Journal of Neural Engineering

PAPER • OPEN ACCESS

o-CLEAN: a novel multi-stage algorithm for the ocular artifacts' correction from EEG data in out-of-the-lab applications

To cite this article: Vincenzo Ronca *et al* 2024 *J. Neural Eng.* **21** 056023

Journal of Neural Engineering

OPEN ACCESS

RECEIVED
22 March 2024

REVISED
2 August 2024

ACCEPTED FOR PUBLICATION
16 September 2024

PUBLISHED
30 September 2024

PAPER

o-CLEAN: a novel multi-stage algorithm for the ocular artifacts' correction from EEG data in out-of-the-lab applications

Vincenzo Ronca[1,2,*], Gianluca Di Flumeri[2,3], Andrea Giorgi[2,4], Alessia Vozzi[2,4], Rossella Capotorto[4], Daniele Germano[1,2,3], Nicolina Sciaraffa[2], Gianluca Borghini[2,3], Fabio Babiloni[5,6] and Pietro Aricò[1,2]

[1] Department of Computer, Control, and Management Engineering, Sapienza University of Rome, 00185 Roma, Italy
[2] BrainSigns Srl, Industrial Neurosciences Lab, 00198 Rome, Italy
[3] Department of Molecular Medicine, Sapienza University of Rome, 00185 Roma, Italy
[4] Department of Anatomical, Histological, Forensic and Orthopaedic Sciences, Sapienza University of Rome, 00185 Roma, Italy
[5] Department of Physiology and Pharmacology 'Vittorio Erspamer', Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[6] College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310005, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: vincenzo.ronca@uniroma1.it

Keywords: EEG, ocular artefacts, signal processing

Abstract

In the context of electroencephalographic (EEG) signal processing, artifacts generated by ocular movements, such as blinks, are significant confounding factors. These artifacts overwhelm informative EEG features and may occur too frequently to simply remove affected epochs without losing valuable data. Correcting these artifacts remains a challenge, particularly in out-of-lab and online applications using wearable EEG systems (i.e. with low number of EEG channels, without any additional channels to track EOG). *Objective.* The main objective of the present work consisted in validating a novel ocular blinks artefacts correction method, named multi-stage OCuLar artEfActs deNoising algorithm (o-CLEAN), suitable for online processing with minimal EEG channels. *Approach.* The research was conducted considering one EEG dataset collected in highly controlled environment, and a second one collected in real environment. The analysis was performed by comparing the o-CLEAN method with previously validated state-of-art techniques, and by evaluating its performance along two dimensions: (a) the ocular artefacts correction performance (IN-Blink), and (b) the EEG signal preservation when the method was applied without any ocular artefacts occurrence (OUT-Blink). *Main results.* Results highlighted that (i) o-CLEAN algorithm resulted to be, at least, significantly reliable as the most validated approaches identified in scientific literature in terms of ocular blink artifacts correction, (ii) o-CLEAN showed the best performances in terms of EEG signal preservation especially with a low number of EEG channels. *Significance.* The testing and validation of the o-CLEAN addresses a relevant open issue in bioengineering EEG processing, especially within out-of-the-lab application. In fact, the method offers an effective solution for correcting ocular artifacts in EEG signals with a low number of available channels, for online processing, and without any specific template of the EOG. It was demonstrated to be particularly effective for EEG data gathered in real environments using wearable systems, a rapidly expanding area within applied neuroscience.

1. Introduction

The electroencephalographic (EEG) signal is one of the most informative electrophysiological biosignals, employed in a variety of research areas and biomedical applications, such as brain–computer interfaces (BCIs), mental states assessment, neurofeedback, and many others [1–10]. The recent effort in developing even more effective EEG wearable devices, with a few number of sensors, enables this technology to an employment outside of the research labs. An example of use consists in the EEG-based BCI systems, that are based on outputs derived from brain activity arising with (i.e. active BCIs, where communication or control is the user's main task) or without the purpose of voluntary control (passive BCIs, to enable communication through measures of ongoing mental/emotional states of the user engaged on a different task). Such systems rely on the EEG signal analysis and interpretation through specific features (in time and frequency domain) estimation. Anyhow, EEG-based features are affected by a low signal to noise ratio, and several confounding factors could distort or hide the desired physiological information.

In this regard, the artefact generated by the ocular movements (i.e. ocular blinks and saccades) contribution represents one of the most confounding factors, much studied in bioengineering EEG processing research, because of two main reasons:

– The power spectrum main content of ocular movements overwhelms informative EEG- related features (i.e. the EEG brain rhythms), since the ocular movements bandwidth is comprised between 3 and 15 Hz [11], which is the same frequency range in which relevant neurophysiological contents are identified, such as the EEG theta and alpha bands in frequency domain. Even in time domain analysis, ocular artifacts could interfere for example with the Evocked Potentials extraction and related assumptions.
– The occurrence of ocular artifacts may be too high to simply try to remove EEG epochs containing them (e.g. 12–18 ocular blinks per minute [12]).

Therefore, the presented study was focused on methodologies to deal with this specific category of EEG signal artefacts, i.e. the eye movement-based artefacts with a particular regard to ocular blinks, by correcting them, therefore without losing any data.

In particular, the ocular blinks contaminate the EEG signal content as a result of major physiological sources: the corneo-retinal dipole, eyelid movements, and extraocular muscles [13, 14]. The corneo-retinal dipole represents the positive charge of the cornea relative to the retina, causing potential changes at EEG sensors during eyeball rotation. Eyelid movements, which is more consistent during the ocular blinks, introduce high-amplitude potential field changes. Additionally, extraocular muscle contractions impact on the EEG signal amplitude. It has to be noted that within the 3–15 Hz frequency range, the corneo-retinal dipole and eyelid-induced artifacts are prominent.

As mentioned before, these kinds of artifacts cannot simply be deleted, by removing the EEG segments containing the artifacts themselves, because most of the EEG signal would be lost. For this reason, the EEG segments containing ocular blinks artifacts need to be corrected, i.e. depurated by the ocular artifacts' contributions.

There are several bioengineering processing techniques, which are efficient in removing the ocular blinks contribution, but usually the number of EEG channels has to be sufficiently high (i.e. high-resolution EEG), and mostly for offline analyses (i.e. non-causal techniques). More specifically, a proficient series of already validated approaches are available for the correction of EEG signals from ocular-based artifacts. In this regard, the independent component analysis (ICA) appears to be the gold-standard [15]. The ICA technique ensures high grade of reliability and accuracy when a high number of EEG channels is available for the EEG signal collection. The technical conceptualization of the ICA does not allow a sufficient grade of efficiency without negatively affecting the physiological content of the EEG signal when its classical implementation is applied to a low number of EEG channels [16]. In fact, a low number of EEG channels corresponds to a likewise low number of independent components, in which the ocular component could not be effectively separated from the EEG signal. Therefore, with a low number of EEG channels the ICA application implies a non-negligible EEG signal distortion. Other techniques for the eye movement-based artefacts correction by having few EEG channels exist, they are based on regression methods. In this regard, the methods based on Gratton & Coles approach [17] reveals to be reliable and compatible with online implementations, if an independent ocular component (e.g. recorded by an electrooculographic (EOG) signal) is available. This information is needed, to be able to remove components of the recorded signals associated to blinks, leaving intact those ones related to EEG.

Indeed, although EOG recording is easy to employ during experiments in lab, it is not suitable for out-of-the-lab applications, since it requires a bipolar channel close to one eye, resulting in a too invasive setup. If ocular activity is instead derived directly from a EEG channel (e.g. located within the frontal cortex), the method induces a distortion in the surviving EEG signal, negatively impacting on the neurophysiological information that can be obtained from [18].

All the above-mentioned limitations result to be relevantly burdensome for the online and out-of-the-labs applications, that require a few numbers of EEG sensors and no other additional sensors for EOG recording. In this regard, it has to be noted that different approaches have been proposed in the context of the eye movement-based artifact correction for the EEG. Recently, Kobler and colleagues [19] developed a novel approach based on the identification and removal of the ocular blinks-related EEG artefacts subspace. Such an approach was firstly proposed in its original implementation by Parra and colleagues [20] and then by Kobler and colleagues in 2017 [21]. In particular, this method is based on the ocular artifacts subspace subtraction for identifying and correcting the EEG ocular artifacts both offline and in real time. The advantages with the respect to the standard methods consist in the possibility of training the algorithm on calibration data, in order to apply it even online for correcting the EEG ocular artifact with a high grade of specificity [19]. Such a method was demonstrated to be reliable in correcting the EEG ocular blink artifacts when using a high EEG channels number, and it was also employed in the current study to compare its accuracy with the proposed methodology for out-of-the-lab applications, i.e. EEG data collection through a wearable system equipped with a low EEG channels number, representing the main aim of this research.

The recent scientific literature proposed other methods for identifying and correcting the EEG ocular blink artifacts. For example, the multichannel Wiener filter (MWF), proposed by Somers and colleagues [22] and developed as a generic EEG artifact removal algorithm, fall within the methods suitable to be used with a low number of EEG sensors. The method is based on the theory of optimal filters, able to dynamically reduce noise (e.g. ocular blink artifact) related frequency components from a signal (e.g. the EEG signal). The method is powerfull, if it is previously calibrated on data containing both the signal with (i.e. EEG plus blink) and without (i.e. EEG without blink) noise. Anyhow the method suffer of two drawbacks. The first one is related to the identification of data for the calibration of the filter: this part of the algorithm is indeed done manually (i.e. by visual inspection of the operator), making the method not suitable for out-of-the-labs applications. In addition, the filtered signal may result to be distorted with respect to the original one, depending on the quality of the generated filter, since of course, the final filter is applied on the whole EEG signal, both with and without noise (ocular artifacts).

Therefore, it appears to be clear that the actual state of art related to the ocular blinks-based EEG artefacts identification and correction, within out-of-the-lab and online applications through EEG wearable systems, presents still open gaps.

In fact, even by considering the most recent advancements in wearable devices and EEG-based applications (i.e. BCI), the employment of a wrong method for the ocular artifact correction could completely hide or distorce the neurophysiological content of the EEG signal, making the application itself useless.

Therefore, the main objective of this work consisted in validating a novel regression-based ocular blinks artefacts correction method compliant with online processing and with a low EEG channels number, and without any additional channel to track EOG activity (i.e. for out-of-the-labs applications). Such a method, named 'Multi-stage OCuLar artEfActs deNoising algorithm' (o-CLEAN), needs a calibration phase, in which ocular blinks from a controlled recording at rest are detected and used to train the algorithm, in order to be able to identify and correct the blinks contribution in other coming data (even online). The novelty of the o-CLEAN method is linked to the generation of a specific template related to the EOG activity from the available EEG channels, able to minimize the mutual contamination issue and signal distortion during the regression phase. In particular, as mentioned before, regression algorithms such as the approach proposed by Gratton and Coles [17], required an additional EOG channel, to properly correct the blinks contribution, without inducing distortions in EEG signal. Indeed, the requirement of an additional EOG channel is not compatible with out-of-the-lab and passive BCI application, in which wearable EEG headsets are employed. Moreover, an additional EOG channel would imply a negative impact on the EEG system invasiveness even in laboratory settings. Evolutions of this kind of approach were recently proposed by the scientific community. In this regard, the REBLINCA method [18], which was selected to be included in the presented comparative study and which will be technically described within the *Methods* section, overtakes the Gratton and Coles-based approaches. In fact, the REBLINCA does not require an additional EOG channels, since it estimates the EOG-like template by band-pass filtering the most frontal EEG channel (Fpz channel is suggested by the authors). Actually, this is just a mitigation strategy, since within the band of the filters the mutual contamination is still present, and even at higher extent, between the EEG channel used for the EOG-like template estimation, and the other EEG channels to be corrected. Moreover, it was demonstrated that such a kind of regression-based methods consistently and negatively impact on the neural content of the frontal and anterofrontal EEG channels, due to their closeness with the EEG channel selected as regressor.

To this regard, the proposed o-CLEAN method aims at fully addressing the abovementioned dimensions through its innovative multi-stage approach and the concomitant use of adaptive filtering theory in place of traditional band-pass filtering. In fact, the o-CLEAN algorithm (i) does not require any additional EOG channels, (ii) it performs an initial automatic and robust blinks detection from an available EEG channel, (iii) it deploys the Multi-Channel Wiener filtering approach to estimate the EOG-like template, and (iv) finally uses this template for the regression and correction stages.

