Computers in Biology and Medicine 114 (2019) 103442

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

A new algorithm for blink correction adaptive to inter- and intra-subject variability

E. Guttmann-Flury *, X. Sheng, D. Zhang, X. Zhu, Member, IEEE

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, 200240, PR China

ARTICLE INFO

Keywords:
Electroencephalography
EEG
Brain computer interface
BCI
Spontaneous blinking
Artifact removal
Sample size calculation
Standardized low-resolution electromagnetic tomography
sLORETA
Riemannian geometry
Independent component analysis
ICA

ABSTRACT

Electroencephalographic (EEG) signals are constantly superimposed with biological artifacts. In particular, spontaneous blinks represent a recurrent event that cannot be easily avoided. The main goal of this paper is to present a new algorithm for blink correction (ABC) that is adaptive to inter- and intra-subject variability. The whole process of designing a Brain-Computer Interface (BCI)-based EEG experiment is highlighted. From sample size determination to classification, a mixture of the standardized low-resolution electromagnetic tomography (sLORETA) for source localization and time restriction, followed by Riemannian geometry classifiers is featured. Comparison between ABC and the commonly-used Independent Component Analysis (ICA) for blinks removal shows a net amelioration with ABC. With the same pipeline using uncorrected data as a reference, ABC improves classification by 5.38% in average, whereas ICA deteriorates by −2.67%. Furthermore, while ABC accurately reconstructs blink-free data from simulated data, ICA yields a potential difference up to 200 % from the original blink-free signal and an increased variance of 30.42 %. Finally, ABC's major advantages are ease of visualization and understanding, low computation load favoring simple real-time implementation, and lack of spatial filtering, which allows for more flexibility during the classification step.

1. Introduction

In neuroscience, electroencephalography (EEG) is one of the prime techniques to observe the dynamics of brain function. Allied with cognitive psychology, neurolinguistics, psychophysiology and psychopathology, EEG may help address difficult questions, such as "What brain regions are associated with the state of depression and stress?" [1] and "Can waveforms associated with performance monitoring be used to diagnose anxiety disorder or major depressive disorder?" [2]. The first paper published in 1929 by Berger [3] enumerated the EEG of 38 trepanned patients and 14 healthy subjects. Interestingly, the effects of blinking in the EEG were already clearly visible in one of the subjects who had the left frontal skull partly removed. Berger examined the influence of muscle movements on the EEG and warned that these movements could be a source of error. With instrumental and computational improvements, EEG has since been applied for a variety of uses, including clinical rehabilitation using neurofeedback therapy, enhancing sport's performance [4], neuromarketing [5] and even space exploration [6].

In addition to answering neurophysiological questions, this brain imaging technique constitutes the predominant technology in Brain-Computer Interfaces (BCI). BCIs provide new output channels that reflect brain activity rather than relying on peripheral nerves and muscles [7]. With an extensive range of possible practical applications, non-invasive EEG-based BCIs may directly control devices or offer an indirect communication channel. Therefore, significant efforts have been devoted to developing wheelchair or robotic guidance, as well as "speller" implementation. Recently, other applications are being considered, such as hands-free control channels for gaming, virtual reality and creative expression. However, after 40 years of research, the market for BCIs is still mostly narrowed to people with severe neurological disabilities as an assistive technology or used as a neuro-rehabilitative tool in clinical environments [8]. Nevertheless, the use of BCIs in the game market for healthy users is expected to expand, as publicity about the technology is growing [9].

As the first step in EEG-based BCI data analysis, pre-processing algorithms focus on minimizing the influence of artifacts in the EEG signal. By definition, "artifacts" have been described as undesired signals that may affect measurements and signal of interest [10]. Not of cerebral origin, such artifacts have traditionally been divided into

* Corresponding author.
E-mail addresses: eva.flury@gmail.com (E. Guttmann-Flury), mexyzhu@sjtu.edu.cn (X. Zhu).

https://doi.org/10.1016/j.compbiomed.2019.103442
Received 4 July 2019; Received in revised form 4 September 2019; Accepted 7 September 2019
Available online 10 September 2019

physiological and non-physiological artifacts. Most algorithms concentrate on the former, since the latter can be easily reduced [11]. On the other hand, there is a myriad of physiological artifacts that can rarely be avoided. Unquestionably, cardiac activity is always present, though usually of low amplitude on the scalp. The electromyogram (EMG) activity caused by contracting muscles can be more delicate to deal with. Depending on which muscle groups are involved, characterization and correction can be strenuous, as EMG might overlap on classic EEG bands [12]. Specifically, oculomotor activity majorly contaminates EEG data by causing a change in the electrical fields that surround the eyes [13]. Typical eyelid movements, notably during blinks, have amplitudes generally much larger than of the background EEG activity, which propagates differently across the scalp [14]. Furthermore, there is a bidirectional interference between Electrooculography (EOG) and EEG contamination [15], which cannot be solved by asking subjects to refrain from blinking, since this instruction may seriously distort brain activity [16].

The goal of this paper is to present the algorithm for blink correction (ABC) that has the advantage of being able to adapt to inter- and intra-subject variability. The main underlying hypothesis is that accurately correcting blinks improves classification accuracy. Using the maximum amplitude on the frontopolar channel (FP1) as a parameter, ABC aims to correct blinks while staying in the temporal domain. Thus, no assumption must be made concerning the independence of the signals and it might be used with a small number of electrodes, provided that at least one is a frontopolar channel. The objective is to develop a "plug-and-play" blink artifact correction algorithm with low computation load and using only EEG channels. As such, it has the potential to be implemented for any EEG user in real-time. So far, ABC has been tested offline and compared to the widely-used Independent Component Analysis (ICA) as a pre-processing tool for blinks removal. The classification is then carried out using a mixture of the standardized low-resolution electromagnetic tomography (sLORETA) for source localization and time restriction [17] followed by Riemannian geometry classifiers. Indeed, the latter are considered the current state of the art for multiple BCI problems [18] and the former allow for restriction of the analysis to where and when the signals are actually task-related.

2. Experiment design

2.1. A priori sample size calculation

Since the classification is carried out by calculating the distance between the trial-to-classify covariance matrix and the condition-related (Left-MI or Right-MI) Riemannian geometry mean, the variable to investigate is the foresaid distance, δ. Using the methodology previously described in Ref. [19], the distribution of δ is first investigated to assess whether it follows a Gaussian distribution. The experiment is first run on a few subjects (four to five) to explore the distribution and estimate the minimum total number of subjects necessary. Arbitrarily, the number of trials was fixed to 40, consisting of 20 left and 20 right motor imagination hand grasping.

The Shapiro-Wilk test examined the null hypothesis that a sample of random variables arises from a normally distributed population. Hereafter, the Type I error rate (or significance level) is fixed at $\alpha = 0.05$, meaning that a p-value of less than 0.05 rejects the null hypothesis that the data are from a normally distributed population. In both cases ($p-value = 0.03542$ for 4 subjects and $p-value = 0.000337$ for 5 subjects), the null hypothesis is rejected, or in other words the δ distribution cannot be considered to be Gaussian. Accordingly, six distributions were tested: Logistic, Lognormal, Gamma, Weibull, Cauchy and Gumbel. By inference of common tests (Kolmogorov-Smirnov test, Akaike information criterion) and visual inspection of the Q-Q and P–P plots, the least less likely distribution is determined. All test and plots were computed using the fitdistrplus library available in the Comprehensive R Archive Network [20]. As Fig. 1 shows, the δ distribution is least less likely to follow a Lognormal distribution. Indeed, if the distance to the Riemannian geometry mean is representative, most distances are expected to be small. Because a distance cannot be smaller than zero but can very well be greater than forecasted (i.e. outliers), it is consistent with a positive skewness (or right-tail) of the δ distribution.

Subsequently, a Monte Carlo (MC) simulation is generated using the parameters of the Lognormal distribution calculated beforehand. The Type II error rate is fixed at $\beta = 0.2$, leading to a power of 0.80. The estimation of Cohen's effect size [21], or degree of departure from the null hypothesis, yield $d \approx 0.81$ for four subjects (resp. $d \approx 0.69$ for five subjects). In the first case, the minimum sample size should be 750 trials, or 19 sessions of 40-trials, whereas in the second the minimum sample size increased to 3750 trials, or 94 sessions. Compared to the same computations carried out with the hypothesis of a normal distribution, the sample requirement decreased by approximately 60% (1750 trials for four subjects, resp. 10250 trials for five subjects with the same other parameters).

