2 - Insert approximately here

3.2.2. Channels interpolated

The percentage of channels interpolated in the RS analysis, for each pre-processing method, is presented in Figure 6(a). A one-way repeated measures ANOVA with a Greenhouse-Geisser correction showed a significant main effect of the pre-processing *Method* ($F(1.575, 105.527) = 15.49$, $P < 0.001$). Post hoc tests using Bonferroni-Holm correction revealed that APP interpolated a similar amount of channels as the supervised scheme (1.20 ± 1.87 % and 1.15 ± 2.51 %, respectively, $p = 0.924$), but significantly less than TAPEEG (2.48 ± 1.33 %, $p = 0.018$) and Prep pipeline (3.98 ± 4.87 %, $p < 0.001$). TAPEEG interpolated significantly fewer channels than Prep pipeline ($p = 0.008$) and significantly more than the supervised scheme ($p = 0.018$). Prep pipeline interpolated significantly more channels than the supervised scheme ($p < 0.001$).

3.2.3. Data segments removed

The percentage of RS data segments removed in the analysis, for each pre-processing method, is presented in Figure 6(b). A one-way repeated measures ANOVA with a Greenhouse-Geisser correction yielded a significant main effect of the pre-processing *Method* ($F(2.720, 182.267) = 57.94$, $P < 0.001$). Post hoc tests using Bonferroni-Holm correction revealed that APP removed significantly more data segments than the supervised scheme (8.32 ± 2.24 % and 4.99 ± 2.73 %, $p < 0.001$) and Prep pipeline (5.13 ± 2.61 %, $p < 0.001$) but significantly less than TAPEEG (10.36 ± 3.53, $p < 0.001$). TAPEEG removed significantly more data than the supervised scheme ($p < 0.001$) and Prep pipeline ($p < 0.001$). Prep pipeline removed a similar amount of data segments as the supervised scheme ($p = 0.771$).

3.2.4. Independent components removed

The percentage of ICs removed by each pre-processing method (APP and TAPEEG) is presented in Figure 6(c). A two-tailed paired-samples *t*-test showed that APP and TAPEEG removed similar number of ICs (6.96 ± 3.29 % and 6.46 ± 3.31 %, respectively); ($t(67) = 0.889$, $p = 0.377$).

3.2.5. Channels interpolated per epoch

The percentage of channels interpolated per epoch rejected by each pre-processing method (APP and TAPEEG) is presented in Figure 6(d). A two-tailed paired-samples *t*-test showed that APP interpolated significantly less channels per epoch than TAPEEG (1.46 ± 0.80 % and 2.09 ± 1.18 %, respectively); ($t(67) = 4.125$, $p < 0.001$).

Figure 6 – Insert approximately here

4. Discussion and Conclusion

EEG data are usually contaminated by numerous artifacts and require expert supervision for artifact identification and removal. However, with the increasing size of available datasets due to increasing numbers of EEG channels and study participants, supervised data pre-processing becomes impractical, paving the way for automatic pre-processing methods.

In this study, we propose a novel automatic pipeline (APP) for EEG pre-processing and artifact detection and removal, which makes use of state-of-the-art EEG signal processing techniques in a step-by-step manner to correct the EEG data for external and internal, i.e. participant generated, noise. First, APP filters the EEG data for non-stationarity and removes the power line noise with minimum distortion of the frequency spectrum. Second, it re-references the data to a robust estimate of the mean of all channels. Then, it detects and interpolates bad channels and removes bad epochs. After that, APP uses ICA to remove eye-movement, muscular and bad-channel related artifacts. Finally, it removes epoch artifacts. At each step, several parameters are estimated from the data and outliers are detected and removed based on a robust data driven outlier detection scheme. APP was validated on two datasets containing real ERP and RS data.

ERP data

The ERP dataset consisted of data from healthy controls and schizophrenia patients performing a vernier discrimination task. This type of visual stimuli usually evokes a strong negative N1 component, around 200 ms after the stimulus onset, as measured by the GFP. However, schizophrenia patients tend to have reduced amplitudes compared to healthy controls (Plomp et al., 2013). The same result, i.e. low N1 amplitudes for patients compared to controls, were found using either APP, supervised inspection by experts, or FASTER (a previously proposed method of artifact detection in ERP data). However, the N1 amplitudes obtained using APP were similar to the ones obtained using the supervised scheme but were significantly larger than the ones obtained using FASTER. Interestingly, the difference between the mean of the controls' N1 amplitudes and the mean of patients' N1 amplitudes were larger for APP and the supervised scheme than the ones of FASTER.