In order to demonstrate the effectiveness of the o-CLEAN method with the respect to the state-of-the art, two EEG datasets were employed for assessing two dimensions of the proposed algorithm: (i) the performance of the proposed methodology in detecting and removing the ocular blinks-related EEG artefacts, and (ii) the performance of the proposed methodology in preserving the neurophysiological content of the EEG signal, i.e. by evaluating how the method impacted on the neurophysiological content of the corrected EEG signal. These two aspects represented the (i) specificity of the methods in detecting the true positive ocular blink artifacts, i.e. when the EEG signal was affected by the ocular blink artifacts (IN-blink), and (ii) the sensibility in detecting the true negative ocular blink artifacts, i.e. when the EEG signal was not affected by the ocular blink artifacts (OUT-blink).

2. Material and methods

Two datasets were employed within the present study. The first EEG dataset (LAB dataset) was collected in a very controlled environment by using high density EEG channels system, and already available online, and employed by other studies (laboratory settings). The second dataset (REAL dataset) was instead collected within a naturalistic (i.e. out-of-the-lab) experiment, through a wearable EEG system, to test the realiability of the method in a realistic settings.

All the EEG signal included in the present work were pre-processed by applying a band-pass filtering [2–28] Hz.

The assessment of the o-CLEAN method performances, in terms of sensibility in detecting and correcting the ocular blinks artifacts and in terms of specificity for reliably preserving the neurophysiological content of the EEG signal, have been compared with the current state of the art available techniques. More technical details are reported in the following subsections.

2.1. EEG dataset in controlled settings (LAB dataset)

This EEG dataset is stored on the public repository OSF and published by Kobler and collegaues [19]. Such a EEG dataset included EEG recordings from thirty-six (36) participants. The experimental paradigm consists in four distinct conditions: Rest, Horz, Vert, and Blink. Participants were instructed, depending on the specific condition: to keep their eyes open and focus on a central stimulus without blinking (Rest), track a moving stimulus along a horizontal or vertical axis (Horz /Vert), or simulate an involuntary blink by lowering and (iv) raising their eyelid when the stimulus size decreases (Blink). For the purposes of the present study, only the Rest and Blink conditions have been considered. The EEG activity was captured using 64 EEG channels located according to the 10–10 international system, with the right mastoid serving as the reference and Afz as the ground. Simultaneously, EOG activity was recorded with six electrodes positioned near the outer canthi, infraorbital, and superior orbital regions. Active EEG channels (actiCAP, Brain Products GmbH, Germany) and a biosignal amplifier (BrainAmp, Brain Products GmbH) were employed for synchronous recording of EEG and EOG activity. The sampling frequency was 200 Hz. The average EEG signal length corresponded to 80 s per subject. Such a dataset was employed as benchmark for the EEG artefacts identication and correction methodologies since the high density EEG channels and the EOG channels allowed to compare the proposed methodology with other available state-of-art techniques, in terms of sensibility and specificity.

2.2. EEG dataset in naturalistic settings (REAL dataset)

This EEG dataset was collected during an experimental protocol performed in a real driving condition [23]. More specifically, thirty-eight (38) participants were involved within the experiments and they performed real driving tasks in standard and monotonous traffic and road conditions. Additionally, the EEG signal was collected during resting conditions. The EEG activity was collected using the Mindtooth Touch EEG system (Brain Products GmbH, www.mindtooth-eeg.com) [29, 30]. It consists of eight Ag/AgCl water-based electrodes placed according to the International 10–20 system (Afz, AF3, AF4, AF7, AF8, Pz, P3, and P4) plus ground and reference electrodes placed respectively on left and right mastoids. The sampling frequency was 125 Hz. The average EEG signal length corresponded to 600 s per each participant, in total. More specifically, it was considered a 60 s-EEG segment collected while the participant kept eyes closed as resting condition, while the remaining part corresponded to the EEG collected while the participants were performing naturalistic real driving.

2.3. o-CLEAN processing steps

This paragraph will present the processing steps of the o-CLEAN algorithm. It has been named 'Multi-stage OCuLar artEfActs deNoising algorithm' (o-CLEAN) and it can be considered as an evolutionary approach fusing the regression-based techniques and the adaptive filtering theory. More specifically, the proposed method can be envisioned as composed by two main blocks: the first one corresponding to the ocular blink automatic detection step, and the second one corresponding to the ocular blinks contribution removal. Regarding the first processing block, the o-CLEAN method was implemented in order to minimize the

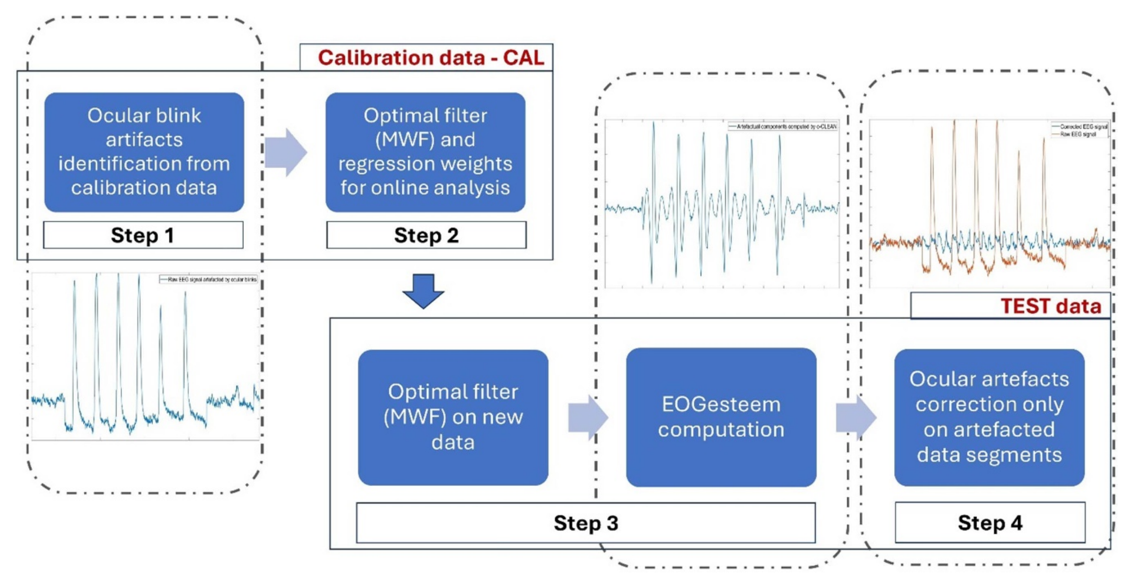


Figure 1. Operational schema of the o-CLEAN method processing steps.

risk of mutual contamination [24] when estimating the EOG pattern from the EEG signal, which characterizes the classical regression-based approach, as the REBLINCA previously proposed by Di Flumeri *et al* [18]. Additionally, the proposed o-CLEAN was conceptualized for avoiding false positive during the ocular blinks automatic detection, which is a further limitation characterizing the regression-based methods, especially when applied on EEG signals collected in non-controlled settings [25].

[18, 24]. Therefore, a detailed description of the different processing stages of the proposed method is provided below (figure 1).

1. The first step consists in the 'ocular blink artifacts identification' along the EEG signal. This identification process is performed through a regression-based detection procedure, derived from the approach firstly proposed by the REBLINCA method [18]. In particular, the original method was designed to identify all 'potential blinks', in order to perform correction only when they occur. Since the correction is performed by using a pre-processed frontal EEG channel, even if the correction is performed on a false positive its impact would be negligible, therefore it was designed to favour false positive instead of false negative. However, in this case the blink template has to be as much reliable as possible, i.e. both false positive and negative should be avoided, otherwise the following step (please refer to step 2) would be negatively affected. To do so, some additional criteria have been implemented in order to discard false positive from the blinks initially identified by the method, i.e. which was previously proposed and validated as the name of REBLINCA. These threshold-based criteria have been developed empirically, by taking advantage from the evidence provided by previous scientific studies [11]. Specifically, for each blink initially detected by the method, three parameters are calculated: length, skewness, and standard deviation. If the length is less than 0.15 s (minimum duration of a blink), or the skewness related to the potential ocular blink event is less than 0.2 (the blink-related data distribution is highly skewned, being a large positive variation of the signal), or the ratio between the standard deviation of the blink and that of the EEG clean signal is less than 3 (the blink signal is almost one order of magnitude higher than EEG signal), the identified blink is considered a 'false positive' and discarded. This first step (i.e. ocular blink artifacts identification) is performed by considering a 60 s-calibration run (CAL), which allows to specifically characterize the individual ocular blinking pattern. In fact, this calibration run is intended as a trial in which the participant is just naturally blinking while looking at a fixation point in front of him/her. Such a calibration run was selected for the EEG signal processing related to each participant in the present study. The outcome of this step is a labelling vector of 0 (no-blink) and 1 (blink) with the same sampling frequency of the original EEG data.
2. In the second step, the algorithm is trained for extracting specific filtering coefficients from the CAL trial, according to the Wiener filtering

approach implemented by Somers and colleagues [22]. Such an approach relies on a data model receiving in input a signal exclusively constituted by the artefacts to be removed (i.e. the ocular blinks),

$$y(t) = n(t) + d(t),$$

3. where, $y(t)$ corresponds to the multi-channel modelled EEG signal, $n(t)$ corresponds to the true neural signal, and $d(t)$ represents artifacts of other origin superimposed on the neural signal. Such a data model generates a template to maximize the differences between the clean signal $n(t)$ (i.e. pure EEG signal) and the artefactual signal $d(t)$ (i.e. pure ocular blinks). Thus, the optimal filter spectrum and related parameters, specific on ocular blinks components—$d(t)$, within a clean EEG signal—$n(t)$ is generated by following the Wiener optimal filter theory.
4. The optimal filter generated from CAL data, is now applied to new EEG data, even online (TEST) in order to extract from each EEG channel an ocular blinks component. In other words, for each channel, the EEG contribution but not the ocular artifact is filtered out. A unique ocular blinks artefactual component (EOGesteem) is then computed by averaging the aforementioned components (i.e. one for each EEG channel), in order to obtain an optimal representation of the ocular blinks contribution over the entire EEG channels set, minimally correlated to the original EEG signals.
5. The final step of the proposed o-CLEAN algorithm corresponds to the ocular blinks artifact correction by applying a regression-based procedure, similar to the one introduced by Gratton and Cole [17], by considering the EOGesteem signal as the EOG component. Therefore, by employing regression between the EOGesteem component and each of the EEG channels, it is estimated a weight that depends on the degree of dependence between the EOG pattern and that specific EEG channel, i.e. how big is the impact of the ocular artefact. Then, the EOGesteem component is subtracted (correction phase) from each EEG channel in a weighted way. For online analysis each weight is computed from CAL data, for offline analysis weights can be directly computed from TEST data. To minimize the eventual distortion introduced by the method the EEG signal has been corrected exclusively in correspondence of blinks, by using blinks detection criteria described in [18].

2.4. State of the art ocular blinks correction algorithms for EEG signals

The proposed o-CLEAN method was compared with the most widespread algorithms for the EEG ocular blinks identification and correction. Therefore, all the algorithms considered within the present research were applied to both the two EEG dataset (i.e. LAB and REAL setting):

- **Gratton**: represents the classical implementation of the artefacts identification and correction method proposed by Gratton and colleagues [17]. Such a method employs multi-channel approach to differentiate ocular artifacts from brain-generated signals. This method requires the recording of additional electrooculogram (EOG) channels, which specifically capture eye-related activity. The Gratton method then employs mathematical algorithms to estimate by regression the blink-related content on EEG channel and to proportionally remove the ocular artifact components from the EEG data. Even if such a method constitutes one of the most reliable and validated EEG ocular artefacts correction methods, indeed it is not compatible with the application within out of the lab environments.
- **AMICA**: the Adaptive Mixture ICA [26] is an advanced form of ICA [27], a signal processing technique, particularly in the context of ocular artifacts. It works by decomposing EEG data into independent components, where each component represents a different underlying source of neural or non-neural activity. The AMICA does not strictly require EOG channels, but its efficiency relevantly depends on the EEG channels number. We have used this method as gold standard, in the controlled dataset.
- **Sgeyesub**: is a recent approach developed and validated by Kobler and colleagues [19]. This method removes the subspaces that explain the variance introduced by the cornea-retinal dipole and eyelid movements from the EEG activity.
- **REBLINCA**: this method, introduced by Di Flumeri *et al* [18] is based on regression and the statistical identification of the ocular blinks occurrences. This method is compliant with online applications, and it requires few EEG channels. However, the previous research which proposed and validated such a method revealed that it is prone to impact the spectral EEG contents in terms of distortion, especially among the frontal and anterofrontal EEG channels, even if such a signal distortion resulted to be lower compared to the

Gratton approach, if EOG is not available, and a frontal EEG channel (i.e. AFz) is used instead.