According to Cohen [21], $d = 0.8$ represents a large effect size, implying that the populations are so separated that almost half of their areas are not overlapping. This only partially holds in the case of distance to the Riemannian mean, as it has been observed that it also shows a great inter- and intra-subject variability. Consequently, a large effect size will probably yield a too small number of subjects. Nonetheless, finding 94 subjects, as would suggest the calculation with a smaller effect size, is a daunting task. In this experiment, a compromise has been made between these two requirements.

2.2. Participants

Thirty-one healthy subjects voluntarily took party in this study (11 women and 20 men, mean age 28.6 ± 6.8 (SD), range 20 − 57). All volunteers signed an informed consent form, approved by the local ethics committee of Shanghai Jiaotong University and stating that EEG signals would be recorded. However, to avoid any unforeseen interference, the real reason for studying the effects of blinking on EEG was not explained to participants until the end of the experiment. All participants were asked to fill out a questionnaire based on [22] to which were added physical characteristics questions (such as height, skull diameter, eye correction) and the Edimburg Handedness Questionnaire [23]. All subjects were kept in the study regardless of handedness or their familiarity to any of the BCI tasks. Moreover, after each task, subjects were asked about their degree of alertness, potential mistakes or exterior distractions. Finally, pictures were taken for each subject in order to capture other physical characteristics that were not already recorded.

2.3. Environment

During this experiment, subjects were seated in an electrically-shielded noise-proof chamber, designed specifically for EEG recordings. A 65 channel Quik-cap acquired signals through a SynAmps2 system connected to an amplifier (Compumedics, Neuroscan). Sixty-two (62) EEG electrodes were placed according to the extended 10/20 system. The reference was located on the right mastoid with the ground electrode on the forehead. The EEG signals were recorded at a 1000 Hz sampling frequency.

At the beginning of the session, subjects were informed to follow the visual cues generated by the E-Prime software on the presentation screen, place on a desk in front of them. During the experiment, subjects were left alone in the chamber to avoid any noise or distractions. The task consisted in imagining a hand grasping movement (Left or Right), also called Motor Imagination (MI). At the beginning of each trial, a white fixation cross appeared in the center of the screen for 2 s. A red rectangle cue then showed up randomly on the left or the right side of the cross for 4 s. Meanwhile, subjects performed kinesthetic motor imagination of grasping for 3 times. Finally, the fixation cross and the red rectangle cue disappeared during a random rest of 1–1.5 s allowing

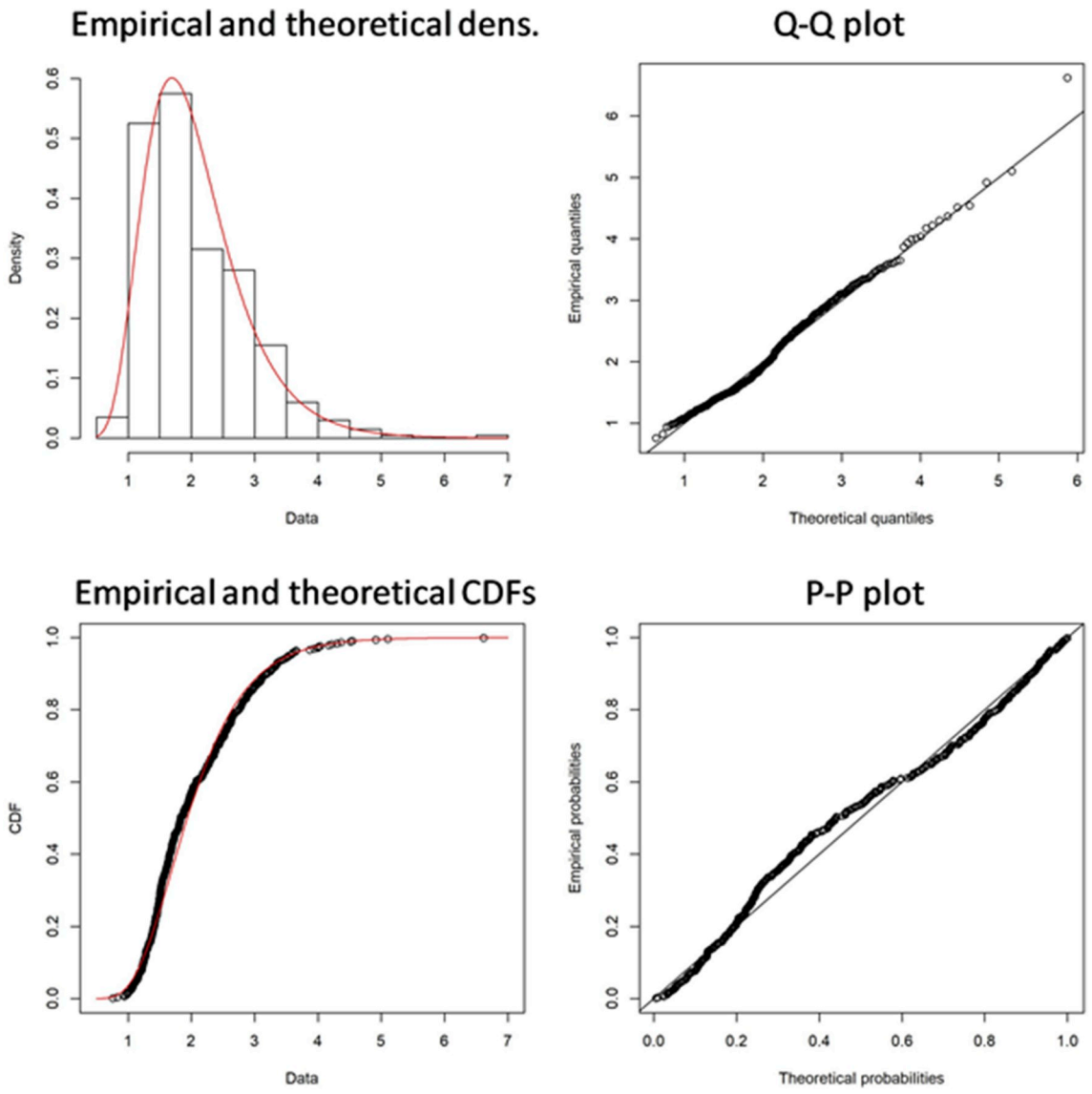

Fig. 1. Visual inspection of the δ distribution fitted with the Lognormal distribution.

relaxation while avoiding subject's adaptation. The experimental paradigm is illustrated in Fig. 2.

This MI task lasted around 6 min per session with a total of 40 trials (20 Left + 20 Right randomly distributed). At the beginning of each session, subjects had to fill in a questionnaire detailing their conditions at the time of recording. For example, the degree of alertness has been divided into four levels: "Rested", "Slightly tired", "Moderate fatigue" and "Extreme fatigue". At the end of each task, similar questions were

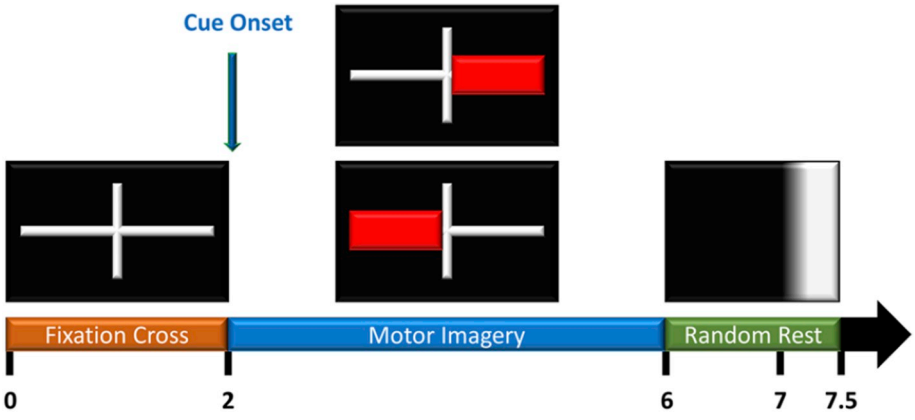

Fig. 2. Experimental paradigm.

asked.

