Semi-synthetic EEG Data for the Evaluation of Linear EEG Cleaning Methods

Wadda du Toit[1], Martin Venter[1], David Vandenheever[2,*]

[1]*Department of Mechanical and Mechatronic Engineering*
Stellenbosch University
Stellenbosch Central, Stellenbosch, South Africa
E-mails: waddabenjamin@gmail.com, mpventer@sun.ac.za

[2]*Department of Agricultural and Biological Engineering*
Mississippi State University
130 Creelman Street, Starkville, Mississippi, USA
E-mail: davidvdh@abe.msstate.edu

[*]*Corresponding author*

Received: October 26, 2022 **Accepted: March 09, 2023**

Published: December 31, 2023

Abstract: *Electroencephalography (EEG) data recordings can be contaminated by artefacts that reduce the quality and make analysis difficult, and therefore cleaning methods are essential for accurate analysis of EEG data. It is not yet well established how to measure performance based on measured contaminated data since there is no established benchmark for comparison. Here we use "clean" EEG data synthetically contaminated by electrocardiography (ECG), electrooculography (EOG) and electromyography (EMG). This introduces fewer assumptions to the comparison between various cleaning methods, providing a clear datum for comparison. Further contamination is controlled, adding artefacts individually and also as a combination of artefacts. The results show that signal to noise ratio (SNR) of the simulated artefacts was within the same ranges as found with measured artefacts from literature. Popular linear cleaning methods were evaluated on the dataset, showing similar results to those in the literature, further validating the usefulness and accuracy of the semi-synthetic dataset. The semi-synthetic dataset showed comparable characteristics to real measured EEG data and proved useful in the assessment of EEG cleaning methods. The cleaning methods showed varied results when performance was evaluated on individual artefacts.*

Keywords: *EEG, EOG, EMG, ECG, Artefacts, Simulation*.

Introduction

Electroencephalography (EEG) plays an essential role in identifying brain activity and behaviour. However, EEG signals are notoriously weak and are easily contaminated by artefacts. In this context, artefacts are activities that do not originate directly from the brain but are still present in the measured EEG data [72]. Some artefacts can imitate cognitive or pathological activity and become a significant problem. These imitated pathologies result in misleading visual interpretations and misdiagnosis of diseases such as sleep disorders and Alzheimer's disease [41, 58].

Artefacts complicate, distort and obscure the measured electrical activity originating from the brain [9, 41, 43]. Artefacts originate from various sources, which can significantly and detrimentally affect EEG due to large variations in temporal and spectral contamination [10, 20, 40, 73]. Some artefacts may contaminate several neighbouring channels, while others contaminate only a single channel. In addition, some artefacts appear as regular periodic events, while others are extremely irregular [40].

One can categorize artefacts into non-physiological and physiological artefacts. Physiological artefacts primarily include electrooculography (EOG), electromyography (EMG) and electrocardiography (ECG) which can cause severe problems for EEG analysis [10, 20, 73]. Non-physiological artefacts in EEG include instrumental and interference artefacts, such as line noise, magnetic fields, electrode movement, and poor electrical ground [17, 40, 58]. Non-physiological artefacts can often be removed using filtering techniques or wavelet-based methods [58, 77]. On the other hand, physiological artefacts pose a more difficult problem for EEG artefact removal compared to non-physiological artefacts, as they typically share the same frequency range as the EEG signals of interest [10, 20, 74]. By focusing our scope on physiological artefacts, we can evaluate the effectiveness of different EEG cleaning methods in handling the most challenging and critical aspect of EEG artefact removal [10, 20, 74].

EOG artefacts mainly originate from eye movements and blinks. Eye blinks have a large inter-subject variability, with naturally occurring eye blinks having smaller amplitudes and shorter duration than forced blinks [4, 30, 31, 47, 50, 83]. Saccade artefacts originate from changes in orientation of the retina and cornea dipole [30, 31, 40, 77]. Saccades and eye blinks both exhibit particular frequency characteristics but differ from each other significantly [30, 31, 53]. Saccades usually display a lower average voltage and lower range in voltage than eye blinks [14].

Furthermore, saccades show a similar average frequency but a higher frequency range than eye blinks [14]. Vertical saccades influence midline electrodes more, while lateral saccades influence lateral electrodes more [30]. Thus, EOG data is a combination of eye blinks and saccades [50, 53]. EOG artefacts are often removed using a reference channel and regression methods. A limitation of these methods is that the EOG reference data can also be cross contaminated by the EEG data, causing a possible removal of valuable cerebral activity from the data by these methods [41].