Moreover, as expected, patients had more bad trials than controls, for all pre-processing schemes. The supervised scheme rejected more trials than APP and FASTER. This might be the case because the experts did not use ICA to reduce artifacts nor interpolated channels that are bad just in specific trials, thus, considering those trials as bad overall. Considering, the two automatic methods, APP removed less bad trials than FASTER; however, the two methods removed similar amount of ICs. In terms of channels interpolated, APP interpolated similar number of channels as the supervised

scheme; however, FASTER interpolated significantly more than the other two methods. In addition, FASTER interpolated more channels per trial than APP.

Resting-state data

For the validation of the proposed pipeline on RS data pre-processing, data from healthy participants performing a 5 minutes eyes-close EEG recording was used. Similar levels of correlation were found for the power across multiple frequency bands, as well as the peak alpha frequency, between each of the three automatic pre-processing (APP, TAPEEG, and Prep pipeline) and the expert supervision. In addition, APP removed and interpolated a similar number of channels as the supervised scheme and fewer than both TAPEEG and Prep pipeline. Regarding the amount of data segments rejected, EEG experts had to prune the data for artifactual segments, following Prep pipeline pre-processing, since this method does not incorporate such feature. Consequently, Prep pipeline and the supervised scheme rejected similar amount of data segments and significantly less than APP and TAPEEG. The latter two methods use fixed epochs to prune the data for artifactual segments, which might cause the difference between the EEG experts supervision and the two fully automated methods. Considering only the two fully automated methods, APP removed significantly less epochs than TAPEEG but no statistically significant difference was found in terms of ICs removed. Additionally, TAPEEG interpolated more channels per epochs than APP.

Unlike FASTER and TAPEEG, APP does not assume a normal distribution of the parameters calculated at each step, which avoids that much data from skewed distributions be considered as outliers and therefore discarded. This is especially important when dealing with patient data, which by nature tend to be noisier but also more difficult to obtain so that we cannot afford to lose artifact-free information.

Limitations of the proposed pipeline

We only tested APP on 64-channels datasets. However, we expect the pipeline to work at least equally well for recordings with more channels, since each parameter computed at each step is treated statistically and more data therefore provide a better sampling of the parameter distribution. For the same reason, we do not recommend the use of APP for recordings with less than 32 channels without supervision. In such cases, APP can still be used, but only as an auxiliary tool to further support an informed decision on whether or not to classify a specific channel, trial or IC as artifactual.

The most challenging part of the proposed pipeline is the correct classification of ICs. So far, no metric can accurately classify artifactual ICs. Thus, in APP to avoid overcorrection, we relied on metrics that have low false alarm rates, while achieving satisfactory hit rates (for a comparison of the classification accuracy of different IC metrics, see Chaumon et al., 2015). Since Chaumon and colleagues used training sets to define thresholds for each metric, we conducted a small analysis to determine how these metrics would perform using APP's outlier detection scheme. Results were inspected by one EEG expert and the metrics reported in Chaumon et al. (2015) to have better results were found to yield satisfactory performance. Furthermore, in the absence of EOG signals the correction of eye-movement related artifacts cannot be performed, since APP uses EOG signal as a reference for the IC selection. Further work on APP could include metrics for improved IC classification, as well as metrics for the detection of eye-movement related ICs that do not resort on reference signals, such as the ADJUST eye movement detector. Finally, APP was only tested in laboratory conditions, in which artifacts are somehow mild and SNR is relatively high when compared to harsher environments, such as natural settings outside laboratories or inside magnetic resonance scanners. Therefore, we do not recommend the use of APP with EEG data acquired under such circumstances.

Summary

In sum, the aim of this paper was to propose and validate APP, a novel method for EEG pre-processing and artifact removal. Our results show that APP performs at the same level as the pre-processing done by EEG experts, while outperforming existing alternatives in many aspects, namely the amount of data lost, and achieving higher ERP amplitudes. Currently, APP is a series of Matlab® scripts that can be obtained by contacting the corresponding author. However, our intention is to integrate APP as a plugin for EEGLAB, and to make the respective source code freely available online. The default parameters in this plugin will be the ones used in the present article; however, the users will have the option to change the parameters at their convenience. We hope that this method will contribute to the EEG research field by aiding researchers deal with the increased amount of data and improving the reproducibility of results.