- **MWF**: corresponds to the implementation of the technique based on the Wiener filtering proposed by Somers and colleagues [22]. The filter is applied to the whole EEG signal (i.e. both with and without the ocular blink contribution). This method relies on the ocular blink artifact multi-channel estimations, by computing a low-rank approximation of the artifact covariance matrix based on the generalized eigenvalue decomposition. The approach was validated to be reliable for the generic EEG artifact removal. However, dependently from the signal—noise overlap, such method could generate potentially relevant signal distortions.

2.5. Performed analyses

Each of the listed methodologies were applied to the above-described dataset in order to evaluate their reliability and effectiveness in detecting and correcting the EEG ocular artifacts. In particular, different metrics were selected among the scientific literature to quantify the performance of the tested methodologies:

- Pearson correlation computed between the non-corrected EEG signal and the EOG channel, and between the corrected EEG signal and the EOG channel during the Blink condition. Such a metric was employed for characterizing the methods' efficiency in detecting and correcting the ocular blinks-related EEG artifacts. In fact, the decrease of correlation after the methods application was hypothesized to be linked with the ocular artifact-based content removal from the corrected EEG signals, since such a content was the most prominent within the signal collected through the EOG channels.
- Mutual information (MI) computed between the corrected and the non-corrected EEG signal and the EOG channel during the BLINK condition. This metric was computed by filtering the EEG signal between 2 and 15 Hz, which is the spectra in which the ocular blinks contribution was demonstrated to be more prominent [28]. The MI as well was employed as an efficiency indicator of the tested methods. The metric was computed by technically implementing the following:

$$MI(X,Y) = \iint_{X,Y} P_{(X,Y)}(x,y)\log\left(\frac{P_{(X,Y)}(x,y)}{P_{(X)}(x)\,P_{(Y)}(y)}\right)\mathrm{d}x\mathrm{d}y$$

- where X and Y corresponded to the ocular blink component and the non-corrected or corrected EEG signal, $P_{(X,Y)}$ was the joint probability density function of X and Y, $P_{(X)}$ and $P_{(Y)}$ were the marginal probability density functions of X and Y respectively.
- Frequency correlation (FC) computed between the corrected and the non-corrected EEG signal and the EOG channel during the BLINK condition. Similarly, the FC was applied after filtering the EEG signal between 2 and 15 Hz. The analytical definition of this metric corresponds to the following [31]:

$$fc = \frac{1}{2}\frac{\sum_{\omega_2}^{\omega_1} x_{\mathrm{corr}} \cdot x_{\mathrm{orig}}^{*}}{\sqrt{\sum_{\omega_2}^{\omega_1} x_{\mathrm{corr}}^{2}} \cdot \sqrt{\sum_{\omega_2}^{\omega_1} x_{\mathrm{orig}}^{2}}}$$

- Power spectral density (PSD) preservation computed between the corrected and non-corrected EEG signal within the resting condition. This metrics was evaluated in three different EEG frequency bands (i.e. theta, alpha, and beta) in order to quantify the signal distortion related to each of the tested methodologies. Such metrics were computed as follows:

$$\mathrm{PSD}_{\mathrm{freq-preservation}} = \frac{\mathrm{PSD}_{\mathrm{freq-corr}}}{\mathrm{PSD}_{\mathrm{freq-uncorr}}}$$

- where $\mathrm{PSD}_{\mathrm{freq-preservation}}$ corresponded to the EEG PSD preservation computed in each specific EEG frequency band (i.e. theta, alpha, and beta), $\mathrm{PSD}_{\mathrm{freq-corr}}$ corresponded to the EEG PSD computed in each specific band of interest along the corrected EEG trials, and $\mathrm{PSD}_{\mathrm{freq-uncorr}}$ corresponded to the EEG PSD computed in each specific band of interest along the non-corrected EEG trials.

Pearson correlation, MI and FC were the measures employed to evaluate the methods specificity, i.e. how the tested methodologies were reliable in correctly removing the ocular blinks contributions from the EEG signal (IN-blink analysis). While the PSD preservation, computed in theta, alpha, and beta EEG frequency bands was selected as measure to evaluate the performance of each tested methodology in terms of sensibility, i.e. which one of the tested methods resulted the best in terms of neurophysiological content preservation of the signal (OUT-blink analysis).

Both the comparison analyses were performed by considering two principal EEG channels subsets: (i) the frontal area, which included all the EEG channels located in anterofrontal and frontal scalp positions, and (ii) the parietal area, which included all the EEG channels located within the parietal scalp positions. Each performance measure, initially computed for the single EEG channel, has been so averaged over each of the two subsets. We took into account these two-scalp area, in order to investigate the behaviour of each method in two extreme conditions, i.e. the frontal area, where the ocular blinks contribution is maximum, and parietal area, where such contribution is minimum.

The above-described analyses were performed on both the LAB and REAL settings EEG datasets. Concerning the statistical analysis, the Shapiro–Wilk test was used to assess the normality of the distribution related to each of the considered measures. If normality was confirmed, the analysis of variance (ANOVA) or, in the case of non-normal distribution, its non-parametric equivalent (Friedman ANOVA) was performed. In order to compare the single tested methodologies, post-hoc analyses were performed by applying the Tukey's approach if the group analyses resulted to be statistically significant. For all tests, statistical significance was set at $\alpha = 0.05$.

3. Results

This section was divided into different subparagraphs. In particular, a first detailed statistical analysis was performed to assess the efficiency of the tested methods in terms of ocular blink artifacts identification and correction (i.e. IN-blink analysis) and in terms of EEG signal preservation (i.e. OUT-blink analysis), and finally a topographical analysis was conducted for clearly resuming the main research outcomes.

3.1. EEG ocular blink detection and correction: IN-blink analysis

3.1.1. EEG dataset in controlled settings

Concerning the Pearson correlation (the lower the better) measure, which was computed between the data channel containing the ocular blink artifact contribution (i.e. the EOG channels for the dataset collected in controlled settings) and the raw (i.e. not corrected, 'Original') and corrected (by the different methods) EEG channels, the Friedman ANOVA revealed that the Pearson correlation statistically decreased after applying the ocular blink artifacts correction algorithms over the frontal EEG channels (Friedman chi-squared = 84.858, $p < 0.001$) (figure 2). Moreover, the post-hoc analysis demonstrated that the Gratton, AMICA, Sgeyesub, and the MWF implementation resulted to be the most effective ones in terms of Pearson correlation decrease, with respect to the 'Original' (i.e. not corrected) case, between the ocular blink components and the corrected EEG data (all $p_{\text{Tukey}} < 0.01$).

A similar pattern was observed when applying the tested methodologies over the parietal EEG channels (Friedman chi-squared = 29.104, $p < 0.001$). In this case, the post-hoc analysis revealed that the proposed o-CLEAN algorithm, the AMICA, and the MWF implementations were associated with the lower Pearson correlation between the ocular blink components and the corrected EEG data (all $p_{\text{Tukey}} < 0.007$) (figure 2).

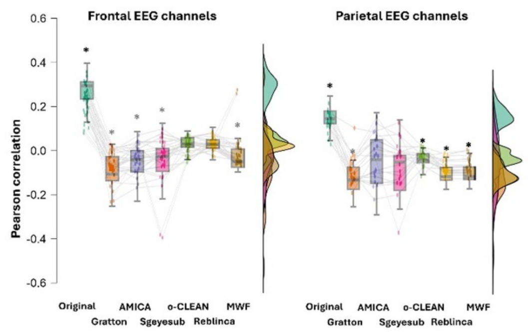


Figure 2. Concerning the frontal EEG channels, the Friedman ANOVA demonstrated that the Gratton, AMICA, Sgeyesub, and the MWF resulted to be the most effective ones in terms of Pearson correlation decrease, with respect to the 'Original' (i.e. not corrected) case, between the ocular blink components and the corrected EEG data (all $p_{\text{Tukey}} < 0.01$). While over the parietal EEG channels, the statistical analysis showed that o-CLEAN, the AMICA, and the MWF were associated with the lower Pearson correlation between the ocular blink components and the corrected EEG data (all $p_{\text{Tukey}} < 0.007$).

Similar statistical analyses were performed by considering the MI and FC efficiency measures (the lower the better). In this regard, the Friedman ANOVA showed a statistically significant decrease in terms of MI between the ocular blink components and the frontal EEG channels after applying all the tested methodologies (Friedman chi-squared = 25.173, $p < 0.001$). The post-hoc analysis showed the same statistical specificity among the tested methods in terms of MI decrease after the ocular blink artifact correction on the EEG dataset collected in controlled settings, except for the REBLINCA method, which resulted to be less specific ($p_{\text{Tukey}} = 0.03$) (figure 3). Similar patterns were observed when conducting the statistical analysis over the parietal EEG channels. The group analyses revealed a statistically significant specificity in correcting the ocular blink artifacts through the tested methodologies (Friedman chi-squared = 15.143, $p < 0.001$). The post-hoc analysis demonstrated that the proposed o-CLEAN algorithm resulted to be among the most efficient approaches in terms of MI decrease after the method application ($p_{\text{Tukey}} < 0.008$) (figure 3).

Regarding the FC measure (the lower the better), the Friedman ANOVA demonstrated a significant effect associated to the correction methods application over the frontal EEG channels (Friedman chi-squared = 29.104, $p < 0.001$). The post-hoc analyses demonstrated that the proposed o-CLEAN algorithm resulted to be the most specific in detecting and correcting the ocular blink artifacts, even when other validated methodologies, such as the Gratton and Sgeyesub, exhibited statistically lower specificity ($p_{\text{Tukey}} < 0.02$) (figure 4). Concerning the statistical analysis performed over the parietal EEG channels, the Friedman ANOVA revealed the same

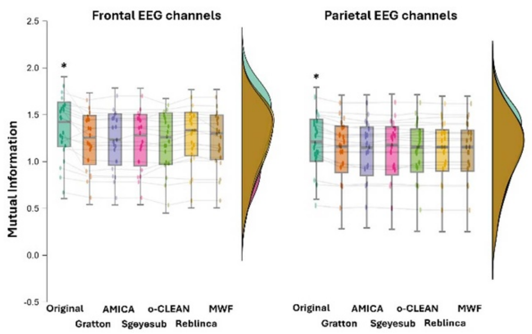


Figure 3. The Friedman ANOVA showed the same statistical specificity among the tested methods in terms of MI decrease after the ocular blink artifact correction on the EEG dataset collected in controlled settings, except for the REBLINCA method, which resulted to be less specific ($p_{\text{Tukey}} = 0.03$). Similar patterns were observed when conducting the statistical analysis over the parietal EEG channels.

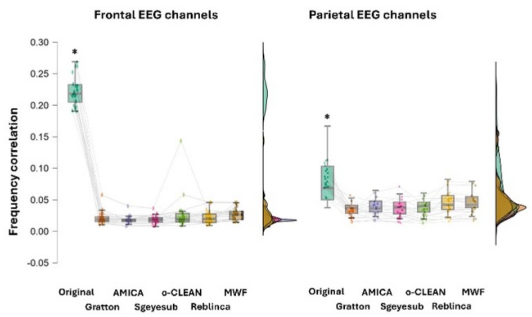


Figure 4. The Friedman ANOVA and the related post-hoc tests (i.e. Tukey) revealed that all the tested methods resulted to be statistically specific in correcting the ocular blinks artefacts in terms of frequency correlation. These outcomes were observed over both the frontal and the parietal EEG channels.

above-mentioned significant effects of the correction methods in terms of FC decrease (Friedman chi-squared = 19.271, $p < 0.001$). Similarly to the outcomes related to the analysis conducted over the frontal EEG channels, the post-hoc analysis showed that the proposed o-CLEAN method was among the most efficient methods when applied to the EEG dataset collected in controlled settings ($p_{\text{Tukey}} = 0.01$) figure 4.