3. Algorithm for blink correction (ABC)

3.1. Problem statement

Human eye blinks are commonly categorized as spontaneous, reflexive or voluntary [24]. Herein, we focus on spontaneous blinks, which are typically shorter in duration as well as smaller in amplitude compared to the others. On the frontopolar channels, blinks are easily recognizable by anyone who has measured EEG signals. As in many signal processing or computer vision problems, the difficulty consists in developing an algorithm that doesn't need human input to detect the shape of interest. Once the blinks have been identified, a preliminary analysis will highlight the huge intra- and inter-subject variability of their characteristic parameter. Consequently, a correction, and by extension a physiologically-based correction, will be deemed appropriate.

3.2. A typical blink

Typically, a human eye blink has an amplitude which can be 10 to 100 times larger than the electrical signals originating from the cerebral cortex and may last up from 200 *ms* to 400 *ms* [25]. This artifact is primarily picked up by frontal electrodes but extends further [14]. Though blink rate is highly task-dependent, its mean was found to be close to 17 blinks per minute at rest [26].

When the eyes are open, the orbicularis oculi muscle (OOM) and the levator palpebral superioris (LPS) show a balanced level of tonic activity. The former is innervated by the cranial nerve VII, also called facial nerve, and can be divided into the pretarsal, preseptal and orbital portions. The first two, corresponding to the upper and lower eyelid respectively, are used during the downward movement of a spontaneous blink. At the end of the closing phase, while LPS is still inhibited, OOM also stops its contraction causing both muscles to be inactive for a short period of time. Then LPS, innervated by the cranial nerve III or by the oculomotor nerve, solely contracts and raises the upper eyelid [27].

Blinks can be divided into two parts in the EEG: the ascending part in Fig. 3 corresponds to the downward motion of the upper eyelid, while the decreasing part represents the upward motion. The upper eyelid accelerates rapidly in its descent, reaching its maximum speed when it crosses the visual axis. During the BCI tasks, it has been observed that most subjects do not blink completely, as the upper eyelid usually comes to rest before attaining the lower eyelid [28]. On the contrary, if subjects are aware that their blinks will be measured, they tend to execute voluntary blinks. These are mostly complete. This is the reason why the real goal of this experiment was kept from the participants until the end of the recordings. The greater muscle fibers number involved with OOM during the downward motion might explain the shorter Anterior Width time compared to the Posterior Width. In ABC, these values are purposely set to correspond to a total blink length time of (see Fig. 3)500*ms*.

3.3. Blink detection

The blink detection is firstly carried out on a single channel. Since a blink occurs on average every 3.5 s, the first step consists in downsampling to 250 *Hz* and lowpass filtering the data with a 4th-order Butterworth filter on 2 s-wide windows over a reference frontopolar channel (FP1). The reference channel is chosen as the one on which the blink amplitude is maximum, which is subject-dependent among FP1, FP2 and FPZ. If no frontopolar channel is available, it might still be possible to detect blinks with a further electrode, however some smaller blinks might remain undetected. The amplitude criterion then selects all data higher than the moving average to which a heuristic threshold is added, as highlighted in Fig. 4. If no data is greater than the threshold, ABC swipes right to the next window. Most of the blink-free windows are already discarded after this first criterion.

The third step consists merely in calculating the maximum amplitude on the selected window. Up to this point, noisy data can be mistakenly considered as a potential blink. For example, electrode "pops" or other kind or artifacts might exceed the foresaid threshold. To remove them, the differences between the maximum amplitude and the average mean before the Anterior Width, respectively after the Posterior Width are computed. If both these values are greater than the threshold, the Pre- and Post-amplitude criterion is validated, assessing the blink shape likeliness. Accordingly, the data around the maximum potential is extracted and zero-centered. Since blinks are known to propagate along the scalp, the area at the reference channel FP1 and the area at a further electrode (CZ) are compared. A ratio comparison allows the elimination of brain-related or eyeball movement event and consists of the last criterion. Similarly, the choice of this further electrode is arbitrary. Any electrode in the vicinity of CZ (e.g. C1, C2) might be chosen to carry out the propagation criterion. When all three criteria have been verified, the blink is extracted for further analysis.

3.4. Blink classification

Heuristically, all blinks' maximum amplitude on FP1 are included in a $[40 - 450]\mu V$ interval. These choices have been made considering that a maximum amplitude lower than 40 μV could be caused by a phenomenon other than a blink, while a one higher than 360 μV is probably caused by a mixture of events, for example two successive blinks or a

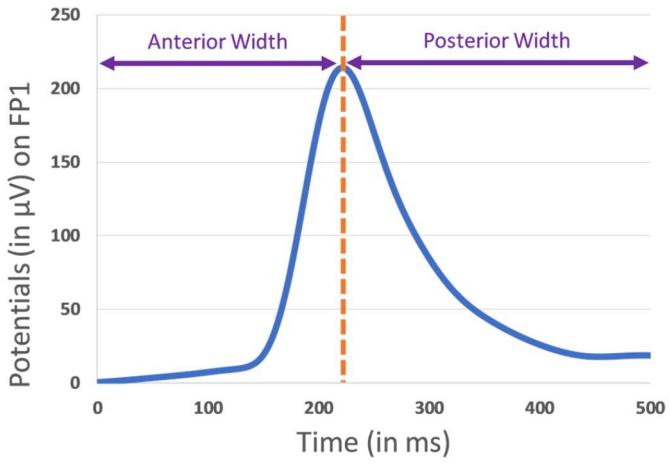

Fig. 3. A typical eyelid on the frontopolar channel FP1.

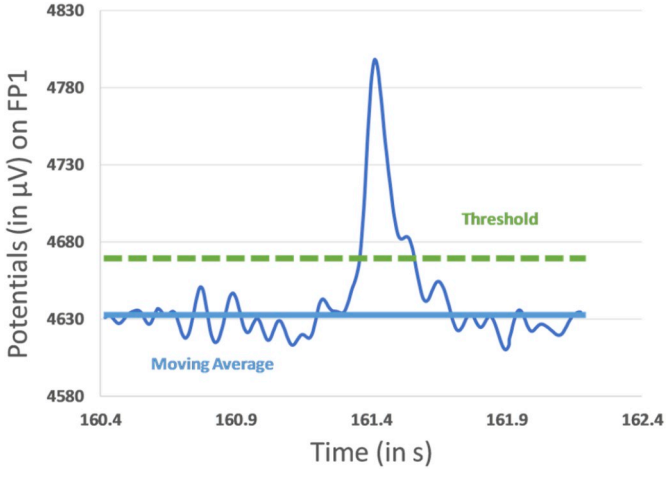

Fig. 4. Step 2: Threshold criterion.

blink accompanied by an eye movement. Though blinks between $[360-450]\,\mu V$ are not "lone" blinks per se, they are still kept in this analysis as they can occur frequently for some subject. As Fig. 5 shows, the blinks distribution varies greatly from one subject to another. The average of the maximum amplitude on FP1 per subject can be as low as $\sim 90\,\mu V$ and as high as $\sim 280\,\mu V$ (with a mean all averages of $\sim 170\,\mu V \pm 60\,\mu V$ and a mean all variances of $\sim 60\,\mu V \pm 20\,\mu V$). As mentioned above, the amplitude distribution is only analyzed on FP1. However, since the propagation remains constant for a specific subject, any distribution analysis over different electrodes would yield similar results.

3.5. Blink correction

Accordingly, the blinks are classified into $20\,\mu V$-range bins, called Class hereafter, and the Grand Averages (GA) per Class are computed. Obviously, the number of blinks per Class is not constant, but it remains similar for one subject from one session to the next. Fig. 6 illustrates a Grand Average of Class across all electrodes for one specific subject. Besides the threshold on FP1, no other heuristic choices are needed for ABC. By computing the GA on real data, noise unrelated to blinks are discarded, allowing the visualization of blinking effects even on the furthermost electrodes (e.g. OZ). Subsequently, ABC can correct blink effects lower than $40\,\mu V$ on all other electrodes, that would have been undetectable with a threshold otherwise.