EMG artefacts originate from any muscle movements. Contamination of EEG data by muscle activity is a well-recognized and complex problem arising from different muscle groups [28, 63]. Any muscle contraction or stretch near an electrode recording site can result in EMG artefacts affecting the “clean” EEG signal [77]. The degree of muscle contraction and stretch affects the amplitude and waveform of EMG artefacts. Regression methods cannot be applied to EMG data as with EOG data because they originate from multiple sources [41]. EMG sources include the movement of many muscles, including muscle groups from the neck and face, such as the cheeks, forehead, jaws and tongue, from head movement, chewing, swallowing, clenching, talking, sniffing, and facial contractions [4, 17, 40, 58]. EMG presents a wide spectral distribution contaminating all the standard frequency bands. It is, however, most significant in the higher frequency bands, with most literature assuming that the EMG artefacts only affect the higher frequencies, starting at 15 to 20 Hz and upwards [4, 17, 40, 58]. The amplitude of EMG data has a peak in the 20 to 30 Hz range in the frontalis location [28]. The time series of EMG signals follow a spontaneous bursting behaviour with a temporal and spectral distribution similar to Gaussian noise [8, 51, 55]. Additionally, EMG and EEG signals have substantial statistical independence both temporally and spatially. This implies that independent component analysis (ICA) methods could effectively identify and remove EMG artefacts [40].

ECG artefacts originate in the heart, occurring in EEG data as a pulse or heartbeat when an electrode is placed on or near a pulsating blood vessel such as a scalp artery [17]. ECG signals display a simple, characteristic, and periodic time-frequency characterization pattern [42, 70, 77]. The amplitude of the ECG artefacts is relatively low compared to the amplitudes of EOG and EMG artefacts. However, the amplitude of the ECG artefacts also greatly depends on the electrode's relative position to the blood vessel and the anatomy of the participant [21, 77].

To mitigate the complexity of contaminated EEG, methods such as current source density (CSD) and blind source separation (BSS) have been developed [13, 77, 80, 82]. CSD is a method that aims to enhance the spatial resolution of EEG data and to remove the effects of volume conduction, which can be a confounding factor in EEG analysis [13, 61]. It achieves this by estimating the current density of the underlying neural sources from the recorded EEG signals. BSS, on the other hand, is a signal processing technique used to separate different sources of EEG signals, making it a useful tool for identifying and removing artefacts from the EEG data [77, 80]. Therefore, the combination of CSD and BSS methods can improve the quality of EEG data analysis by enhancing the spatial resolution of EEG data and reducing the effects of artefacts [61, 71].

Methods for removing artefacts are primarily developed and tested for removing only one type of artefact at a time [4, 6, 17, 20, 37, 77]. However, artefacts are plentiful, diverse and contaminate EEG data with a large variety of intensities, types, locations, combinations, and durations. Furthermore, participant variability is another factor to consider when removing artefacts [40]. EEG data processing that generalizes to multiple types of artefacts remains a significant challenge [27, 54, 69, 73, 84]. Therefore, artefact removal methods must handle a large variance in artefacts and EEG characteristics [40].

In addition to the high variability and complexity of artefacts, artefact removal methods are also limited by inherent constraints [41, 81]. A primary limitation is the removal of some of the real EEG signal alongside the artefacts, as artefacts may not be completely separable from the neural activity [41, 81]. Artefact removal techniques, such as ICA, use linear decomposition to extract artefacts from the neural activity [44, 77]. However, these techniques assume linear independence between the artefacts and neural activity, which may not always hold true [24]. Consequently, valuable cerebral activity may also be removed alongside the artefacts when they are not completely separable [24, 41, 49, 81]. Furthermore, artefact removal methods may add noise or distortion to the EEG signal, reducing the accuracy of the cerebral activity estimates [49]. Since the noiseless EEG signal is not known, the loss of valuable cerebral activity cannot be fully quantified and may have significant repercussions, particularly in clinical and research settings where accurate measurement and interpretation of neural activity is critical [7].