The following table 1 resumes the presented results in terms of methods specificity in detecting and correcting the ocular blink artifacts over the frontal and parietal EEG channels associated to the dataset collected in controlled settings.

3.1.2. EEG dataset in naturalistic settings

The above-described statistical analysis was replicated by considering the EEG dataset collected in real settings. Starting from the measures to evaluate the specificity of the investigated methods, the Pearson correlation computed between the data channel containing the ocular blink artifact component (i.e. the Fpz EEG channel in this case) and the raw ('Original') and corrected (by the different methods) EEG channels, statistically decreased after applying the ocular blink artifacts identification and correction algorithms over the frontal EEG channels according to the Friedman ANOVA (Friedman chi-squared = 38.227, $p < 0.001$). Interestingly, the post-hoc analysis performed when considering the EEG dataset collected in real settings demonstrated that the proposed o-CLEAN approach and the regression-based method (i.e. REBLINCA) resulted to be the most specific in correcting the ocular blink artifacts (all $p_{\text{Tukey}} < 0.005$), while the same post-hoc analysis performed by considering the dataset collected in controlled settings did not underline any statistical differences between the considered methodologies in terms of specificity to the ocular blink artifact recognition and correction (figure 5).

Concerning the case in which the tested methodologies were applied over the parietal EEG channels, the statistical analysis revealed a significant effect in terms of ocular blink identification and correction (Friedman chi-squared = 27.289, $p < 0.001$). In this case, the post-hoc analysis did not reveal any statistical differences between the tested methods in terms of Pearson correlation decrease (figure 5).

Analogously, the statistical analyses were performed by considering the MI and FC efficiency measures. In this regard, the Friedman ANOVA showed a statistically significant MI decrease between the ocular blink components and the frontal EEG channels after the application of the tested methodologies (Friedman chi-squared = 20.985, $p < 0.001$). Additionally, the post-hoc analysis demonstrated that the proposed o-CLEAN method resulted to be the most specific in terms of MI decrease related to the ocular blink artifacts correction as the Gratton, AMICA, Sgeyesub, and the classical MWF implementation (all $p_{\text{Tukey}} < 0.01$) (figure 6). Similar behaviours were observed when conducting the statistical analysis over the parietal EEG channels. The analyses revealed a statistically significant specificity in correcting the ocular blink artifacts through the tested methodologies (Friedman chi-squared = 11.157, $p < 0.001$). The post-hoc analysis demonstrated that the proposed o-CLEAN algorithm resulted to be among the most specific approaches in terms of MI decrease after the method application ($p_{\text{Tukey}} = 0.004$) figure 6.

Regarding the FC measure, the Friedman ANOVA revealed a significant statistical effect associated to the correction methods application over the frontal EEG channels (Friedman chi-squared = 27.881, $p < 0.001$). As observed in the controlled settings analysis, the post-hoc tests showed that the proposed o-CLEAN algorithm resulted to be the most specific in detecting and correcting the ocular blink artifacts ($p_{\text{Tukey}} < 0.007$) (figure 7). Similar results were observed when assessing the methods specificity through the FC over the parietal EEG channels. The

Table 1. The resuming table representing the statistical results obtained in terms of methods specificity by performing the statistical analysis (i.e. Friedman ANOVA) considering the frontal and the parietal EEG channels separately. Each tested method is associated to the specificity metrics (i.e. Pearson correlation, MI, and FC) median values computed between the ocular blink components (i.e. the EOG channel) and the corrected EEG channels. The 'rank' rows represent the statistical ranking associated with the tested methods specificity. The table includes also the specificity metrics median value computed by considering the ocular blink components and the non-corrected EEG channels (i.e. Original).

IN-blink analysis		Pearson correlation		MI		FC	
Methods	Efficiency metrics	Fronetal	Frontal	Parietal	Frontal	Parital	Parietal
Original	Median	0.39	0.21	0.80	0.62	0.18	0.04
	Rank	NA	NA	NA	NA	NA	NA
Gratton	Median	0.02	−0.09	0.61	0.54	0.03	0.02
	Rank	1	2	1	1	1	1
AMICA	Median	0.05	0.01	0.61	0.52	0.02	0.02
	Rank	1	1	1	1	1	1
Sgeyesub	Median	0.04	−0.07	0.61	0.59	0.03	0.02
	Rank	1	2	1	1	1	1
o-CLEAN	Median	0.10	0.01	0.52	0.50	0.04	0.02
	Rank	2	1	1	1	1	1
REBLINCA	Median	0.12	−0.08	0.66	0.53	0.03	0.03
	Rank	3	2	2	1	1	2
MWF	Median	0.06	0.02	0.60	0.49	0.03	0.02
	Rank	1	1	1	1	1	1

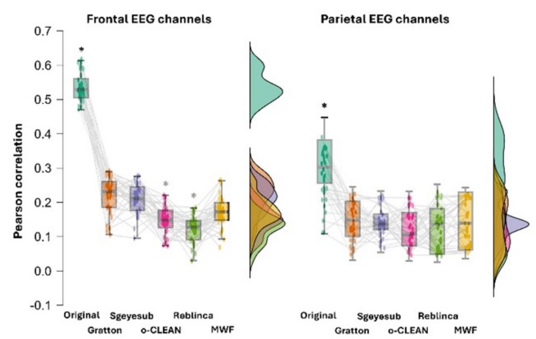


Figure 5. The Friedman ANOVA performed by considering the parameter computed over the frontal EEG channels, revealed that the proposed o-CLEAN approach and the regression-based method (i.e. REBLINCA) resulted to be the most specific in correcting the ocular blink artifacts (all $p_{\mathrm{Tukey}} < 0.005$).

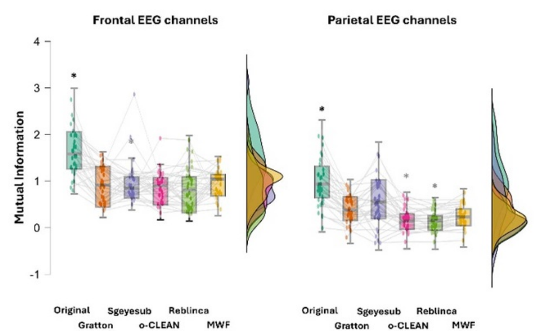


Figure 6. The Friedman ANOVA related to the methods performance in terms of ocular blinks artefact correction showed that the o-CLEAN methods resulted to be among the most specific ones, over both the frontal and the parietal EEG channels (all $p_{\mathrm{Tukey}} < 0.01$).

Friedman ANOVA showed a significant FC decrease after the methods application with the respect to the *Original* condition (Friedman chi-squared = 12.527, $p < 0.001$). While the post-hoc analysis showed that the o-CLEAN method was among the most specific methods in terms of ocular blink identification and correction (all $p_{\mathrm{Tukey}} < 0.008$) (figure 7).

The following table 2 resumes the presented results in terms of methods specificity in detecting and correcting the ocular blink artifacts when considering the EEG dataset collected in real settings.

In conclusion, the following table 3 represents the topographical analysis performed for visualizing the ocular blink artifacts detection and correction efficiency associated with each tested method.

The above topographical representations show how all the tested methodologies generally identified and corrected efficiently the ocular blink artifacts when considering the LAB EEG dataset. While the topographical analysis performed on the Real EEG dataset highlighted a slight, but still statistically significant (see table 3), higher performance of the proposed o-CLEAN method, the REBLINCA and the classical implementation of the MWF in terms of ocular blink artifacts correction. In fact, the topographical maps related to the three considered performance measures (i.e. the Pearson correlation, FC,

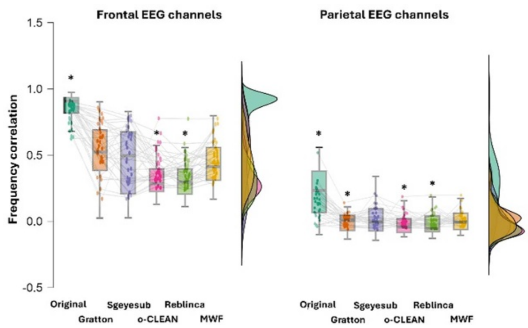


Figure 7. The post-hoc analyses (i.e. Tukey) demonstrated that the o-CLEAN method was the most specific in detecting and correcting the ocular blink artifacts ($p_{\text{Tukey}} < 0.007$). Concerning the parietal EEG channels, the Friedman ANOVA showed a significant FC decrease after the methods application with the respect to the Original condition (Friedman chi-squared = 12.527, $p < 0.001$). While the post-hoc analysis showed that the o-CLEAN method was among the most specific methods in terms of ocular blink identification and correction (all $p_{\text{Tukey}} < 0.008$).

Table 2. The resuming table representing the statistical results obtained in terms of methods specificity by performing the statistical analysis (i.e. Friedman ANOVA) considering the EEG dataset collected in real settings. Each tested method is associated to the specificity metrics (i.e. Pearson correlation, MI, and FC) median values computed between the ocular blink components and the corrected EEG channels. The 'rank' rows represent the statistical ranking associated with the tested methods specificity. The table includes also the specificity metrics median value computed by considering the ocular blink components and the non-corrected EEG channels (i.e. Original). It has to be noted that the AMICA algorithm was not tested over the EEG dataset collected in real settings because of the low EEG channels number (i.e. 5 and 3 frontal and parietal EEG channels, respectively).

IN-blink analysis		Pearson correlation		MI		FC	
Methods	Efficiency metrics	Frontal	Parietal	Frontal	Parietal	Frontal	Parietal
Original	Median Rank	0.54 NA	0.33 NA	0.93 NA	0.24 NA	0.86 NA	0.59 NA
Gratton	Median Rank	0.22 2	0.15 1	0.39 1	0.14 2	0.52 2	0.21 1
AMICA	Median Rank	Not tested NA	Not tested NA	Not tested NA	Not tested NA	Not tested NA	Not tested NA
Sgeyesub	Median Rank	0.21 2	0.14 1	0.44 2	0.16 3	0.46 2	0.25 2
o-CLEAN	Median Rank	0.12 1	0.11 1	0.35 1	0.06 1	0.31 1	0.18 1
REBLINCA	Median Rank	0.12 1	0.14 1	0.39 1	0.05 1	0.33 1	0.2 1
MWF	Median Rank	0.21 2	0.15 1	0.39 1	0.12 2	0.34 1	0.23 2

and the MI) show that the most prominent differences between the corrected and the non-corrected (i.e. *Original*) occurred when the above-mentioned methodologies were applied.