As expected, the blinking signal decreases as it propagates along the scalp. However, it does not keep a same shrinkage ratio across subjects. This emphasizes yet again the necessity of a subject-dependent correction. Different Classes yield different shapes of GA, highlighting the requirement of a class-dependent correction, which is a direct consequence of intra-subject variability. Finally, Fig. 7 shows the EEG signals on FP1 before and after ABC correction.

Contrarily to others artifact correction methods, because the signals always remain in the temporal domain, there is no difference outside of the blinks. This is the reason why FP1 and the ABC correction of FP1 perfectly overlaps outside of the blinking width.

3.6. Flowchart

Fig. 8 presents a succinct flowchart presenting the main steps of ABC.

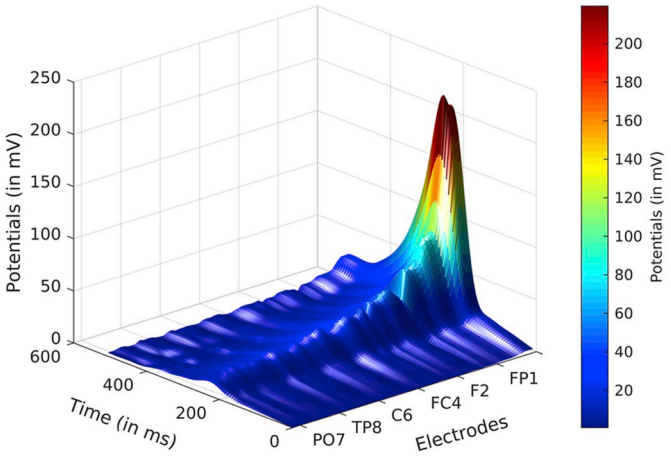

Fig. 6. Grand Average of Class $[240\ 260]\mu V$ on all electrodes for Subject S01.

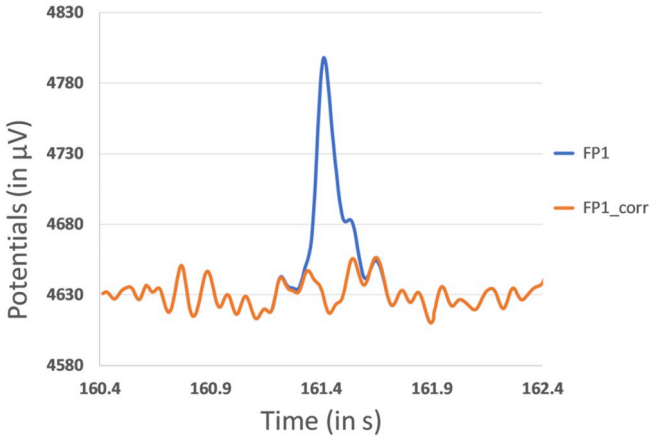

Fig. 7. Before and after ABC correction.

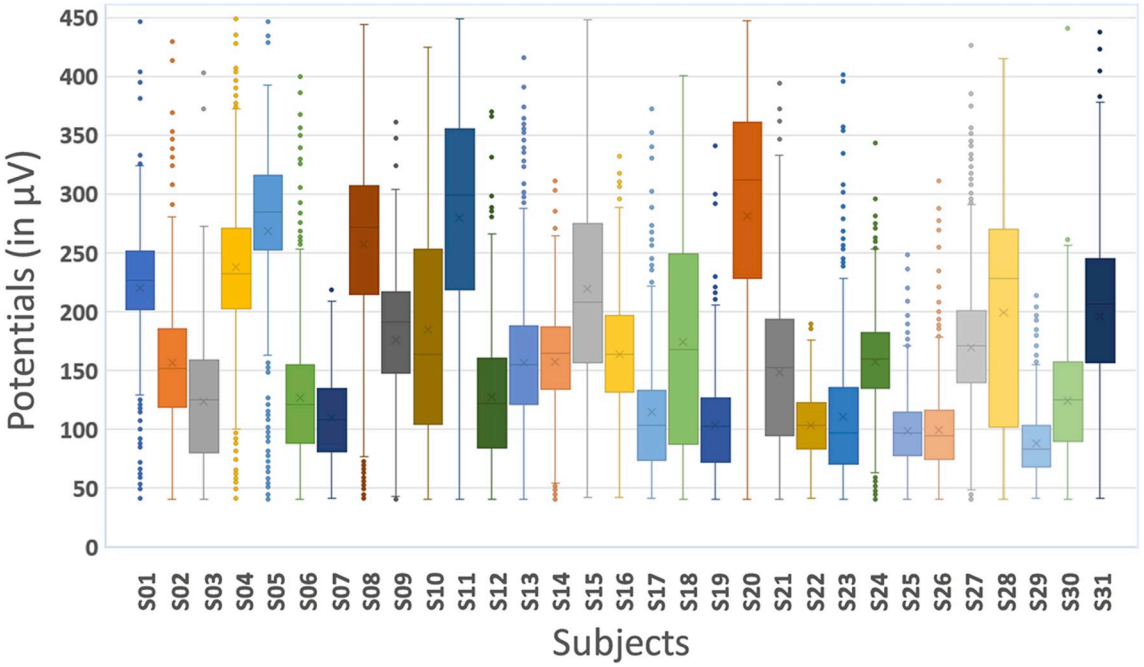

Fig. 5. Blinks distribution per subject.

In a nutshell, the correction process consists of three steps: (1) blink detection, (2) blink classification and (3) subtraction with the corresponding adjusted GA.

4. Classification

4.1. Source localization

EEG signals are "contingent to the impressed current density unleashed by cortical pyramidal neurons undergoing post-synaptic processes" [17]. As such, these signals can be used to estimate the electric neuronal activity distribution (current density vector field) on the cortex. The standardized low-resolution electromagnetic tomography (sLORETA) represents a solution to the inverse problem with zero-localization errors [29]. Event-Related Potentials (ERPs) of Left-MI and Right-MI EEG signals per subject are analyzed with the Loreta-Key software. These ERPs are converted to sLORETA in time domain per 100 *ms* range. Fig. 9 shows the differences in brain regions activation depending on the condition.

For this subject, the contralateral occipital is first activated between 0 *ms* and 100 *ms*. Then, the frontal and parietal lobes are successfully activated between 100 *ms* and 400 *ms*. Images are then compared two by two using the structural similarity index (SSIM) [30]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

Where x and y are two image signals with a μ mean intensity and a σ standard deviation, used as an estimate of the signal contrast. C_1 and C_2 are two constants with the role of avoiding instability when either $\mu_x^2 + \mu_y^2$ or $\sigma_x^2 + \sigma_y^2$ are close to zero. The timings corresponding to the smallest SSIM, i.e. presenting the highest dissimilarities, are then extracted for each subject and represent the first classification step. One might argue that activation of the brain regions should be shorter than 100 *ms*. This is true on a single-trial basis, but since the signal activation is here investigated on 40 trials average, it is more accurate to extract longer time intervals.

4.2. Riemannian Geometry

In MI-based BCI, covariance matrices are usually considered to be a reliable representation of brain signals and these spatial covariances have been used to conveniently manipulate EEG data [31]. Because covariance matrices are symmetric positive-definite (SPD), they belong to Riemannian geometry, meaning that the distance function is defined as the arc length of a minimizing geodesic. Thus, classification can be achieved by measuring the Riemannian distance between the geometric Riemannian mean of the two conditions on reference epochs, and the covariance matrices of the epochs to classify. Let $x_t \in \mathbb{R}^n$ denote the EEG signal vector at a specific time point t, with n denoting the number of recording channels. The epochs of the EEG signal, or trials, can be represented as a matrix $X_i = [x_{t+T_i} \dots x_{t+T_i+T_s-1}] \in \mathbb{R}^{n \times T_s}$ which corresponds to the i-th trial of MI started at time $t = T_i$. For this trial, the spatial covariance matrix is estimated using the Sample Covariance Matrix (SCM) $P_i \in \mathbb{R}^{n \times n}$ where [32]:

$$P_i = \frac{1}{T_s - 1} X_i X_i^T \quad (3)$$

Consider the case $n = 2$. Let $x_1(t)$ and $x_2(t)$ be the two EEG time series recorded at electrodes C3 and C4. For the i-th trial, the SCM is [33]:

$$P_i = \begin{pmatrix} Var(x_{1\,i}) & Cov(x_{1\,i},\ x_{2\,i}) \\ Cov(x_{2\,i},\ x_{1\,i}) & Var(x_{1\,i}) \end{pmatrix} \quad (4)$$

Since $Cov(x_{1\,i},\ x_{2\,i}) = Cov(x_{2\,i},\ x_{1\,i})$, 2×2 SPD matrices can be represented as points in $\mathbb{R}^3$, where each point represents a trial of length T_s. Fig. 10 shows an example of classification using only two electrodes (or $\mathbb{R}^3$ points) for visualization. For the same subject as represented in the previous paragraph, covariance matrices calculated on the appropriate time restriction interval are easily categorized. In this example, half the data is used for training and the other half is used for testing. The Riemannian geometry means are calculated for both condition on the training dataset and are then used for classification on the testing epochs. For a specific trial-to-classify, whichever distance is smaller from its covariance matrix to one of the Riemannian averages, will represent which class it belongs to.