Evaluating the performance of artefact removal methods presents a significant challenge as the noiseless signal is not known a priori [73, 77]. A primary advantage of simulated EEG data is that the quality of the signal can be evaluated before and after artefact removal using standard evaluation measurements such as the signal to noise ratio (SNR). Simulations have historically played a significant role in developing cleaning methods. They can be generated using techniques ranging from very simple to more complex [5, 77]. Simulated contaminated EEG data enables the use of SNR, which compares the energy of the frequency domain of the "clean" EEG data to that of the artefacts [77]. It is possible to simulate some characteristics of recorded EEG data relatively accurately. However, characteristics such as synchronization between channels, volume conduction, combination of different artefacts and the effect of artefacts on physiological sources are more challenging to simulate. Volume conduction in EEG refers to the way in which the electrical potentials generated by neural activity spread throughout the conductive medium of the head and scalp, causing a mixing of signals between neighbouring channels [5, 34, 48]. When simulating artefacts, it is important to consider the effect of volume conduction on the resulting signal and to account for the spatial distribution of the artefact to accurately reflect the true nature of the contamination [35]. Furthermore, when simulating artefacts, the combination of EEG and physiological artefacts can be assumed to be linear or non-linear [41, 67, 77].

The linear assumption regarding physiological artefacts such as EMG, EOG, and ECG assumes that the contributions of each signal to the overall signal are proportional to their respective coefficients, and that they are independent from each other. The linear combination can be represented by Eq. (1) [30, 41, 77, 79]:

$$y(t) = a_1 x_1(t) + a_2 x_2(t) + a_3 x_3(t) + a_4 x_4(t), \tag{1}$$

where $y(t)$ is the recorded signal at time t, $x_1(t)$ is the EEG signal, $x_2(t)$ is the EMG signal, $x_3(t)$ is the ECG signal, $x_4(t)$ is the EOG signal, and a_1, a_2, a_3, and a_4 are the coefficients that determine the contribution of each signal to the overall signal.

The assumption of linearity is often challenged by non-linear interactions between the EEG signal and contaminating signals. For example, EMG activity can cause non-linear changes in scalp potentials, which makes it difficult to separate EMG from EEG signals using linear methods alone. In practice, the quality of recorded signals may also limit this assumption, as noise, drift, and other artefacts can affect the linearity of the mixing process. Linear-based artifact removal techniques may not be sufficient, as they do not account for the correlation between non-physiological and physiological artifacts, thus limiting their applicability [28, 30, 41, 53, 77].

Assuming that artefacts such as EMG, EOG, and ECG are added to each other and EEG non-linearly means that their contributions to the recorded signal are not proportional to their respective amplitudes and their amplitudes can vary over time and interact with each other and the EEG signal in a non-linear fashion. The non-linear combination is represented by Eq. (2) [1, 23, 34]:

$$y(t) = f(x_1(t), x_2(t), x_3(t), x_4(t)), \tag{2}$$

where $y(t)$ is the recorded signal at time t, $x_1(t)$ is the EEG signal, $x_2(t)$ is the EMG signal, $x_3(t)$ is the ECG signal, $x_4(t)$ is the EOG signal, and f is a non-linear function that maps the input signals to the output signal.

To estimate the non-linear function accurately, advanced optimization and statistical techniques are required due to the complexity of non-linear interactions. More complex models, such as feedforward models, simulate scalp EEG data as generated by dipolar sources by solving the electromagnetic forward problem using 3D models of the brain, skull and scalp. These models face many challenges, such as the sufficient modelling of source mixing caused by head tissue conductivity and the correlation of noise sources [1, 23, 34].

There is a need to develop tools that allow objective measurement and comparison of the performance of new and current EEG cleaning algorithms. In this quest, the current study develops a realistic semi-synthetic contaminated EEG dataset to evaluate and aid in the development of adequate cleaning methods. These cleaning methods will be benchmarked according to their ability to identify and remove individual artefacts and the combination of several artefacts from EEG data. To evaluate the performance of the cleaning methods, an original physiological artefact signal was linearly and realistically modified and added to the EEG data. As the focus of the research question is on assessing the effectiveness of the cleaning methods in removing the artefacts from the EEG data, rather than understanding the underlying neural processes, the CSD method is not directly relevant to the study and will be excluded from the scope.

The scope of cleaning methods investigated was limited to linear approaches, and more specific, BSS techniques, specializing in biomedical signals and removing physiological artefacts [18, 77]. Currently, BSS methods are the most popular category for artefact removal in EEG research [26, 32, 41, 77]. Under the BSS banner of methods, ICA and canonical correlation analysis (CCA) are some of the most popular methods [77]. The two most popular ICA based methods used in research are Extended Infomax and SOBI [53, 77]. CCA is a classic BSS method and is frequently used in brain-computer interface (BCI) research, which is associated with commercial products and, therefore, with a less controlled participant who has more artefact-inducing behaviours [10, 68, 77].