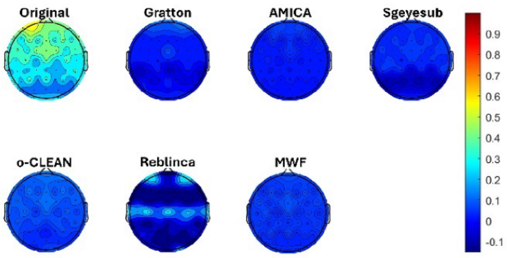
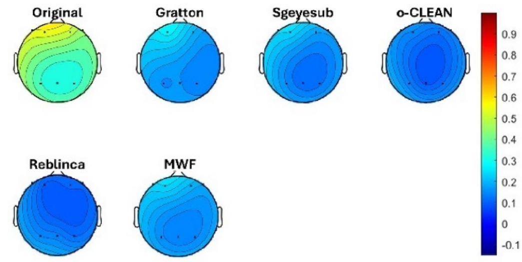
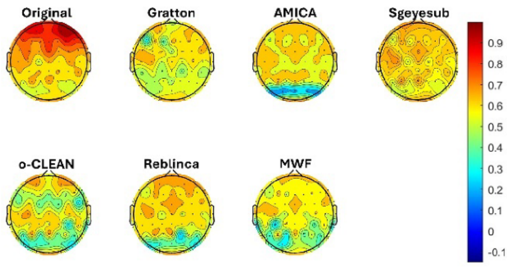
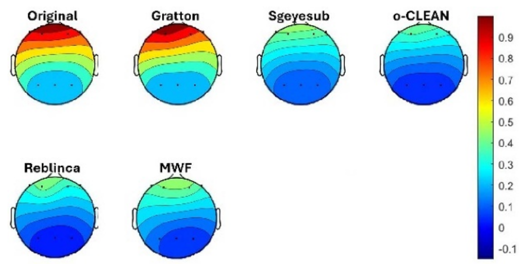
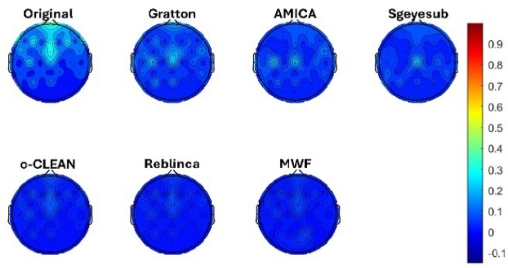
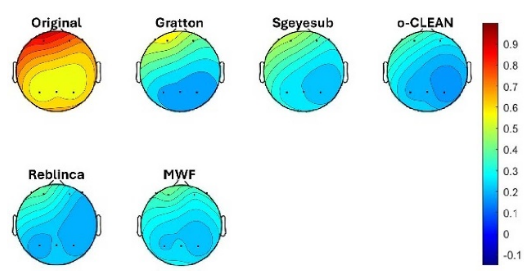
3.2. Performance in terms of EEG signal preservation: OUT-blink analysis

3.2.1. EEG dataset in controlled settings

To investigate the EEG signal preservation outside of the blinks (i.e. OUT-blink), for each of the employed methods it has been calculated the PSD preservation with the respect to the resting state condition (the highest the better), i.e., the EEG data collected along the experimental conditions in which no ocular blinks occurred, computed in theta, alpha, and beta EEG frequency bands. In this regard, the Friedman ANOVAs calculated for the o-CLEAN method, showed the highest PSD preservation in theta, alpha, and beta EEG frequency bands, when considering both the frontal (figure 8) and the parietal EEG channels (figure 9) (Frontal EEG channels: Friedman chi-squared = 8.791, $p < 0.001$; Parietal EEG channels: Friedman chi-squared = 6.183, $p = 0.02$).

The statistical results highlighted that the tested methodologies behaved statistically similar when the PSD preservation was computed in beta EEG frequency band, while the more prominent differences between methods were observed within the theta and alpha EEG frequency bands, i.e. the bands where the blinks spectral content is higher and therefore the risk

Table 3. Topographical maps representing the spatial median distribution of the computed efficiency metrics (i.e. the Pearson correlation, FC, and MI) related to the different tested methodologies for identifying and correcting the ocular blinks artifacts.

IN-blink analysis	LAB EEG dataset	REAL EEG dataset
Pearson correlation		
MI		
FC		

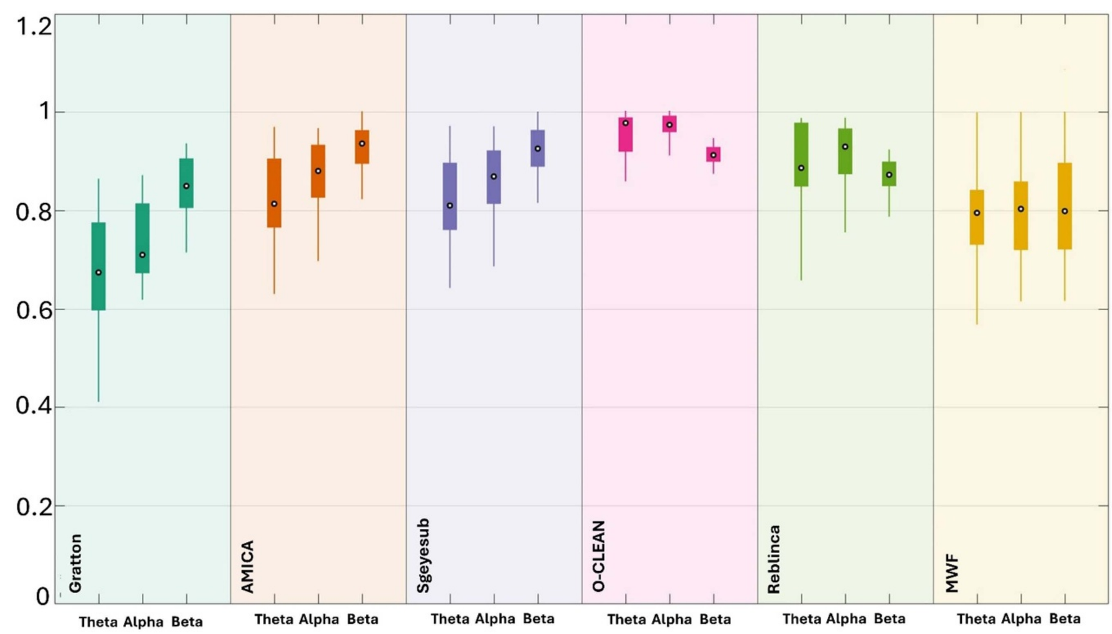


Figure 8. The statistical analysis (i.e. Friedman ANOVA) demonstrated that o-CLEAN method showed the highest PSD preservation in theta, alpha, and beta EEG frequency bands, when considering the frontal EEG channels, (Friedman chi-squared = 8.791, $p < 0.001$).

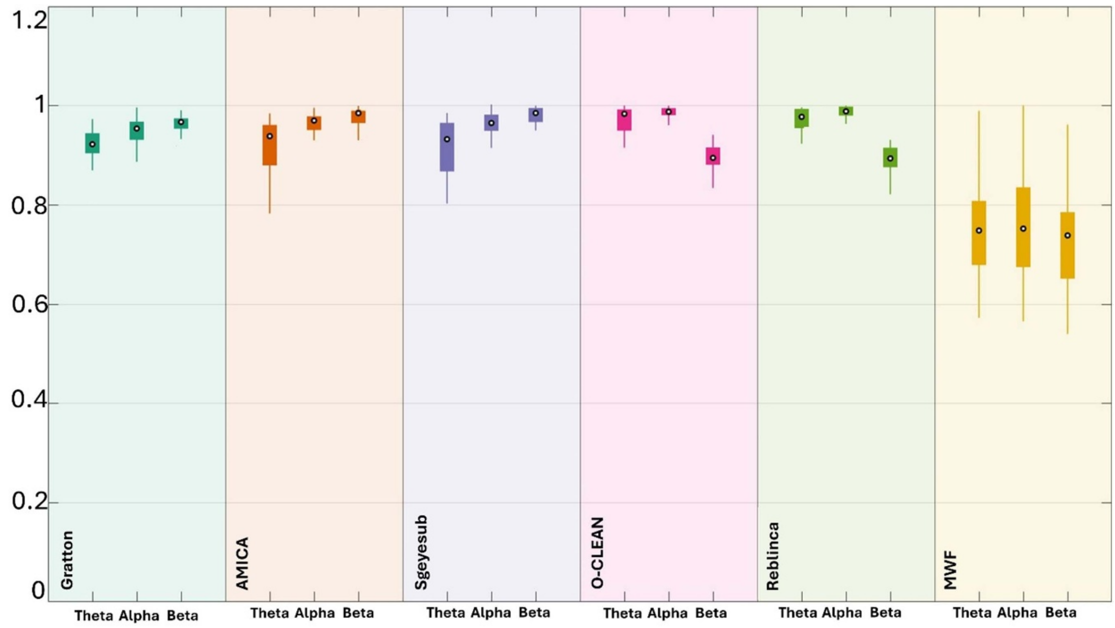


Figure 9. The statistical analysis (i.e. Friedman ANOVA) performed on the methods' performance in terms of EEG signal preservation over the parietal EEG channels highlighted similar results to the ones observed over the frontal EEG channels. The o-CLEAN methods resulted to be the most performing within all the considered EEG frequency bands (Friedman chi-squared = 6.183, p = 0.02).

of inducing distortion by the correction procedure is higher as well. In this regard, the post-hoc statistical analysis showed that the lowest PSD preservation, i.e. the highest EEG signal distortion associated to the correction method impact, was observed within the theta and alpha EEG frequency bands when applying the Gratton, AMICA, and Sgeyesub methodologies (all $p_{\mathrm{Tukey}} < 0.02$). Such PSD preservations were observed to be statistically lower when applying the above-mentioned methodologies on the Real EEG datasets ($p_{\mathrm{Tukey}} < 0.001$). On the contrary, the post-hoc analysis highlighted that the o-CLEAN method resulted to be the best among all the tested approaches in terms of EEG signal preservation within all the considered frequency bands (all $p_{\mathrm{Tukey}} < 0.01$).

The following table 4 represents a summary of the presented results in terms of methods performance in correcting the ocular blink artifacts from the frontal and parietal EEG channels.

3.2.2. EEG dataset in real settings

The same analysis, regarding the EEG signal preservation for the employed methods, was performed by considering the dataset collected in real settings. Similarly to the outcomes of the controlled settings, the Friedman ANOVAs showed how the proposed o-CLEAN method was characterized by the higher PSD preservation in theta, alpha, and beta EEG frequency bands, when applied to both over the frontal and the parietal EEG channels (Frontal EEG channels: Friedman chi-squared = 15.102, $p < 0.001$; Parietal EEG channels: Friedman chi-squared = 11.072, $p < 0.001$). In particular, the post-hoc statistical analysis showed that the lowest PSD preservation, i.e. the highest EEG signal distortion, was observed within the theta and alpha EEG frequency bands when applying the Gratton, MWF, and Sgeyesub methodologies (all $p_{\mathrm{Tukey}} < 0.02$). Similarly to what observed within the LAB dataset, the post-hoc analysis showed that the o-CLEAN method induced the highest EEG signal preservation within all the considered frequency bands over both the frontal and the parietal EEG channels (all $p_{\mathrm{Tukey}} < 0.005$) (figures 10 and 11).

The following table 5 reports the summary of the statistical analysis, by representing the statistical analysis performed on the performance metrics for each method, on the EEG dataset collected in real settings.

The topographical representation performed on the performance measures in terms of EEG signal preservation, i.e. the PSD preservation computed with the theta, alpha, and beta EEG frequency bands calculated between the corrected and the non-corrected EEG resting experimental trials, are represented in table 6.

The above-presented topographical representation showed how all the tested methodologies did not induce a consistent distortion to the EEG signal when they were applied to the LAB EEG dataset. In fact, it can be observed that the PSD preservations computed within the theta, alpha, and beta EEG frequency bands were above 0.7, and in most of the cases they were higher than 0.9. Similarly to the efficiency

Table 4. The resuming table representing the statistical results obtained in terms of methods performance by performing the statistical analysis considering the EEG dataset collected in controlled settings. Each tested method is associated to the performance metrics (i.e. PSD preservation computed within the theta, alpha, and beta EEG frequency bands) median values computed by considering the corrected and the non-corrected EEG signal collected during the resting state experimental conditions. The 'rank' rows represent the statistical ranking associated with the tested methods performance.