Conflict of Interest Statement

None of the authors have declared any conflict of interest.

Acknowledgment

This work was partially funded by the Fundação para a Ciência e a Tecnologia under grants FCT UID/EEA/50009/2013 and FCT PD/BD/105785/2014, and the National Centre of Competence in Research (NCCR) Synapsy (The Synaptic Basis of Mental Diseases) under grant 51NF40-158776.

Appendix

The biweight estimate (Hoaglin et al., 1983, 1985) is a weighted average, in which the weights decrease nonlinearly as going away from the center of the distribution. The weighting function returns zero after a certain distance from the center of the distribution and this distance is controlled by a censoring parameter c. First, the median M and the median absolute deviation MAD are determinate. Second, a weight $\boldsymbol{u}(i)$ is assigned to each of the N observations $\boldsymbol{x}(i)$ as follows:

$$\boldsymbol{u}(\boldsymbol{i}) = \frac{\boldsymbol{x}(\boldsymbol{i}) - \boldsymbol{M}}{\boldsymbol{c} \times \boldsymbol{MAD}}$$

To censor the extreme values, $\forall\, |\boldsymbol{u}(i)| \geq 1$, $\boldsymbol{u}(i) = 1$. In this study, c is set to 7.5, which censors values more than 5 standard deviations away from the center of the distribution.

Then, the biweight estimate of the mean is given by:

$$\overline{\boldsymbol{x}}_{\boldsymbol{biweight}} = \boldsymbol{M} + \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x}(\boldsymbol{i}) - \boldsymbol{M})\,(\mathbf{1} - \boldsymbol{u}(\boldsymbol{i})^{\mathbf{2}})^{\mathbf{2}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\mathbf{1} - \boldsymbol{u}(\boldsymbol{i})^{\mathbf{2}})^{\mathbf{2}}}$$

The biweight estimate of the standard deviation is calculated in a similar fashion:

$$\boldsymbol{s}_{\boldsymbol{biweight}} = \frac{\sqrt{\boldsymbol{N} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x}(\boldsymbol{i}) - \boldsymbol{M})^{\mathbf{2}} (\mathbf{1} - \boldsymbol{u}(\boldsymbol{i})^{\mathbf{2}})^{\mathbf{4}}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\mathbf{1} - \boldsymbol{u}(\boldsymbol{i})^{\mathbf{2}})\,(\mathbf{1} - \mathbf{5}\boldsymbol{u}(\boldsymbol{i})^{\mathbf{2}})}$$

References

Abreu, R., Leite, M., Jorge, J., Grouiller, F., van der Zwaag, W., Leal, A., Figueiredo, P., 2016a. Ballistocardiogram artifact correction taking into account physiological signal preservation in simultaneous EEG-fMRI. NeuroImage 135, 45–63. https://doi.org/10.1016/j.neuroimage.2016.03.034

Abreu, R., Leite, M., Leal, A., Figueiredo, P., 2016b. Objective selection of epilepsy-related independent components from EEG data. J. Neurosci. Methods 258, 67–78. https://doi.org/10.1016/j.jneumeth.2015.10.003

Bach, M., 1996. The Freiburg Visual Acuity test--automatic measurement of visual acuity. Optom. Vis. Sci. Off. Publ. Am. Acad. Optom. 73, 49–53.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., Moulines, E., 1997. A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. 45, 434–444. https://doi.org/10.1109/78.554307

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., Robbins, K.A., 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front. Neuroinformatics 9, 16. https://doi.org/10.3389/fninf.2015.00016

Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal Analysis of Multichannel EEG: CARTOOL. Intell Neurosci. 2011, 2:1–2:15. https://doi.org/10.1155/2011/813870

Chaumon, M., Bishop, D.V.M., Busch, N.A., 2015. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J. Neurosci. Methods. https://doi.org/10.1016/j.jneumeth.2015.02.025

Daly, I., Nicolaou, N., Nasuto, S.J., Warwick, K., 2013. Automated Artifact Removal From the Electroencephalogram A Comparative Study. Clin. EEG Neurosci. 44, 291–306. https://doi.org/10.1177/1550059413476485