As expected for a right-handed subject, cortical source activation will be more extended when thinking of the non-dominant hand (i.e. left hand) compared to the dominant hand (i.e. right hand). Consequently, for electrodes belonging to each hemisphere, the signal is supposed to coincide better in the former case, and thus have greater covariance matrices. This is confirmed for this specific subject on Fig. 11. Considering the source localization indications in the first step of this classification, four electrodes are selected for this second step: the two electrodes located on top of the primary somatosensory cortex, C3 and C4, and two electrodes on the parietal lobe, P3 and P4, corresponding to the angular gyrus. Because this region of the brain is responsible for transferring visual information, it seems plausible that the recordings of these electrodes would hold important information to be decoded.

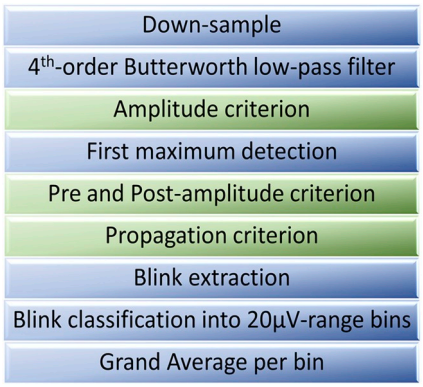

Fig. 8. ABC flowchart.

4.3. Comparison: ABC vs. ICA

Comparative studies evaluating the performance of eyeblinks artifacts removal algorithms are often carried out either on real and/or simulated data. Popular comparison techniques may use visual inspection [15], residual mutual information [34], or analysis of known ERPs (e.g. P3 amplitude [35], visual evoked potential [36]). However, when the purpose of the current paradigm is to identify a specific BCI task, herein the right vs. left hand in a Motor Imagery task, evaluating the accuracy of the signal of interest is more pragmatical. Consequently, we chose the classification accuracy as the comparison metric.

4.4. Real data

Using the same classification pipeline detailed beforehand, accuracy results are compared using ABC vs. ICA as preprocessing tools for blinks removal. ICA is carried out using EEGLAB [37] by manually identifying and removing eye artifacts Independent Components (IC). The remaining components are then converted back to the time temporal domain and saved for later classification. By design, ABC only corrects EEG signals when a blink is detected. To the contrary, ICA filters all data and thus, sometimes induces accurately classified data to be wrongly labeled. Fig. 11 shows an example of such an inaccurate change in

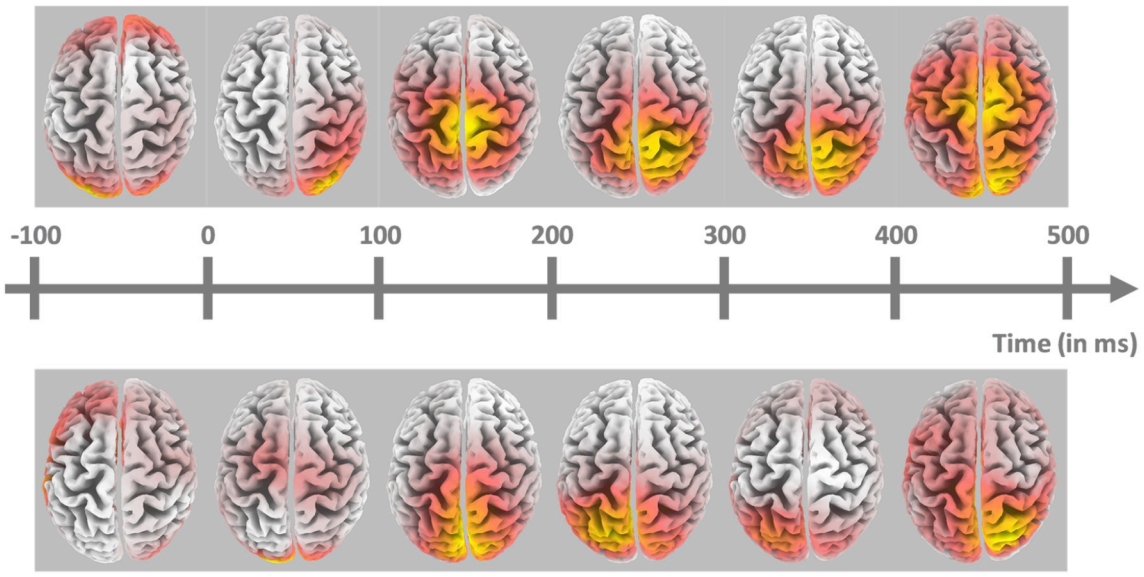

Fig. 9. Source localization with sLORETA comparing MI-Left and MI-Right per 100 ms range for one specific subject.

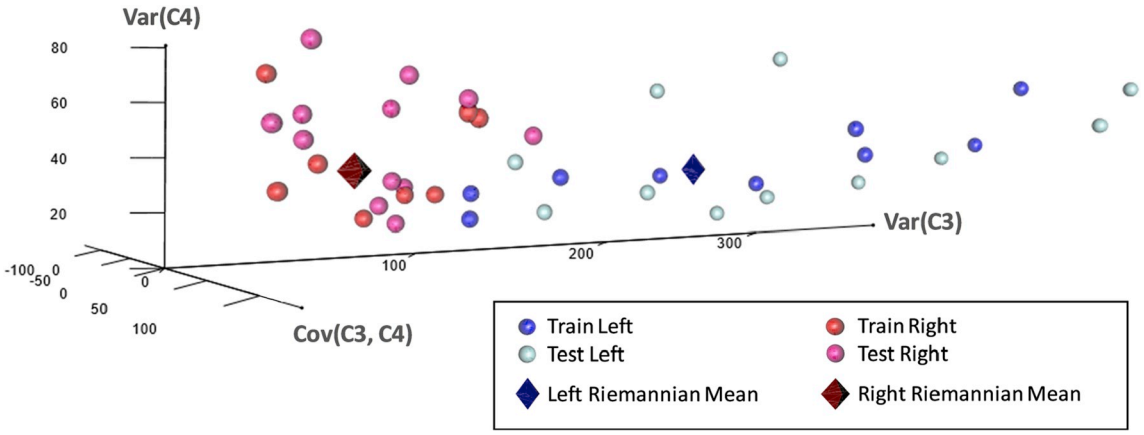

Fig. 10. Classification using Riemannian Geometry classifiers: Left MI in blue and Right MI in red.

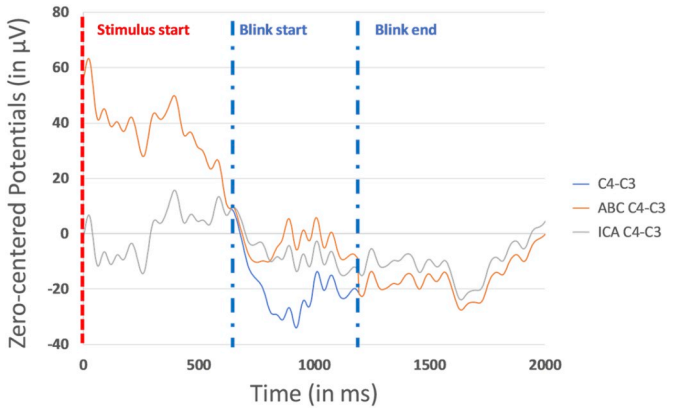

Fig. 11. Difference between C4 minus C3 in three cases: (1) No correction, (2) Correction with ABC, (3) Correction with ICA.

labelling. For comparison, the difference of zero-centered EEG signals between C3 and C4 is displayed in three cases: (1) no correction, (2) correction with ABC and (3) correction with ICA.