OUT-blink analysis		PSD preservation in theta band		PSD preservation in alpha band		PSD preservation in beta band	
Methods	Performance metrics	Frontal	Parietal	Frontal	Parietal	Frontal	Parietal
Gratton	Median	0.69	0.91	0.73	0.94	0.85	0.95
	Rank	4	2	3	1	1	1
AMICA	Median	0.81	0.91	0.85	0.95	0.92	0.99
	Rank	3	2	2	1	1	1
Sgeyesub	Median	0.79	0.91	0.84	0.95	0.91	0.99
	Rank	3	2	2	1	1	1
o-CLEAN	Median	0.94	0.95	0.96	0.96	0.91	0.89
	Rank	1	1	1	1	1	2
REBLINCA	Median	0.86	0.96	0.88	0.97	0.86	0.89
	Rank	2	2	2	1	2	2
MWF	Median	0.72	0.76	0.78	0.74	0.78	0.72
	Rank	3	3	2	2	2	3

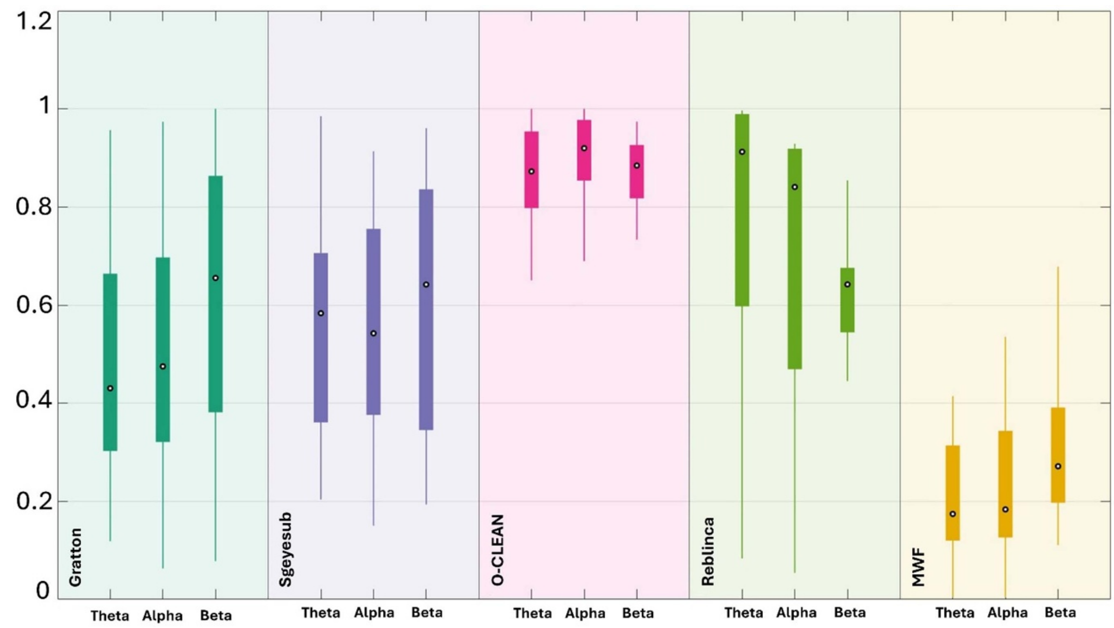


Figure 10. When considering the EEG dataset collected in real settings (i.e. REAL), the statistical analysis (i.e. Friedman ANOVA) highlighted how the o-CLEAN methodology resulted to be significantly more reliable in minimizing the signal distortion with the respect to the other tested methods over the frontal EEG channels (Friedman chi-squared = 15.112, $p < 0.001$; all $p_{\mathrm{Tukey}} < 0.005$).

assessment, the differences between the tested methodologies in terms of performance were highlighted on the Real EEG dataset. In fact, tables 3 and 6 demonstrate that the PSD preservations resulted to be lower than 0.55 for the Gratton, Sgeyesub, and MWF methods, especially when they were applied for correcting the ocular blinks artifacts over the frontal EEG channels. Concerning the proposed o-CLEAN approach, the topographical representations demonstrated that such a method was the one inducing the lower signal distortion all over the scalp.

Finally, the followings scatterplots represent the tested methodologies performance in terms of specificity (i.e. effective ocular artefacts correction) and sensibility (i.e. the EEG signal preservation maximization). Such scatterplots were built by assigning

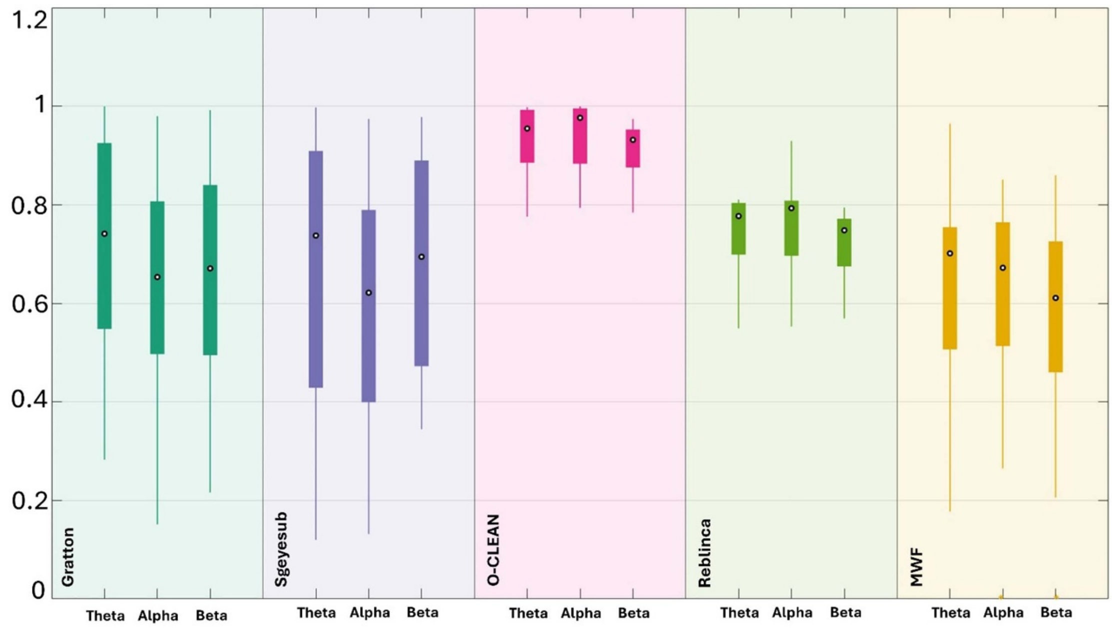


Figure 11. Similarly to the results observed when considering the frontal EEG channels, the o-CLEAN method exhibited the best performance in terms EEG signal preservation when applied to the parietal EEG channels (Friedman chi-squared = 11.072, $p < 0.001$; all $p_{\mathrm{Tukey}} < 0.005$).

Table 5. The resuming table representing the statistical results obtained in terms of methods performance by performing the statistical analysis (i.e. Friedman ANOVA) considering the EEG dataset collected in real settings. Each tested method is associated to the sensibility metrics (i.e. PSD preservation computed within the theta, alpha, and beta EEG frequency bands) median values computed by considering the corrected and the non-corrected EEG signal collected during the resting state experimental conditions. The 'rank' rows represent the statistical ranking associated with the tested methods performance. It has to be noted that the AMICA algorithm was not tested over the dataset collected in real settings because of the low EEG channels number (i.e. 5 frontal and 3 parietal EEG channels respectively).

OUT-blink analysis		PSD preservation in theta band		PSD preservation in alpha band		PSD preservation in beta band	
Methods	Performance metrics	Frontal	Parietal	Frontal	Parietal	Frontal	Parietal
Gratton	Median Rank	0.45 2	0.8 2	0.48 3	0.65 3	0.62 2	0.68 2
AMICA	Median Rank	Not tested NA	Not tested NA	Not tested NA	Not tested NA	Not tested NA	Not tested NA
Sgeyesub	Median Rank	0.55 2	0.67 2	0.51 3	0.58 3	0.61 2	0.69 2
o-CLEAN	Median Rank	0.81 1	0.9 1	0.86 1	0.91 1	0.83 1	0.86 1
Reblinca	Median Rank	0.72 1	0.71 2	0.66 2	0.76 2	0.6 2	0.71 2
MWF	Median Rank	0.42 2	0.61 2	0.38 3	0.59 3	0.38 3	0.54 3

one positive score (i.e. +1) to each method resulting among the best ones in terms of each of the three considered specificity metrics, i.e. the Pearson correlation, the MI, and the FC. The following table 7 contains such representations when considering the LAB dataset:

The same final scatterplots (table 8) were computed by considering the methods' performance in terms of specificity and sensibility when they were applied to the REAL dataset:

An additional performance analysis was performed exclusively for the o-CLEAN method. The objective of such analysis was to assess the performance of the method's first processing block, in order to evaluate its performance in identifying the ocular blinks along the EEG signal collected during the

Table 6. Topographical maps representing the spatial median distribution of the computed performance metrics in terms of EEG signal preservation (i.e. the PSD preservation computed within the theta, alpha, and beta EEG frequency bands by considering the corrected and the non-corrected EEG trials collected during the resting experimental trials) related to the different tested methodologies for identifying and correcting the ocular blinks artifacts.

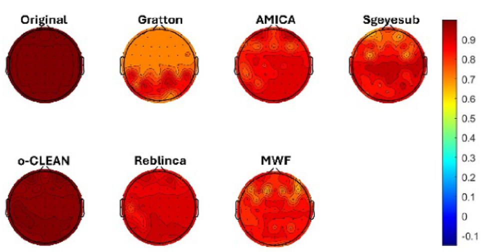
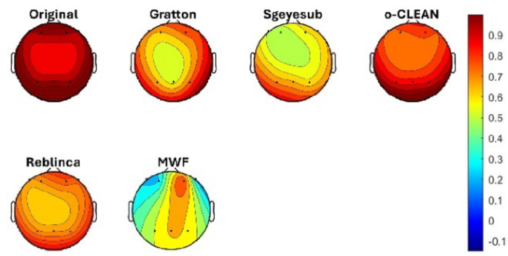
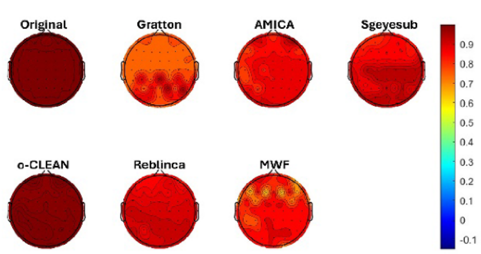
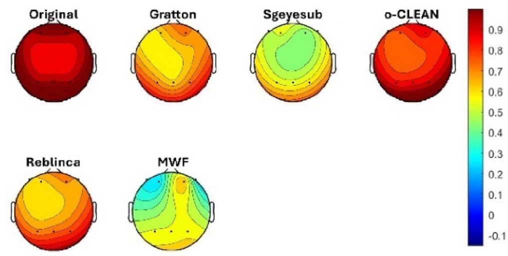
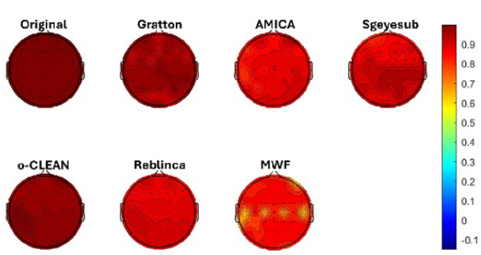
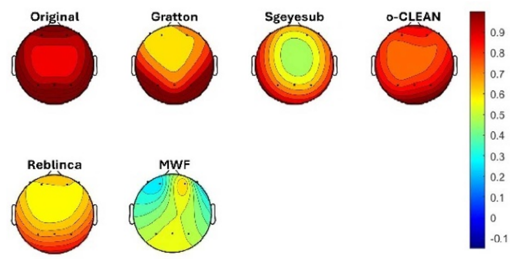
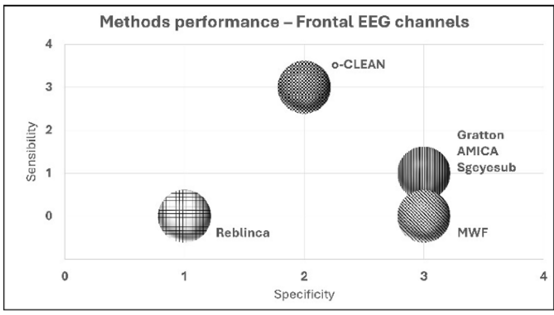
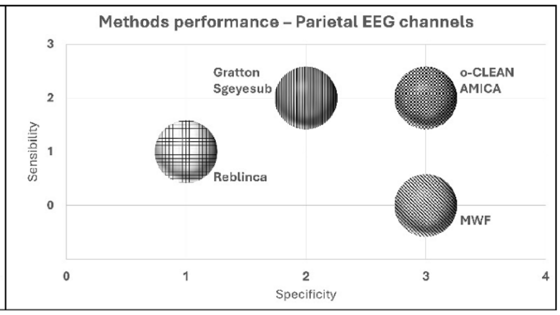
OUT-blink analysis	LAB EEG dataset	REAL EEG dataset
PSD preservation in theta band		
PSD preservation in alpha band		
PSD preservation in beta band		

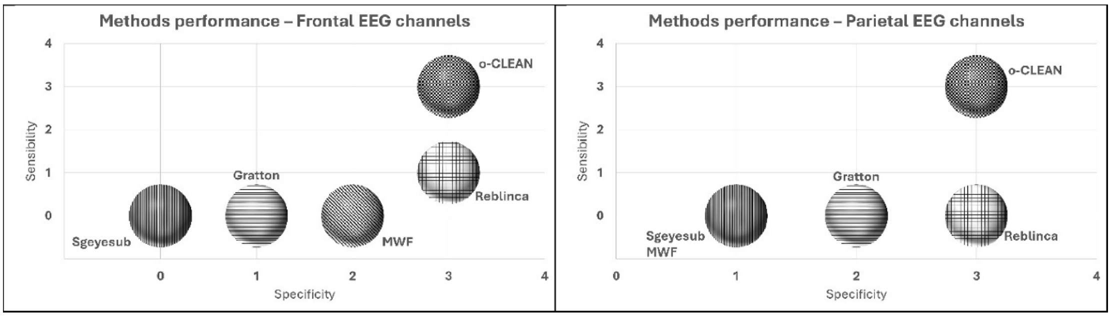
Table 7. Scatterplots representing the methods' performance in terms of specificity and sensibility when they were applied to the EEG dataset collected in controlled environment (i.e. the LAB dataset). Either over the frontal EEG channels and over the parietal ones, the o-CLEAN method resulted to be among the best performing.





calibration. Such a step is crucial, since it allows to compute the EOG template for each subject, which is functional to subsequently identify and correct the ocular blinks artifact along the test EEG data. Therefore, the sensibility and specificity of the first o-CLEAN processing block was computed by considering the identified ocular blinks and the "true" ocular blinks identified by visual inspection. Such a preliminary analysis was conducted on both the controlled and naturalistic (i.e. real driving) settings, and it resulted that the o-CLEAN method was highly sensible in identifying the ocular blinks artifacts along the calibration data. In fact, the observed sensibility corresponded to [98.39% ± 1.61%] and [96.74% ± 3.66%] for the controlled and the naturalistic EEG dataset respectively. Otherwise, the specificity of the method (Detected Blink-free EEG/True Blink-free EEG) was 100% in both the cases.