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Fletcher, E.M., Kussmaul, C.L., Mangun, G.R., 1996. Estimation of interpolation errors in scalp topographic mapping. Electroencephalogr. Clin. Neurophysiol. 98, 422–434. https://doi.org/10.1016/0013-4694(96)95135-4

Hatz, F., Hardmeier, M., Bousleiman, H., Rüegg, S., Schindler, C., Fuhr, P., 2015. Reliability of fully automated versus visually controlled pre- and post-processing of resting-state EEG. Clin. Neurophysiol. 126, 268–274. https://doi.org/10.1016/j.clinph.2014.05.014

Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), 1985. Exploring Data Tables, Trends, and Shapes, 1 edition. ed. Wiley, New York.

Hoaglin, D.C., Mosteller, F., Tukey, J.W. (Eds.), 1983. Understanding robust and exploratory data analysis. Wiley, New York.

Hubert, M., Vandervieren, E., 2008. An adjusted boxplot for skewed distributions. Comput. Stat. Data Anal. 52, 5186–5201. https://doi.org/10.1016/j.csda.2007.11.008

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Iglewicz, B., Hoaglin, D.C., 1993. How to Detect and Handle Outliers. ASQC Quality Press.

Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. 48, 609–621. https://doi.org/10.1016/0013-4694(80)90419-8

Luck, S.J., 2014. An introduction to the event-related potential technique, Second edition. ed. The MIT Press, Cambridge, Massachusetts.

Lynn, P.A., 1989. Introduction to the Analysis and Processing of Signals, 3rd ed. Hemisphere Publishing Corporation.

Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M., 2011. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48, 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Mullen, T., 2012. CleanLine. NITRC.

Niedermeyer, E., Lopes da Silva, F.H., 2005. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins.

Nolan, H., Whelan, R., Reilly, R.B., 2010. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. J. Neurosci. Methods 192, 152–162. https://doi.org/10.1016/j.jneumeth.2010.07.015

735 Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M.,
736 Silberstein, R.B., Cadusch, P.J., 1997. EEG coherency: I: statistics, reference
737 electrode, volume conduction, Laplacians, cortical imaging, and interpretation
738 at multiple scales. Electroencephalogr. Clin. Neurophysiol. 103, 499–515.
739 https://doi.org/10.1016/S0013-4694(97)00066-7
740 Onton, J., Westerfield, M., Townsend, J., Makeig, S., 2006. Imaging human EEG
741 dynamics using independent component analysis. Neurosci. Biobehav. Rev. 30,
742 808–822. https://doi.org/10.1016/j.neubiorev.2006.06.007
743 Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2010. FieldTrip: Open Source
744 Software for Advanced Analysis of MEG, EEG, and Invasive
745 Electrophysiological Data. Comput. Intell. Neurosci. 2011, e156869.
746 https://doi.org/10.1155/2011/156869
747 Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp
748 potential and current density mapping. Electroencephalogr. Clin. Neurophysiol.
749 72, 184–187. https://doi.org/10.1016/0013-4694(89)90180-6
750 Picton, T. w., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson, R., Miller, G.
751 a., Ritter, W., Ruchkin, D. s., Rugg, M. d., Taylor, M. j., 2000. Guidelines for
752 using human event-related potentials to study cognition: Recording standards
753 and publication criteria. Psychophysiology 37, 127–152. https://doi.org/null
754 Plomp, G., Roinishvili, M., Chkonia, E., Kapanadze, G., Kereselidze, M., Brand, A.,
755 Herzog, M.H., 2013. Electrophysiological Evidence for Ventral Stream Deficits
756 in Schizophrenia Patients. Schizophr. Bull. 39, 547–554.
757 https://doi.org/10.1093/schbul/sbr175
758 Romero, S., Mañanas, M.A., Barbanoj, M.J., 2008. A comparative study of automatic
759 techniques for ocular artifact reduction in spontaneous EEG signals based on
760 clinical target variables: A simulation case. Comput. Biol. Med. 38, 348–360.
761 https://doi.org/10.1016/j.compbiomed.2007.12.001
762 Royston, P., 1993. A Toolkit for Testing for Non-Normality in Complete and Censored
763 Samples. J. R. Stat. Soc. Ser. Stat. 42, 37–43. https://doi.org/10.2307/2348109
764 Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic
765 inverse problem. Phys. Med. Biol. 32, 11. https://doi.org/10.1088/0031-
766 9155/32/1/004