To properly investigate BSS methods, it is important to mention the assumptions regarding the signals that are being targeted for extraction. Linear approaches, such as SOBI, InfoMax, and CCA, assume that the EEG, EMG, ECG, and EOG signals are mixed as in Eq. (1). Additionally, these methods rely on the independent sources being statistically independent, non-Gaussian, and stationary over the duration of the recording. With these assumptions, linear techniques can successfully separate the independent sources and eliminate physiological artefacts [77, 80].

Materials and methods

Simulation data

The "clean" EEG data in this study was downloaded from Klados and Bamidis [48]. This dataset, which is made available without restriction for research purposes, contains pre-contaminated EEG data [48]. The EEG data were recorded from 27 healthy subjects (14 males with mean age of 28.2 years and 13 females with mean age of 27.1 years) from 19 electrodes during an eyes-closed session referenced to the left and right mastoids. Signals were sampled at 200 Hz, bandpass filtered at 0.5-40 Hz and notch filtered at 50 Hz. The final dataset consists of 54 samples of 30 s duration each which have been carefully inspected to ensure no significant contamination by physiological or external artifacts [48]. For our simulation data we used 50 of these EEG samples which we refer to as "participants" for the remainder of the paper. With our group's long-term goal of investigating EEG of ADHD individuals, we opted to focus on the following 16 EEG locations: Fz, F3, F4, F7, F8, Cz, C3, C4, P3, P4, P7, P8, T7, T8, O1, O2. These locations were chosen to find a balance between being relevant for ADHD diagnosis and not being too severely or too minimally affected by artefacts [19].

The EOG artefacts were also obtained from Klados and Bamidis [48]. Vertical-EOG (VEOG) and horizontal-EOG (HEOG) were measured from the same 27 participants during an eyes-opened (EO) session, using four electrodes placed above and below the left eye and another two on the outer canthi of each eye. The VEOG data is calculated as the upper minus the lower EOG electrode recordings and the HEOG data is equal to the left minus the right EOG electrode recordings. The data is bandpass filtered between 0.5-5 Hz. For our application the VEOG and HEOG data were used in combination with propagation factors. The distribution of the EOG artefacts vary over the scalp and can be described in terms of propagation factors. The propagation factors are percentages used to determine the fraction of the VEOG and HEOG artefacts at a particular electrode location. The propagation factors are based on EOG and EEG data recorded on 23 subjects (14 females, mean age of 32 years) by Lins et al. [50]. Due to the propagation factors being based on real data, it is assumed that the effects of volume conduction is included in these values. The data was sampled at 200 Hz and bandpass filtered at 0.15-70 Hz. The EOG propagation factors were determined as the slopes of the best-fit straight lines relating the EOG with the EEG signal at any electrode location [50]. This resulted in different propagation factor ranges (for the individual differences) at different locations over the scalp (these can be seen in [50]). To simulate a variety of EOG artefacts we randomly chose a propagation intensity using a symmetric probability distribution. The propagation intensity was used to select the corresponding propagation factors in the given ranges and apply these to the VEOG and HEOG data for each electrode location. The resulting VEOG and HEOG data were then linearly combined to form the EOG data. This was done for each of the 50 participants.

The EMG data are based on the work by Goncharova et al. [28] wherein they defined the spectral and topographical characteristics of frontalis and temporalis muscle EMG artefacts over the entire scalp. EEG (64 channels) signals and EMG signals from four facial locations (right and left frontalis and anterior temporalis muscles) were recorded on 25 healthy adults (12 males, mean age of 35 years) during weak (15% of maximum) contractions of the frontalis (produced by raising eyebrows) and temporalis muscles (produced by jaw clenching). The frontalis and temporalis muscles are believed to be the most common source of EMG over the frontal and central head regions [28]. These EMG signals typically have a broad frequency distribution (0-200 Hz) which are attenuated and broadened centrally. A further investigation by Goncharova et al. [28] was conducted on 10 of the subjects to investigate EMG contamination as a function of the strength of frontalis muscle contraction. During this experiment the subjects contracted the frontalis muscle at 15% of maximum, 30% of maximum, 70% of maximum and maximum contraction. Visual feedback was presented on a screen indicating the EMG amplitude to control for muscle strength. Data was sampled at 512 Hz and bandpass filtered at 0.1-200 Hz. Fig. 1(a) shows the amplitude spectra of the EMG data at 15% contraction for channel F8 as extracted from Goncharova et al. [28]. Fig. 1(b) shows the amplitude spectra at four different contraction strengths measured at the left frontalis position as extracted from [28]. We only consider the spectra up to a frequency of 50 Hz although it is available up to 200 Hz. Using a linear regression approach, we fitted a straight line (by minimizing the squared error) between the amplitudes at the four contraction strengths for each frequency. This enabled the calculation of the gradient, m_f, relating the change in amplitude (ΔV_f) with the percentage change in contraction strength (ΔP_f) at each frequency (f), as represented in Eq. (3):