Table 8. The scatterplots representing the methods' performance when considering the EEG dataset collected in real settings (i.e. the REAL dataset) highlighted even more consistently, with the respect to the methods' performance evaluated when considering the LAB dataset, that the o-CLEAN method was the best performing.



4. Discussion

In the recent years, the EEG wearable industry has consistently revolutionized the fields of applied neuroscience, psychology, and in general of all those research areas where the focus is the investigation of brain activity related to the human behaviour. In fact, given the recent technological and methodological advancements, such solutions will soon be largely employed in many application fields (e.g. BCI), where brain activity has to be analysed online and in a seamless and robust way. This was made possible thanks to the relevant advancements in terms of hardware miniaturization and wearability improvements, which allowed to obtain EEG systems compatible with out-of-the-lab applications. In this regard, it has to be noted that the most recent works in scientific literature demonstrated how it is possible to compute specific and reliable EEG-based indicators by using a few EEG channels [23, 29, 32–36], which can be crucial in different operational application fields [6, 37, 38]. In this regard, the biggest challenge is generating robust signal processing techniques, able to maximize the signal to noise ratio (i.e. reducing, ideally deleting the contribution of artifacts). In particular, the ocular blink correction represents one of the biggest confounding factors that could arise during the EEG signal processing.

The presented study introduced a novel method for ocular artifacts correction, compliant with the constraints of real settings (i.e. online processing, with few EEG sensors, for out of the lab applications). The main objectives of the presented study consisted in assessing (i) the performance of the proposed methodology (i.e. the o-CLEAN) in removing the contribution of the ocular blinks-related EEG artifacts, and (ii) the performance of the proposed methodology in preserving the neurophysiological content of the EEG signal after applying the ocular blinks artifact correction method, i.e. by evaluating how the method impacted on the neurophysiological content of the corrected EEG signal, in comparison with the most widely used state of the art techniques. The analysis was conducted by computing three efficiency measures (i.e. the Pearson correlation, the MI, and the FC), in order to evaluate how the method specifically removed the ocular blinks components from the EEG signal, and the EEG PSD preservation measure within the theta, alpha, and beta EEG frequency bands, in order to investigate how the method impacted in terms of EEG signal distortion when applied on EEG trials not affected by the ocular blinks artifacts. Such metrics were computed on two EEG dataset, one in a controlled settings (i.e. LAB) and the other one in a real settings (REAL). The assessment of the proposed o-CLEAN method was performed by comparing it with the identified state-of-the-art related to the EEG processing. In particular, ICA, regression-based methods, and adaptive filtering techniques were selected. Additionally, the Sgeyesub method validated by Kobler and colleagues was identified to be compared with o-CLEAN, since it derives from the Artifact Subspace Reconstruction [39, 40]. In this regard, we underline that the implementation developed specifically for ocular artifact identification and correction was chosen because the primary objective of this study was to focus on correcting eyeblink artifacts in the EEG. Therefore, the discussion section was organized into two subparagraphs representing the two above-mentioned dimensions of the presented analysis.

4.1. Specificity assessment

The statistical analysis demonstrated that all the considered methodologies for the ocular blinks artifacts correction allowed to consistently reduce the ocular blinks contribution from the EEG signal. This was observed especially within the LAB EEG dataset, where the Pearson correlation between the ocular signal component and the EEG channels relevantly decreased to values below 0.1 after the correction methods applications (table 1), for both the frontal and parietal areas. Obviously, the main contribution of each method was observed when applied over the frontal EEG channels, since this was the area in which the ocular blinks signal content is demonstrated to be more prominent [41]. Similarly, the analysis conducted on the FC and MI metrics confirmed that all the tested methodologies allowed to specifically correct the ocular blinks artifacts contribution from the EEG signal. In this regard, few statistical differences among the different tested methods were highlighted by the presented analysis. Considering the Pearson correlation metric, the widely validated Gratton and AMICA approaches and the most recent Sgeyesub algorithm resulted to be the most effective. The proposed o-CLEAN and the MWF implementation (i.e. MWF) resulted to be slightly, but still statistically significantly, less effective in correcting the ocular blinks artifacts, while the regressive based method (i.e. the REBLINCA) were associated to a negative correlation between the EEG channels and the ocular blinks components. This could be linked to the fact that such methods could produce a negative deflection of the EEG signal in correspondence to an ocular blink peak. Such an aspect was visually observed along the performed data processing, especially over the frontal EEG channels.

A different, aspect emerged from the analysis on the REAL EEG dataset. Although the efficiency of all the methods for correcting the ocular artifacts contents from the EEG signal was still good, the statistical analysis highlighted that the proposed o-CLEAN and the REBLINCA methods were the most effective in both the scalp positions (i.e. frontal, and parietal EEG channels). While the Sgeyesub and the Gratton techniques were associated to statistically lower efficiency. In other words, the methods showed a different efficiency in removing the ocular blinks artifacts from the EEG signal, if applied to a less controlled setup (REAL EEG dataset). This seems in fact to be linked to the intrinsic nature of the collected data. The LAB dataset was in fact collected in laboratory settings through a well-defined and accurate experimental paradigm, in which the ocular artifact has been artificially induced. Each participant was in fact asked to artificially (i.e. voluntary) blink her/his eyes in specific moments, triggered by a recurrent cue in the task. On the contrary, the REAL EEG dataset was collected during a real driving task. In that context, blinks naturally (i.e. involuntary) occurred, so having a less reproducible template [42]. Natural blinks in fact, could change in intensity, frequency and duration, depending on many factors (e.g. attentional levels, task difficulty) [43]. This aspect could have played a relevant role in amplifying the differences between the proposed o-CLEAN and the other investigated approaches, with particular regard to the Gratton and Sgeyesub ones. The presented topographical maps in tables 3 and 6 confirmed such an aspect. It can be observed how, when considering the REAL EEG dataset, the FC and the MI did not consistently decrease over the frontal area (i.e. the one most affected by the ocular blinks signal components) when applying the Gratton and Sgeyesub techniques. On the contrary, the o-CLEAN resulted to be the only method exhibiting the higher efficiency in terms of Pearson correlation, MI and FC decrease after the algorithm application. In this regard, it has to be also observed that the methods performance assessment conducted by considering the REAL dataset highlighted one of the most relevant aspects of the o-CLEAN method, which is related to the calibration phase (i.e. the ocular blink artifact parameters computation). In fact, results achieved on the REAL dataset, suggest a big advantage of the proposed method with respect to the other ones. Unlike other methods that perform correction directly on the testing dataset, o-CLEAN involves a preliminary calibration on a controlled dataset, specifically one without artifacts other than ocular blinks. This approach makes the method more robust compared to other techniques, even when applied to a realistic dataset where the variability of EEG signals may be significantly higher than in laboratory settings.

4.2. Sensibility assessment

To investigate the preservation of the EEG signal outside the blinks, the EEG PSD preservation in theta, alpha, and beta EEG frequency bands has been calculated. This measure was computed by considering the corrected and the non-corrected EEG trials during the resting state, i.e. the experimental trials in which no ocular blinks artifact occurred. With respect to the previous analysis (i.e. blink correction), the preservation results showed that the o-CLEAN resulted to introduce less EEG signal distortion with respect to the other methods. In fact, the topographical maps reported in table 6, and the statistical summary reported in tables 4 and 5, confirmed that the o-CLEAN application on the EEG trials without blinks produced the lowest EEG PSD distortion (i.e. the highest EEG PSD preservation) among the three investigated EEG frequency bands. More specifically, the analysis revealed a similar behaviour of the methods within the beta EEG frequency band when considering the frontal area. The EEG PSD preservation in beta band over the parietal EEG channels resulted to be even higher for all methods. Such an outcome was indeed expected since the ocular blinks signal components are more prominent over the frontal area. By considering the EEG PSD preservation in beta band computed on the REAL EEG dataset, it can be observed how the difference between the methods was increased. In this regard, the o-CLEAN resulted to be the best method, since it exhibited a median signal preservation of 0.82 and 0.86 when applied over the frontal and parietal EEG channels respectively.

Concerning the theta EEG frequency band, the statistical analysis revealed that, even when considering the LAB EEG dataset, the Gratton approach generated a considerable EEG signal distortion, since the computed PSD preservation did not exceed 0.7. In this context, the o-CLEAN and REBLINCA resulted to be the most performing approaches, since they exhibited a signal preservation corresponding to 0.93 and 0.94 respectively over the frontal EEG channels and corresponding to 0.95 and 0.96 when applied to the parietal EEG channels. The negative impact of the Gratton, Sgeyesub, and the MWF methods on the signal distortion was most evident on the REAL EEG dataset. In fact, by considering the aforementioned methods, the EEG PSD preservation computed within the theta band over the frontal EEG channels did not exceed 0.55. These outcomes were expected, since the aforementioned methods were developed to be used only if an EOG signal (with no EEG coupled) was actually present. Anyhow, the EOG signal was not recorded within the REAL EEG dataset and, the Fpz EEG channel was considered for the ocular blinks artifact detection. In this regard, recording an additional EOG channel in out-of-the-lab contexts could be not feasible.

On the contrary, the o-CLEAN and the REBLINCA resulted to be the most performing in terms of EEG signal preservation. Such differences resulted to be less evident, but still statistically significant, when applying the selected methodologies over the parietal EEG channels on the REAL EEG dataset.

Finally, the EEG PSD preservation computed in alpha band when considering the LAB EEG dataset demonstrated that the o-CLEAN was among the most performing algorithms. In fact, the related signal preservation, in terms of PSD preservation measure, was 0.95 and 0.96 when considering the frontal and the parietal EEG channels respectively. More importantly, the o-CLEAN was the outperforming method, in preserving EEG signal in the REAL EEG dataset. In this regard, the PSD preservation was 0.83 and 0.91 when the method was applied to the frontal and the parietal EEG channels respectively.

4.3. o-CLEAN strength and robustness

Summing up the results, the o-CLEAN seems to provide superior artifact correction advantages compared to ICA, regression-based methods, and adaptive filtering techniques. Unlike ICA, which often requires a large number of channels and is in general not compliant with online analysis, o-CLEAN achieves robust performance with fewer channels and in real-time applications. This advantage makes o-CLEAN particularly valuable for wearable EEG systems used in mobile and out-of-lab applications. o-CLEAN's robust performance opens up new possibilities for various scenarios. In mobile EEG studies, where participants move freely in their environment, o-CLEAN ensures blinks free data collection, crucial for studying brain activity. In clinical settings, particularly for continuous monitoring of patients, o-CLEAN can enhance the detection of neural signals related to cognitive and emotional states, by a most effective blinks correction procedure. Additionally, in BCI applications, where real-time data processing is essential, o-CLEAN offers a practical solution for maintaining signal integrity for online analysis.