In this example, the stimulus is MI-Right, thus the covariance matrix is expected to be small (see Fig. 10). By filtering all signals, ICA induces a higher similarity between these two signals, hence causing the covariance matrix to increase and become wrongly classified. Moreover, Fig. 12 shows the continuous correction carried out by ICA.

Firstly, it should be noted that in average only 8% ± 4.8% of the detected blinks occur during the BCI task. Similarly, the percentage of blinks during the BCI task is again characterized by a very high subject-dependent variability, with a range from 0% to 18%. For subjects with a low blinking frequency during the BCI task, ABC has little influence over EEG signals, as the correction is only carried out when a blink is detected. On the other hand, ICA indiscriminately provides a faulty correction between the blinks. Indeed, this correction should be constantly at 0 μV, and ICA corrects with noise ranging from ± 20 μV (min −14.9 μV max 17.43 μV average 0.23 μV between the two blinks in Fig. 12). Unfortunately, this happens to be the same amplitude as the background EEG in which lies the signal of interest. Using the uncorrected data as the basis for comparison, classification is carried out and a 1-fold cross-validation is averaged over five runs. Fig. 13 shows the evolution of average accuracy for ABC and ICA correction over the whole training dataset. Similarly, the 4-fold classification averaged five times show similar percentage increase, with a 4.67% increase with ABC and a −7.91% decrease with ICA.

Furthermore, the confusion matrices for each of the three

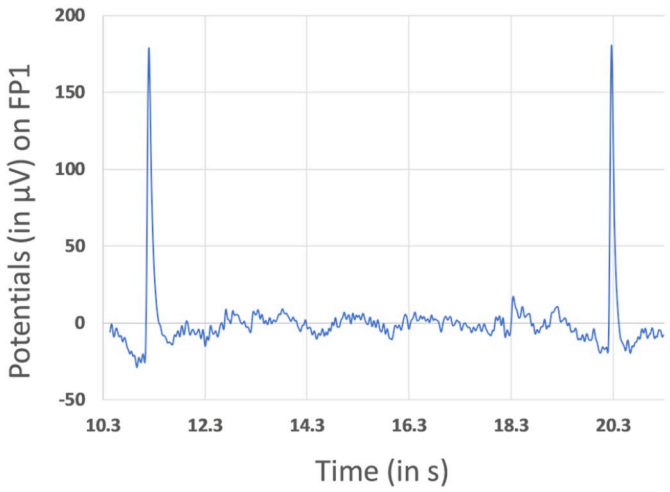

Fig. 12. Blink only Independent Component in the temporal domain, calculated with EEGLAB. Illustration of ICA continuous correction, with significant noise between two blinks.

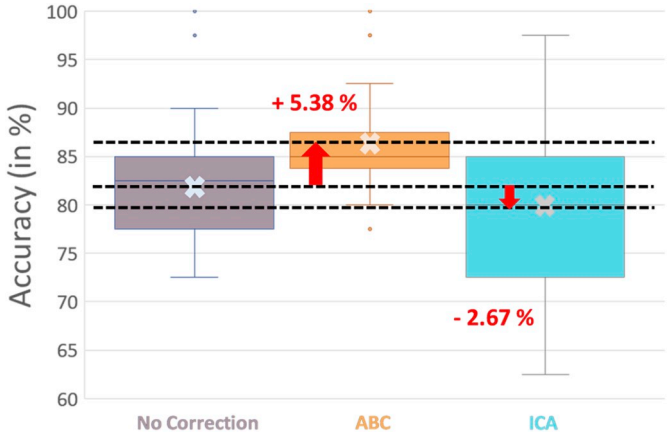

Fig. 13. Classification Accuracy Improvement when using the whole dataset with uncorrected data used as the basis of comparison.

preprocessing choices – No Correction (NC), ABC or ICA – can be represented as ellipses, as shown in Fig. 14. Each ellipse's locus of points corresponds to the regularized coefficients of the confusion matrix, where the upper-right quadrant details the positives – False Negative (FN) or True Positive (TP) – and the lower-left the negatives – True Negative (TN) of False Positive (FP). The major axis represents the percentage of correct prediction, and as such constitutes the accuracy. Moreover, the minor axis characterizes the percentage of incorrect prediction. It follows that a wide and flattened ellipse depicts a high accuracy. The overlapping of the representation of these confusion matrices allows for a clear and simple visualization of the different corrections. Since the number of positive (right MI) is the same as the negative (left MI), the initial sample data is unbiased. Nonetheless, the ellipses are represented on regularized axes (percentage of prediction).

Noteworthily, there is nearly no difference between the positives (FN and TP) of No Correction and ICA. On the contrary, ICA deteriorates the negatives (TN and FP). This is also visible when the confusion matrices are represented as relative ellipses. In this case, only the dissimilarities between two predictions are plotted, which represent the choices dilemmas. The relative sample sizes are smaller than in the absolute case, and improvement or deterioration are more clearly detectable. Indeed, horizontal ellipses display an amelioration, whereas vertical ones are characteristic of a degradation. Fig. 15 outlines the outperforming ABC preprocessing over both No Correction or ICA. On the contrary, ICA deteriorates data regarding No Correction (vertical major axis) and ABC (ICA/ABC would also have a vertical major axis, 90° rotation of ABC/ICA).

4.5. Simulated data

In addition to comparing these two algorithms with the classification accuracy as a metric, simulated data to which blinks have been artificially added can also be used. Using ABC to detect blinks, only blink-free data is extracted. Then, blink Grand Averages are artificially added with the same frequency as found in real data. Accordingly, ABC and ICA are run on the same artificial data and their results are compared in Fig. 16. This example has been chosen as it clearly shows the addition of a blinking signal over a background EEG, that may have important variations.

Outside of the blinks, the blink-free simulated data and ABC correction completely overlap. During a corrected blink, residual activity is sometimes still observed (Fig. 17). On the other hand, ICA corrected data never overlaps with the free-blink simulated data. This continuously added variability on the reconstructed EEG signal is in average close to the original data. However, compared to a typical EEG signal that ranges from $[-20;20]\,\mu V$, wrongfully adding a $\pm 20\,\mu V$ potential difference represents a potential false difference up to 200 % of the original blink-free signal.

This increase in variance within the ICA-corrected data might be caused by inconsistent solutions between and underlying EEG sources [35]. Though the reconstructed data still keeps the shape of the original,

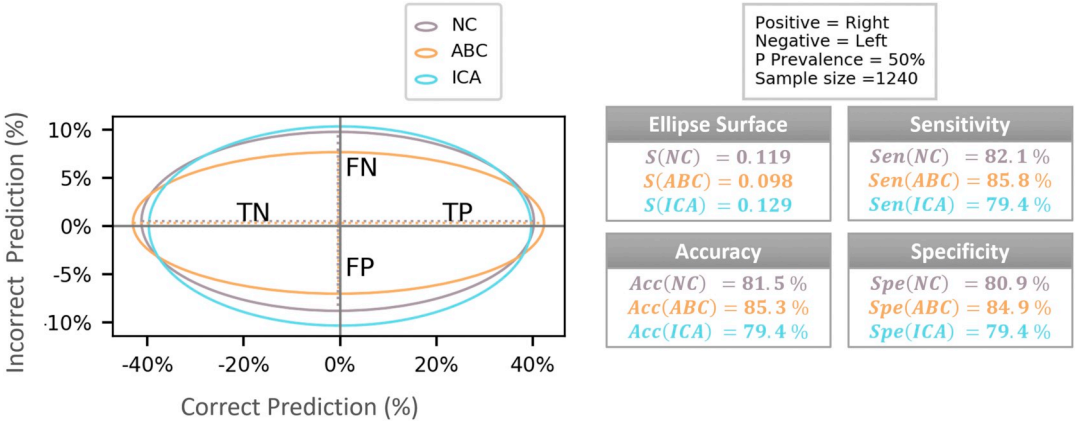

Fig. 14. Elliptic representation of the normalized confusion matrices for the three preprocessing choices (No Correction, ABC or ICA) with corresponding surfaces, sample sizes, accuracies and specificities.