Tang, A.C., Sutherland, M.T., McKinney, C.J., 2005. Validation of SOBI components from high-density EEG. NeuroImage 25, 539–553. https://doi.org/10.1016/j.neuroimage.2004.11.027

Tatum, W.O., Dworetzky, B.A., Schomer, D.L., 2011. Artifact and Recording Concepts in EEG: J. Clin. Neurophysiol. 28, 252–263. https://doi.org/10.1097/WNP.0b013e31821c3c93

Urigüen, J.A., Garcia-Zapirain, B., 2015. EEG artifact removal—state-of-the-art and guidelines. J. Neural Eng. 12, 031001. https://doi.org/10.1088/1741-2560/12/3/031001

Wan, F., Cruz, J.N. da, Nan, W., Wong, C.M., Vai, M.I., Rosa, A., 2016. Alpha neurofeedback training improves SSVEP-based BCI performance. J. Neural Eng. 13, 036019. https://doi.org/10.1088/1741-2560/13/3/036019

Tables Captions

Table 1 - Average Statistics (±SD) of Participants in the Event-Related Potential (ERP) and Resting-State (RS) analysis.

Table 2 - Pearson correlation coefficients of the EEG relative band amplitude in each frequency band and the individual alpha band (IAB) and the peak alpha frequency (PAF), between APP, TAPEEG or Prep Pipeline and the supervised scheme.

Figure Captions

Figure 1 - Example of the EEG spectral power of a single channel (FCz) before (top) and after power line noise removal (bottom) using CleanLine (Mullen, 2012). The power line line power (50 Hz) and its harmonics (150 and 250 Hz) are greatly reduced.

Figure 2 - Comparison of common mean reference (red) and the proposed biweight estimate of the mean (blue). In the case, with no bad channels (top), the two behave similarly, reflected by a small difference (black) and high correlation. However, when a bad channel is included (bottom), the common mean is greatly affected by the bad channel, while the biweight estimate of the mean remains the same, which is reflected in a large difference between them (black) as well as a lower correlation.

Figure 3 - The global field power (GFP) time series to the target stimulus for healthy controls (left) and patients (right) after pre-processing with APP (red), FASTER (blue) and the Supervised scheme (black). Dashed lines represent standard error of the mean.

Figure 4 - Average GFP N1 amplitudes at the peak latencies for each *Group* (Controls and Patients), as a function of the pre-processing *Method* (APP, FASTER and Supervised). The interaction effect between *Group* and *Method* indicates that the GFP difference between patients and healthy controls is not as large when using FASTER compared to using APP or Supervised pre-processing. Error bars indicate the standard error of the mean.

Figure 5 - Channels interpolated (a), trials removed (b), independent components removed (c) and channels interpolated per trial (d), for healthy controls and schizophrenia patients, using APP, FASTER and supervised artifact rejection (except for channels interpolated per trial). Error bars indicate the standard error of the mean.

Figure 6 - Channels interpolated (a), data segments removed (b), independent components removed (c) and channels interpolated per epoch (d) using APP, TAPEEG, Prep pipeline and supervised artifact reject (except channels interpolated per epoch for the latter two methods). Error bars indicate the standard error of the mean.

Tables

Table 1

	ERP Participants		RS Participants
	Schizophrenia Patients	Healthy Controls	Healthy Participants
Gender (F/M)	11/33	29/32	29/39
Age	33.4 ± 8.3	35.1 ± 9.6	35.0 ± 8.2
Education	13.4 ± 2.7	15.0 ± 2.8	14.1 ± 2.7
Illness duration	9.2 ± 7.1		
SANS*	10.4 ± 5.1		
SAPS**	9.0 ± 2.9		
Handedness (R/L)	43/1	56/5	64/4
Visual acuity	1.40 ± 0.34	1.51 ± 0.41	1.62 ± 0.39

* SANS: Scale for the Assessment of Negative Symptoms

** SAPS: Scale for the Assessment of Positive Symptoms

818 Table 2

	APP vs Supervised	TAPEEG vs Supervised	Prep pipeline vs Supervised
Delta	0.918	0.892	0.905
Theta	0.927	0.940	0.938
Low Alpha	0.960	0.969	0.976
High Alpha	0.949	0.967	0.947
Beta	0.952	0.971	0.966
IAB	0.870	0.861	0.882
PAF	0.878	0.864	0.885