$$m_f = \frac{\Delta V_f}{\Delta P_f}. \quad (3)$$

The gradient, m_f, for each frequency was then used in combination with the data from the frontalis and temporalis contractions at 15%, which is represented by V_{15f} in Eq. (4):

$$V_{xf} = m_f(x - 15) + V_{15f}, \quad (4)$$

to calculate the new amplitude at a certain percentage of contraction and frequency V_{xf}, where x represents the new percentage of contraction. Since the amplitudes of the EMG at different frequencies and percentage of contraction is based on real EMG data, it is reasonable to assume that the effect of volume conduction is inherently accounted for in these values.

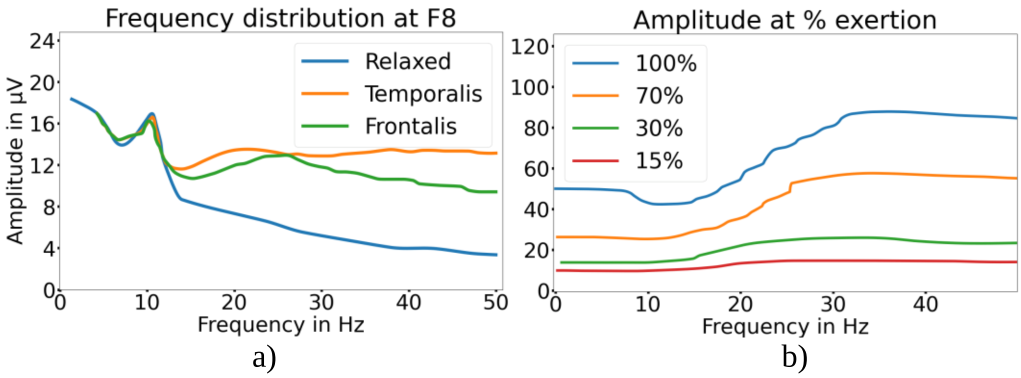


Fig. 1 (a) the amplitude spectra at F8 for the relaxed state and at 15% contraction of the frontalis and temporalis muscles; (b) the amplitude distribution at four different contraction strengths measured at the left frontalis position. Figures are adapted from [21].

Now, EMG data was simulated at each channel by selecting a random contraction strength and using Eq. (4) to determine the EMG contaminated EEG at the specific electrode location and for the selected contraction strength. The relaxed state EEG at the particular electrode location was then subtracted leaving only the contribution of the EMG. This was done separately for the frontalis and temporalis data which was then added to the first and second half of the "clean" EEG data, respectively. We decided to keep the frontalis and temporalis EMG data separate to separately investigate their contributions to the contaminated EEG. The duration of the EMG artefacts as they were added to the EEG data were also controllable. EMG artefacts were therefore added to the EEG data of the 50 participants by adjusting the durations and the contraction strengths.

The amplitude of cardiac activity is usually low and relatively easy to correct [77]. However, pulse artefacts, which occur when the EEG is placed over a pulsating blood vessel such as a scalp artery, is much harder to correct as it may resemble EEG activity [3, 77]. We therefore decided to simulate ECG contamination in the form of pulse artefacts which will typically only affect one EEG channel (due to it being unlikely that more than one electrode will lie directly over a scalp artery). Ten ECG recordings were selected from the 300 samples available for research purposes from Khamis et al. [45]. We selected ten recordings by visual inspection to ensure "clean" data with minimal contamination. The selected recordings were further chosen based on visual inspection of their shape, frequency, shift, and amplitude to ensure high variance between them. The data was additionally bandpass filtered at 3-5 Hz (based on visual inspection) to ensure "clean" pulse data. For each of the 50 participants a random sample was chosen, the amplitudes were scaled by a randomly chosen percentage and the ECG artefact was added to one randomly selected channel. The ECG was concatenated to match the 30 s length of the EEG.