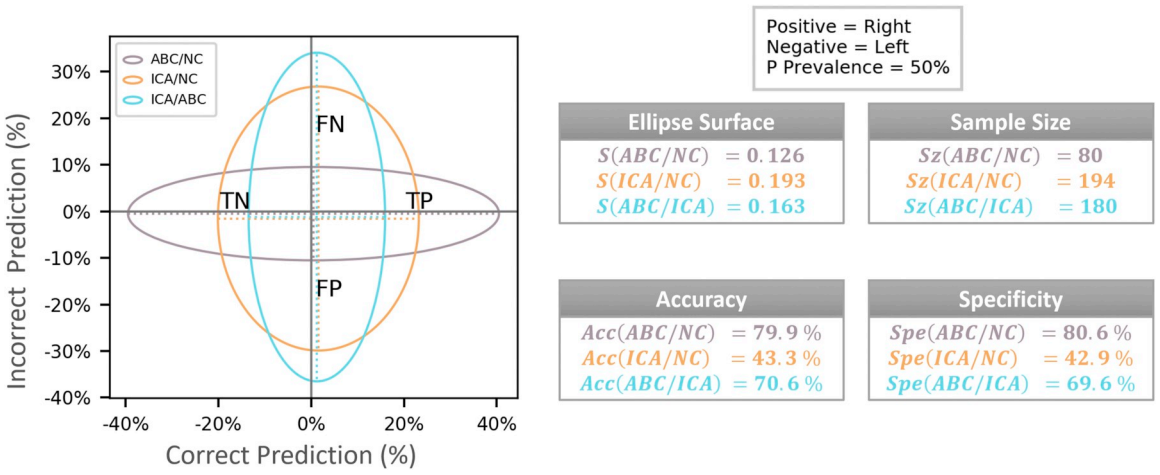

Fig. 15. Relative confusion ellipses comparing (1) ABC to No Correction, (2) ICA to No Correction and (3) ICA to ABC, with corresponding surfaces, sample sizes, accuracies and specificities.

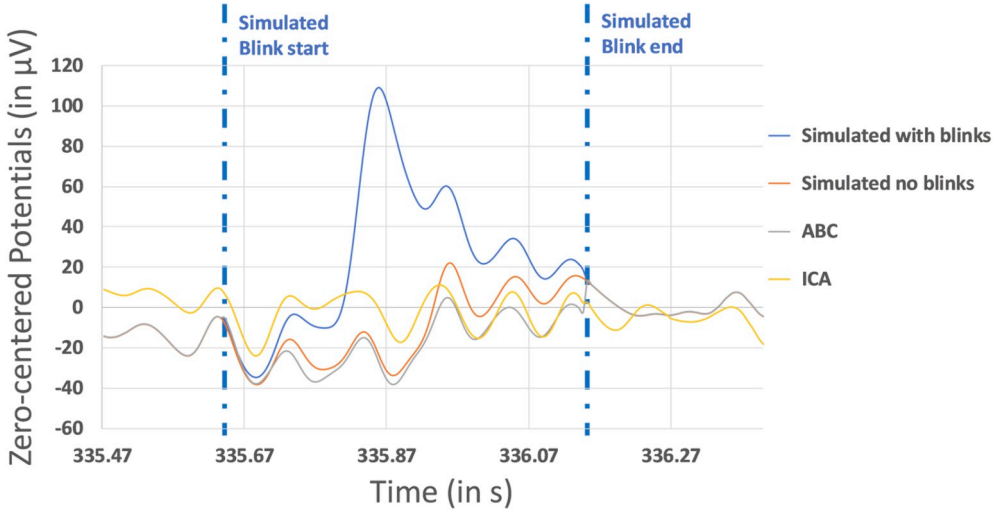

Fig. 16. Comparison ABC vs. ICA for simulated data on FP1.

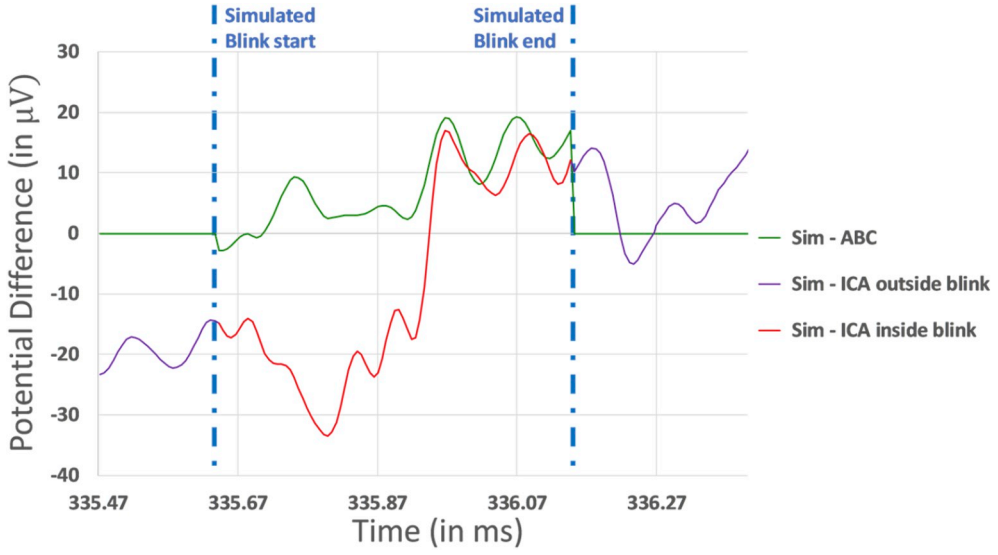

Fig. 17. Potential Difference between Simulated data vs. ABC- and ICA-corrected data on FP1.

as Fig. 16 shows, ICA uncertainty decomposition when used to remove eyeblink artifacts may influence the reconstructed EEG signal. Hence, blink-removal using ICA should probably be avoided for ERP analyses and single-trial classification. This might also explain why the former calculated classification accuracy decreases when using ICA.

5. Discussion

The complete pipeline employed in this paper is highlighted as follows: (1) choose a BCI task (or several); (2) determine the variable of interest based on the intended signal analysis process; (3) calculate the a priori sample size based on the variable observed for a few subjects; (4) carry out the experiment on the selected sample size; (5) preprocess the raw EEG data to eliminate artifacts, and; (6) process the data using the classification method chosen in step (2). Notwithstanding that there are many ways to assess the validity of an algorithm, we arbitrarily opted to restrict the comparison between ABC and ICA using two prevailing techniques. The first approach is based on the classification results of real data as an evaluation metric to decode the user's intent. The second is assessed by the accurate reconstruction of simulated data. The main drawback of the former option is that it depends on both the preprocessing and processing methods, whereas the latter's disadvantage lies in the inherent model-based blink shape. However, with the exception of the different preprocessing steps, the pipeline is completely identical in the first case. Consequently, it seems a fair measure to compare the two preprocessing algorithms. Moreover, as the added simulated blinks are formed from the real GA and as such are physiologically-based, it does not rely on any additional theoretically fitting method. Ergo, it might be more legitimate to presume similarities between a real blink and a GA than a real blink and a fitted one.

Ultimately, blink-free EEG data computed with ABC showed an average increase of 5.38 % in classification accuracy compared to the original data. Hence, our main underlying hypothesis stating that accurately correcting blinks improves classification accuracy appears to be reasonable. Additionally, the exact reconstruction of blink-free data from simulated data demonstrates the efficiency of ABC. As a comparison, the average classification accuracy decreased by 2.67 % when using ICA to remove the blinking artifact components. Reconstruction of simulated data with ICA was also inaccurate both inside and outside the blink width. Indeed, decomposing EEG data with ICA involves a linear change of basis using spatial filters. Moreover, even when no blink is detected, the ICA filtered data yielded differences up to 200 % of the original blink-free signal. Finally, the ICA reconstructed signal during a blink showed an increased variance of 30.42 % compared to the original simulated data. Therefore, ICA might not be the first choice when used for isolating and removing eyeblink-related activity, as previous research has suggested [35].