819

Figures

Figure 1

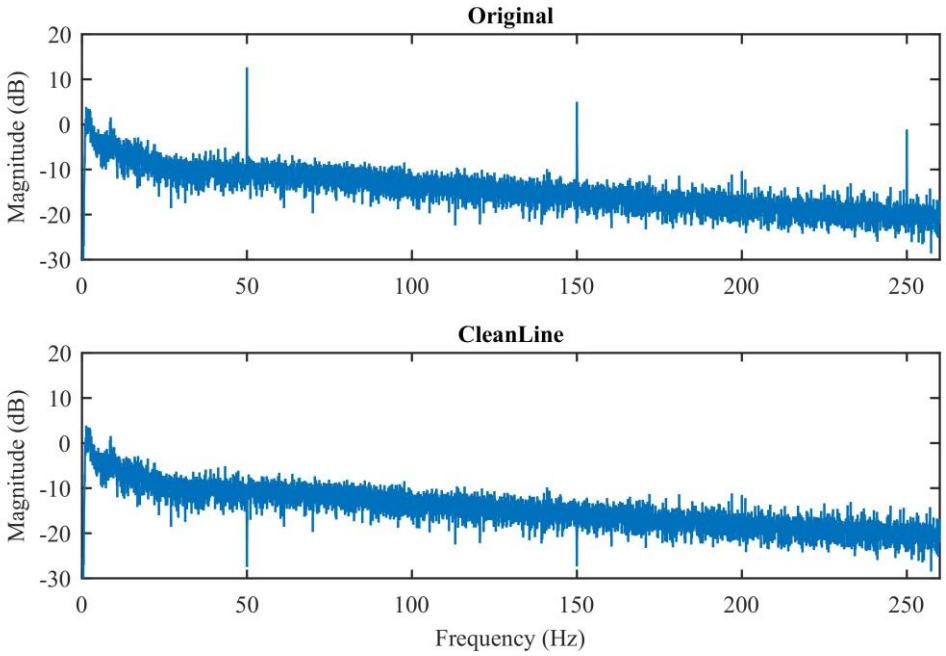


Figure 2

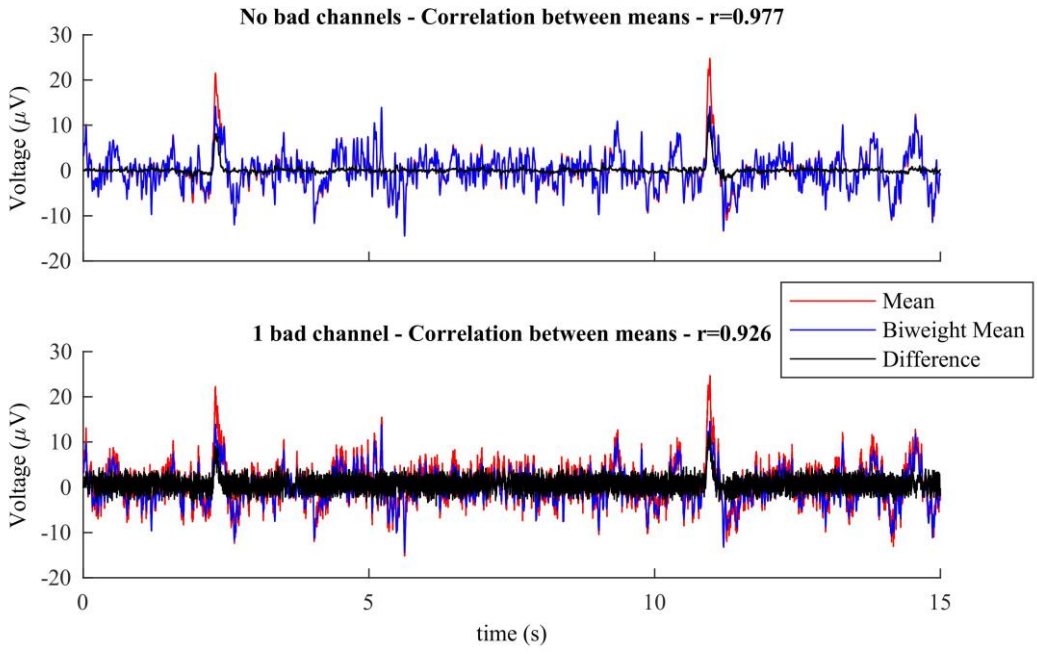


Figure 3