4.4. Limitations and possible improvements

Indeed, the presented results must not be envisioned as an ending point within the field of the EEG data preprocessing for blinks contribution removal, but they pave the way for furtherly optimize and transversally validate the proposed o-CLEAN method. The validation was conducted on two specific datasets, and future research should test o-CLEAN across a broader range of populations, tasks, and environments. This would also furtherly validate the empirical approach chosen for the ocular blinks identification performed by o-CLEAN on the calibration data. In fact, even if the o-CLEAN processing block which identifies the ocular blinks along the calibration was extensively and successfully employed for the EEG signal processing within several studies in simulated driving and aircraft management [29, 30, 32, 44], further demonstrating its effectiveness when applied to different EEG naturalistic datasets would positively contribute to its transversal validation. Moreover, as of now, the proposed algorithm relies on the ocular blinks' identification through a regressive-based procedure which uses a user-specific ocular blinks pattern. Future studies could investigate the development of a method that do not need a calibration phase (e.g. not user dependent, or unsupervised, i.e. calibrated on online data). Furthermore, future studies will be necessary for investigating the reliability of the proposed method in detecting and correcting other ocular EEG artifacts, i.e. the saccades, having a negative impact on the EEG signal neurophysiological interpretation, especially within out-of-the-lab applications.

5. Conclusions

This work presented a novel method for the EEG ocular artifacts correction, by using few EEG channels, compliant with out-of-the-labs application requirements, and with an online use.

The study was conducted by employing a controlled EEG dataset (i.e. LAB), with trials containing exclusively artificial (i.e. voluntary) ocular blink artifacts coupled with the EEG signal (i.e. Blink trials) and trials with non-contaminated EEG signal (i.e. Rest trials), and a real settings EEG dataset (REAL), in which the participants were involved in a driving tasks, and they naturally (involuntarily) blinked during the task execution. The analysis was performed by applying all the methods over the frontal and the parietal EEG channels, to assess their reliability in correcting the ocular blink artifacts and their impact on the neurophysiological content within areas which are more (i.e. the frontal one) or less (i.e. the parietal one) prone to ocular blink signal distortion.

The following results have been achieved:

- Firstly, the proposed o-CLEAN algorithm resulted to be, at least, significantly efficient as the most standard approaches identified in scientific literature, such as the Gratton, AMICA, and Sgeyesub algorithms, in terms of EEG ocular blink artifacts detection and correction.
- Secondly, the proposed o-CLEAN resulted to be the best algorithm in terms of EEG signal preservation.
- More importantly, such a result was observed for both the LAB and REAL EEG datasets, representing the fact that, even within an out-of-the-lab settings, the proposed o-CLEAN method was demonstrated to be reliable in detecting and correcting the ocular blinks-based EEG artifacts, and preserving the EEG signal outside of the blinks.

This aspect confirms the reliability of the o-CLEAN algorithm, even when the ocular blink signal components are characterized by non-regular patterns and amplitudes (natural blinks), as it happens in out-of-the-lab applications.

Besides these potential further improvements mentioned in the Discussion section, the present study clearly demonstrated that the proposed o-CLEAN represents a reliable solution for the ocular blink artifact identification and correction on the EEG signal, especially for neurophysiological data collected within real and/or realistic environments through wearable EEG systems, a field that is consistently growing in the wide range of the applied neuroscience.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—Project Title FIT2WORK—CUP B53D23024030001- Grant Assignment Decree No. 1383 adopted on 01/09/2023 by the Italian Ministry of Ministry of University and Research (MUR); the 'GURU' project, within the program BRIC, funded by the National Institute for Insurance against Accidents at Work, CUP B53C22008970005; the project GR-2019-12369824, funded by by the Italian Ministry of Health (MoH) under the Programme—Giovani Ricercatori 2019; the 'CODA' project funded by the HORIZON—SESAR EU (GA: 101114765), 'TRUSTY' project provided by the HORIZON—SESAR EU (GA: 101114838), and 'FITDRIVE: Monitoring devices for overall FITness of Drivers' (GA n. 953432). The individual grants, 'Fit2Fly' provided by 'Sapienza-Rome Technopole per Attrazione di early-career researchers MSCA Fellowships 2023' CUP B83C22002820006 to Vincenzo Ronca.

Ethics statement

The study in which the naturalistic EEG dataset was collected was approved by the University of Rome 'RomaTre'. The study was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. This human study was approved by University of Rome 'RomaTre' Ethics Committee—approval: 2023/02. All adult participants provided written informed consent to participate in this study for the publication of any potentially identifiable images or data included in this article.

ORCID iDs

Vincenzo Ronca https://orcid.org/0000-0002-7174-6331
Gianluca Di Flumeri https://orcid.org/0000-0003-4426-051X
Rossella Capotorto https://orcid.org/0009-0005-0842-2915
Nicolina Sciaraffa https://orcid.org/0000-0002-3380-0181
Gianluca Borghini https://orcid.org/0000-0001-8560-5671

References

[1] Arico P, Borghini G, Di Flumeri G, Sciaraffa N and Babiloni F 2018 Passive BCI beyond the lab: current trends and future directions *Physiol. Meas.* **39** 08TR02
[2] Aricó P, Borghini G, Di Flumeri G, Sciaraffa N, Colosimo A and Babiloni F 2017 Passive BCI in operational environments: insights, recent advances, and future trends *IEEE Trans. Biomed. Eng.* **64** 1431–6
[3] Borghini G, Astolfi L, Vecchiato G, Mattia D and Babiloni F 2014 Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness *Neurosci. Biobehav. Rev.* **44** 58–75
[4] Young M S, Brookhuis K A, Wickens C D and Hancock P A 2015 State of science: mental workload in ergonomics *Ergonomics* **58** 1–17
[5] Charles R L and Nixon J 2019 Measuring mental workload using physiological measures: a systematic review *Appl. Ergon.* **74** 221–32
[6] Borghini G *et al* 2020 A multimodal and signals fusion approach for assessing the impact of stressful events on air traffic controllers *Sci. Rep.* **10** 1–18

[7] Al-shargie F *et al* 2016 Mental stress assessment using simultaneous measurement of EEG and fNIRS *Biomed. Opt. Express* **7** 3882–98

[8] Choi Y, Kim M and Chun C 2015 Measurement of occupants' stress based on electroencephalograms (EEG) in twelve combined environments *Build. Environ.* **88** 65–72

[9] Islam M R *et al* 2020 A novel mutual information based feature set for drivers' mental workload evaluation using machine learning *Brain Sci.* **10** 551

[10] Arico P *et al* 2014 Towards a multimodal bioelectrical framework for the online mental workload evaluation *2014 36th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBC 2014* pp 3001–4

[11] Kleifges K, Bigdely-Shamlo N, Kerick S E and Robbins K A 2017 BLINKER: automated extraction of ocular indices from EEG enabling large-scale analysis *Front. Neurosci.* **11** 231855

[12] Wang L and Liu S Q 2011 Neural circuit and its functional roles in cerebellar cortex *Neurosci. Bull.* **27** 173–84

[13] Gawne T J, Killen J F, Tracy J M and Lahti A C 2017 The effect of saccadic eye movements on the sensor-level magnetoencephalogram *Clin. Neurophysiol.* **128** 397–407

[14] Keren A S, Yuval-Greenberg S and Deouell L Y 2010 Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression *Neuroimage* **49** 2248–63

[15] Urigüen J A and Garcia-Zapirain B 2015 EEG artifact removal—state-of-the-art and guidelines *J. Neural Eng.* **12** 031001

[16] Artoni F, Delorme A and Makeig S 2018 Applying dimension reduction to EEG data by principal component analysis reduces the quality of its subsequent independent component decomposition *Neuroimage* **175** 176–87

[17] Gratton G, Coles M G H and Donchin E 1983 A new method for off-line removal of ocular artifact *Electroencephalogr. Clin. Neurophysiol.* **55** 468–84

[18] Di Flumeri G, Arico P, Borghini G, Colosimo A and Babiloni F 2016 A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* (Institute of Electrical and Electronics Engineers Inc) pp 3187–90

[19] Kobler R J, Sburlea A I, Lopes-Dias C, Schwarz A, Hirata M and Müller-Putz G R 2020 Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals *Neuroimage* **218** 117000

[20] Parra L C, Spence C D, Gerson A D and Sajda P 2005 Recipes for the linear analysis of EEG *Neuroimage* **28** 326–41

[21] Kobler R, Sburlea A I and Müller-Putz G 2017 A comparison of ocular artifact removal methods for block design based electroencephalography experiments *Proc. 7th Graz Brain-Computer Interface Conf.* (https://doi.org/10.3217/978-3-85125-533-1-44)

[22] Somers B, Francart T and Bertrand A 2018 A generic EEG artifact removal algorithm based on the multi-channel Wiener filter *J. Neural Eng.* **15** 036007

[23] Giorgi A *et al* 2023 Neurophysiological mental fatigue assessment for developing user-centered artificial intelligence as a solution for autonomous driving *Front. Neurorobot.* **17** 1240933

[24] Romero S, Mañanas M A and Barbanoj M J 2008 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** 348–60

[25] Gratton G 1998 Dealing with artifacts: the EOG contamination of the event-related brain potential *Behav. Res. Methods Instrum. Comput.* **30** 44–53

[26] Palmer J, Kreutz-Delgado K and Makeig S 2011 AMICA: an adaptive mixture of independent component analyzers with shared components

[27] Comon P 1994 Independent component analysis, a new concept? *Signal Process.* **36** 287–314

[28] Roy R N, Charbonnier S and Bonnet S 2014 Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms *Biomed. Signal Process. Control* **14** 256–64

[29] Sciaraffa N *et al* 2022 Validation of a light EEG-based measure for real-time stress monitoring during realistic driving *Brain Sci.* **12** 304

[30] Sciaraffa N *et al* 2022 Evaluation of a new lightweight EEG technology for translational applications of passive brain-computer interfaces *Front. Hum. Neurosci.* **16** 458

[31] Lachaux J P, Chavez M and Lutz A 2003 A simple measure of correlation across time, frequency and space between continuous brain signals *J. Neurosci. Methods* **123** 175–88

[32] Ronca V *et al* 2024 A novel EEG-based assessment of distraction in simulated driving under different road and traffic conditions *Brain Sci.* **14** 193

[33] Di Flumeri G *et al* 2023 A neuroergonomic approach fostered by wearable EEG for the multimodal assessment of drivers trainees *Sensors* **23** 8389

[34] Vozzi A *et al* 2023 Time-dependent analysis of human neurophysiological activities during an ecological olfactory experience *Brain Sci.* **13** e.1242

[35] Vozzi A *et al* 2021 The sample size matters: to what extent the participant reduction affects the outcomes of a neuroscientific research. A case-study in neuromarketing field *Sensors* **21** 6088

[36] Di Flumeri G *et al* 2017 EEG-based approach-withdrawal index for the pleasantness evaluation during taste experience in realistic settings *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* pp 3228–31

[37] Zeng H *et al* 2021 An EEG-based transfer learning method for cross-subject fatigue mental state prediction *Sensors* **21** 2369

[38] Borghini G *et al* 2020 Stress assessment by combining neurophysiological signals and radio communications of air traffic controllers *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* (Institute of Electrical and Electronics Engineers Inc) pp 851–4

[39] Kumaravel V P, Kartsch V, Benatti S, Vallortigara G, Farella E and Buiatti M 2021 Efficient artifact removal from low-density wearable EEG using artifacts subspace reconstruction *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* pp 333–6

[40] Mullen T R *et al* 2015 Real-time neuroimaging and cognitive monitoring using wearable dry EEG *IEEE Trans. Biomed. Eng.* **62** 2553–67

[41] Croft R J and Barry R J 2000 Removal of ocular artifact from the EEG: a review *Neurophysiol. Clin./Clin. Neurophysiol.* **30** 5–19

[42] Tran D-K, Nguyen T-H and Nguyen T-N 2021 Detection of EEG-based eye-blinks using a thresholding algorithm *Eur. J. Eng. Technol. Res.* **6** 6–12

[43] Goldstein R, Bauer L O and Stern J A 1992 Effect of task difficulty and interstimulus interval on blink parameters *Int. J. Psychophysiol.* **13** 111–7

[44] Di Flumeri G *et al* 2022 EEG-based index for timely detecting user's drowsiness occurrence in automotive applications *Front. Hum. Neurosci.* **16** 866118