It is generally accepted that EEG signals are all extremely subject-dependent. As the blink classification analysis has shown, blinks are also highly variable among the same subject. Accordingly, ABC – as a preprocessing step – takes into account these inter- and intra-subject variabilities by providing an adaptive correction per subject and per class. Similarly, the classification step should also be adapted to this intrinsic subject-dependent variability. In this paper, classification was evaluated according to different time restrictions depending on the subject. Using source localization as a time restriction method is neither convenient, nor applicable to real-time processing. From a theoretical point of view, time restriction seems a logical choice because being as close as possible to the dynamics of the brain should probably yield more accurate results. Previously, this restriction was computed on averages of both conditions (Left-MI or Right-MI), which represents a decrease of precision. Efforts will be made in the future to improve the time selection method. Concerning the selection of electrodes, the simplified option of using similar channels across subjects was employed. Indeed, the source localization visual analysis seemed to give consistent cortical activations, corresponding to the C3, C4, P3 and P4 electrodes. By improving this first classification step for single trials, we can surely also change the selection of electrodes and apply it for real-time classification. Finally, the second classification step simply computed the minimum distance to Riemannian mean. This last part could also be improved by using other techniques such as either projecting the data in the tangent space [18] or calculating a hyperplane or surface, which would separate the two convex hulls. The latter technique would be an adaptation of support vector machine (SVM) in Riemannian geometry.

For all hypotheses and decisions made throughout this analysis, the overall conclusion is that ABC globally improves the classification. As a blink-targeting correction, ABC's advantages are ease of visualization and comprehension, possibility to be implemented in real-time, and lack of spatial filtering, which allows for more flexibility during the classification step. Currently, ABC has been developed and tested using the Comprehensive R Archive Network (CRAN R) software. Future implementations to MATLAB and/or Python are being considered.

Acknowledgment

This work was supported, in part, by the National Science Foundation of China (No.51620105002), and the Science and Technology Commission of Shanghai Municipality (No.17JC1402700).

Conflicts of interest

No competing interests to declare.

References

[1] A.S. Malik, H.U. Amin, Designing EEG Experiments for Studying the Brain, Academic Press, 2017.
[2] S.A. Baldwin, M.J. Larson, P.E. Clayson, The dependability of electrophysiological measurements of performance monitoring in a clinical sample: a generalizability and decision analysis of the ERN and Pe, Psychophysiology 52 (6) (2015) 790–800.
[3] H. Berger, Uber das Elektrenkephalogramm des Menschen, Archiv für Psychiatrie und Nervenkrankheiten 87 (Issue 1) (December 1929) 527–570.
[4] T. Thompson, T. Steffert, T. Ros, J. Leach, J. Gruzelier, EEG applications for sport and performance, Methods 45 (4) (September 2008) 279–288.
[5] M. Yadava, P. Kumar, R. Saini, P.P. Roy, D.P. Dogra, Analysis of EEG signals and its application to neuromarketing, Multimed. Tools Appl. 76 (18) (September 2017) 19087–19111.
[6] E. Niedermeyer, F.L. Da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, fifth ed., Lippincott Williams & Wilkins, 2005.
[7] N.J. Hill, J.R. Wolpaw, Brain-Computer Interface, Encycl. Neurosci., 2009, pp. 429–437.
[8] D.S. Tan, A. Nijhold, "Brain-Computer Interfaces. Applying Our Minds to Human-Computer Interaction", Human-Computer Interaction Series, Springer, London, 2010.
[9] M. Ahn, M. Lee, J. Choi, S.C. Jun, A review of brain-computer Interface games and an opinion survey from researchers, developers and users, Sensors 14 (Issue 8) (2014) 14601–14633.
[10] J.A. Urigüen, B. Garcia-Zapirain, "EEG artifact removal—state-of-the-art and guidelines", J. Neural Eng. 12 (2015) 031001.
[11] P. Anderer, S. Roberts, A. Schlögl, G. Gruber, G. Klösch, W. Herrmann, P. Rappelsberger, O. Filz, M.J. Barbanoj, G. Dorffner, B. Saletu, "Artifact processing in computerized analysis of sleep EEG – a review", Pharmacoelectroencephalography 40 (1999) 150–157.
[12] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, I. Merlet, Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches, EURASIP J. Adv. Signal Process. 127 (2012) 127.
[13] R.J. Croft, R.J. Barry, Removal of ocular artifact from the EEG: a review, Clin. Neurophysiol. 30 (Issue 1) (February 2000) 5–19.
[14] S. Romero, M.A. Mañanas, M.J. Barbanoj, A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case, Comput. Biol. Med. 38 (2008) 348–360.
[15] G. Gratton, Dealing with artifacts: the EOG contamination of the event-related brain potential, Behav. Res. Methods Instrum. Comput. 30 (Issue 1) (1998) 44–53.
[16] R. Verleger, The instruction to refrain from blinking affects auditory P3 and N1 amplitudes, Electroencephalogr. Clin. Neurophysiol. 78 (1991) 240–251.
[17] R.D. Pascual-Marqui, Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details, Methods Find. Exp. Clin. Pharmacol. 24 (Suppl D) (2002) 5–12.
[18] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update, J. Neural Eng. 15 (2018) 031005.

[19] E. Guttmann-Flury, X. Sheng, D. Zhang, X. Zhu, A priori sample size determination for the number of subjects in an EEG experiment, in: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2019.
[20] M.L. Delignette-Muller, C. Dutang, fitdistrplus: an R package for Fitting Distributions, J. Stat. Softw. 64 (4) (2015) 1–34.
[21] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, second ed., Academic Press, 1988.
[22] R. Reeves, F.A. Struve, G. Patrick, A comprehensive questionnaire for subjects undergoing quantitative research EEGs, Clin. EEG Neurosci. 29 (2) (1998) 67–72.
[23] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, Neuropsychologia 9 (Issue 1) (March 1971) 97–113.
[24] K. Kaneko, K. Sakamoto, Evaluation of three types of blinks with the use of electro-oculogram and electromyogram, Percept. Mot. Skills 88 (1999) 1037–1052.
[25] K. Naraharisetti, Removal of ocular artefacts from EEG signal using joint approximate diagonalization of eigen matrices (JADE) and wavelet transform, Can. J. Biomed. Eng. Technol. 1 (4) (2010).
[26] A.R. Bentivoglio, S.B. Bressman, E. Cassetta, D. Carretta, P. Tonali, A. Albanese, "Analysis of blink rate patterns in normal subjects", Mov. Disord. 12 (Issue 6) (1997) 1028–1034.
[27] A.A.V. Cruz, D.M. Garcia, C.T. Pinto, S.P. Cechetti, Spontaneous eyeblink activity, Clin. Sci. 9 (Issue 1) (2011) 29–41.
[28] M.G. Doane, Interaction of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink, Am. J. Ophthalmol. 59 (1980) 507–516.
[29] R.D. Pascual-Marqui, D. Lehmann, M. Koukkou, K. Kochi, P. Anderer, B. Saletu, H. Tanaka, K. Hirata, E.R. John, L. Prichep, R. Biscay-Lirio, T. Kinoshita, Assessing interactions in the brain with exact low-resolution electromagnetic tomography, Phil. Trans. A Math. Phys. Eng. Sci. 369 (1952) (2011) 3768–3784.
[30] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (No. 4) (2004) 600–612.
[31] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, "Riemannian geometry applied to BCI classification", in: 9th International Conference Latent Variable Analysis and Signal Separation, 2010, pp. 629–636.
[32] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain-computer Interface classification by riemannian geometry, IEEE Trans. Biomed. Eng. Ins. Electr. Electr. Eng. 59 (4) (2012) 920–928.
[33] M. Congedo, A. Barachant, R. Bhatia, Riemannian Geometry for EEG-Based Brain-Computer Interfaces; a Primer and a Review, Brain-Computer Interfaces, 2017.
[34] S. Hoffmann, M. Falkenstein, The correction of eye blink artefacts in the EEG: a comparison of two prominent methods, PLoS One 3 (8) (2008) e3004.
[35] M.B. Pontifex, K.L. Gwizdala, A.C. Parks, M. Billinger, C. Brunner, Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts, Psychophysiology 54 (2017) 386–398.
[36] S. Blum, N.S.J. Jacobsen, M.G. Bleichner, S. Debener, A riemannian modification of artifact subspace reconstruction for EEG artifact handling, Front. Hum. Neurosci. 13 (2019) 141–156.
[37] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21.