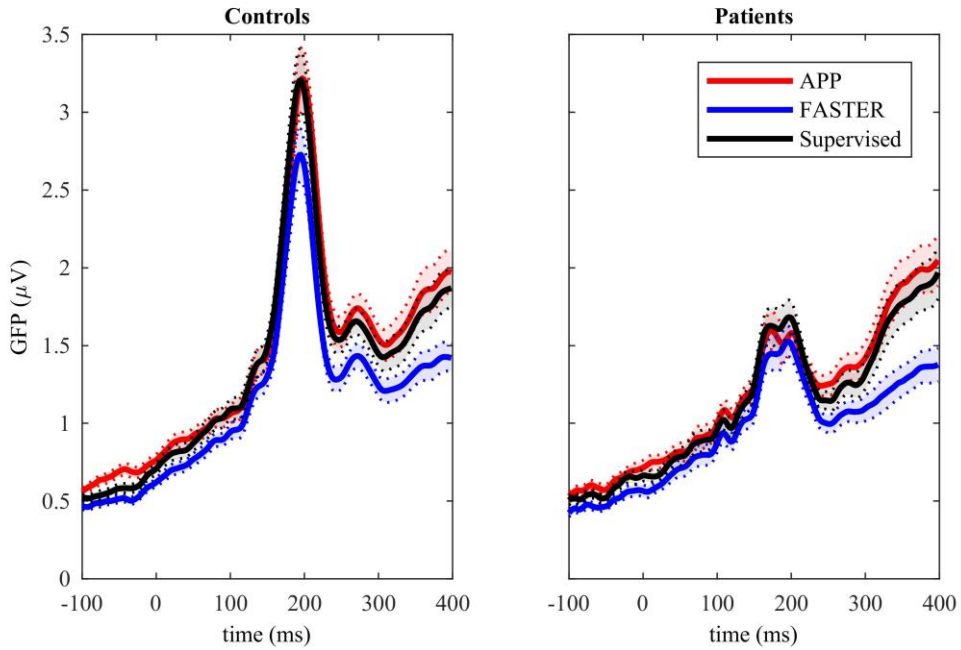
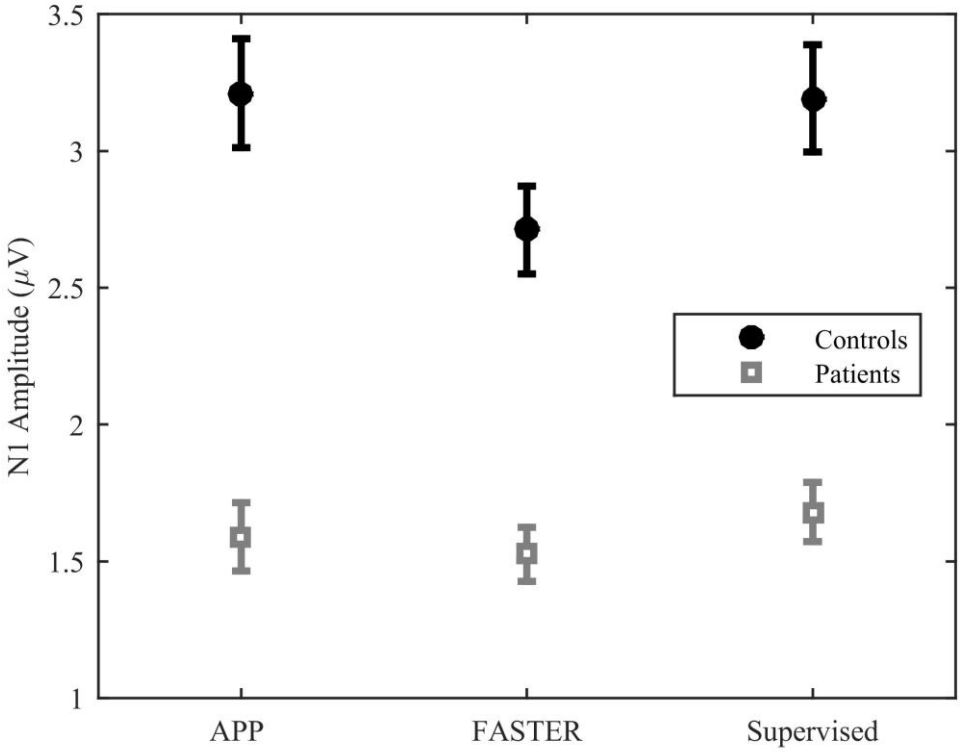
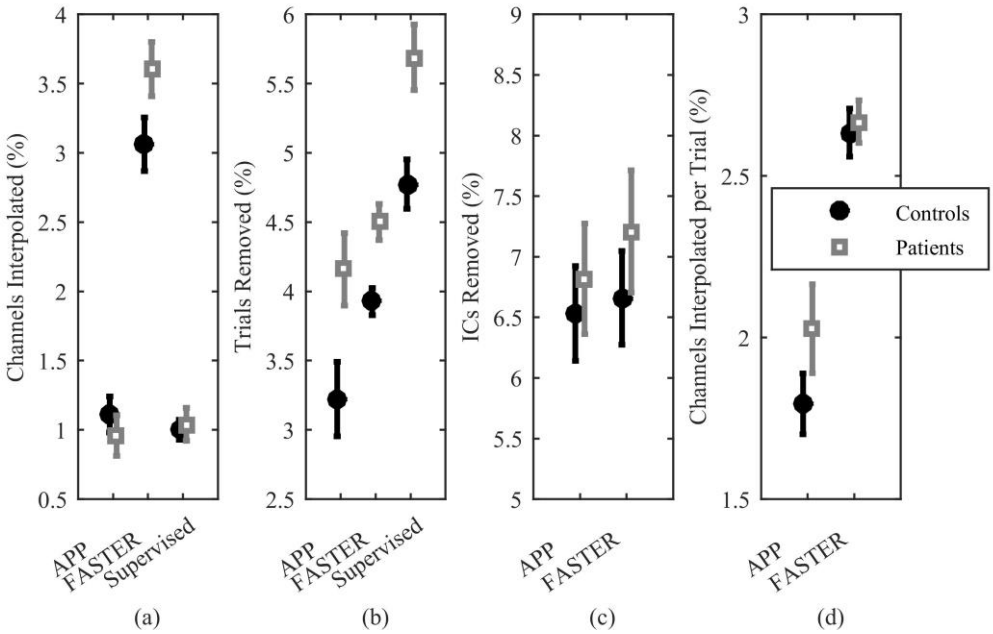


Figure 4

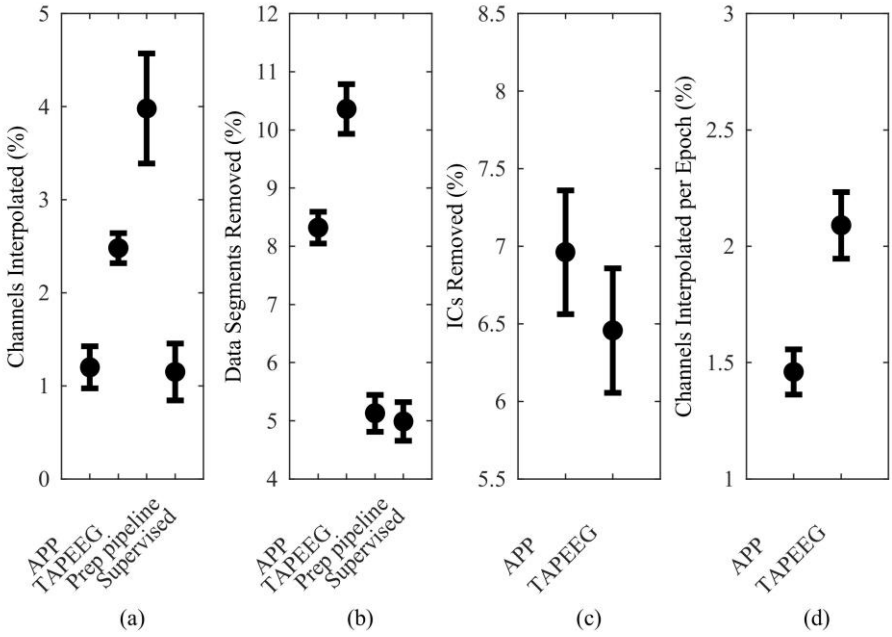


831 Figure 5



832

833 Figure 6



834