Our approach as discussed above makes it possible to vary the contribution of the different artefacts in terms of location and frequency. We varied the EOG, EMG and ECG data within realistic limits as found in literature and linearly added them to the clean EEG to produce 50 sets of contaminated EEG data. In Eq. (5)

$$y(t) = x_{i,E}(t) + a_{i,O}x_O(t) + a_{i,M}x_M(t) + a_{i,C}x_C(t) + N, \tag{5}$$

$y(t)$ represents the result of the linear contamination of the clean EEG, $x_{i,E}(t)$, at a specific location, i, and time, t, with the coefficients, a, and artefact based, O (EOG), M (EMG), and C (ECG), conversions of the physiological artefacts. Finally, the total number of non-physiological noise sources, denoted as N, is assumed to be zero due to the preprocessing and manual data rejection applied on the EEG, EOG, EMG, and ECG signals. Specifically, Klados and Bamidis [48] applied preprocessing and manual data rejection on the EEG and EOG, and we applied preprocessing on the EMG and ECG signals. In addition, the contributions of the EOG, EMG, and ECG signals were investigated separately and together in this study. Therefore, in the case of individual analysis, the coefficients, a, of the other artefacts were set to zero.

This method culminated in a large set of contaminated EEG data, with realistically varying levels of contamination, and comprising different types of artefacts, where the noiseless signal is known a priori. This facilitates the evaluation of EEG cleaning methods and their performance dealing with different kinds of artefacts which is of great significance [37, 77].

Signal to noise ratio for performance evaluation

One advantage of using simulated EEG data is that one can assess the quality of the signal before and after the artefact removal through standard performance measures. The metric most commonly employed to represent the signal's energy, compared to the artefacts' energy, is the SNR. More specifically, the signal to noise ratio can be defined as the ratio of the power spectral density (PSD) of the clean EEG data to the PSD of the artefacts [46, 77].

The SNR is based on the linear mixture model stating that the contaminated data, $X^{(c)}$ is a linear mixture of the clean EEG data, $X^{(s)}$, and artefacts, $X^{(a)}$ [46, 77]. Eq. (6) describes the SNR for one channel, i, where $x_{i,j}$ represents the sample point at a certain channel for the PSD of the EEG data. N is the total number of sample points and n the number of electrode channels. For the simulation of the data, the artefact data, $X^{(a)}$ is known, and therefore it is simple to quantify the amount of contamination using the SNR as shown in Eq. (6)

$$SNR_i = 10log_{10}\left(\frac{\sum_{j=1}^{N} x_{i,j}^{(s)}}{\sum_{j=1}^{N} x_{i,j}^{(a)}}\right). \tag{6}$$

When testing the effectiveness of the cleaning methods, we do not directly know the amount of artefacts, $X^{(a)}$, that are still present when the simulated contaminated data is cleaned. $X^{(a^*)}$ can however be calculated as shown in Eq. (7):

$$X^{(a*)} = X^{(k)} - X^{(s)}, \tag{7}$$

where $X^{(k)} = \left[x_{i,j}^{(k)}\right]_{n\times N}$ is the cleaned data.

Therefore, to calculate the SNR of the cleaned data and the amount of artefact data removed, one can use Eq. (8), where the ideal situation would be for the denominator to be zero, meaning the SNR would be infinity, and the cleaned data matches the "clean" data [46, 77]:

$$SNR_i = 10log_{10}\left(\frac{\sum_{j=1}^{N} x_{i,j}^{(s)}}{\sum_{j=1}^{N}(x_{i,j}^{(k)} - x_{i,j}^{(s)})}\right). \tag{8}$$

BSS cleaning methods

Three popular BSS methods, namely Extended Infomax, SOBI and CCA, were tested. Their ability to increase the SNR of the contaminated data was investigated. This was done to validate that the semi-synthetic dataset developed produced similar evaluation results to those in the literature.

With the BSS methods, we first identified all the components. After the algorithms estimated the components, the artefact components were manually identified, marked, and removed from the mixing and component matrices before calculating the cleaned data. This process was repeated for all 50 participants, for each contamination type and each cleaning method. The cleaning methods were developed in Python 3.8.5. They were either developed from scratch or partially, using relevant libraries. The libraries used included MNE-Python 0.23.0 [29], NumPy 1.20.1 [33], Matplotlib 3.3.3 [38] and scikit-learn 0.24.2 [60]. To test for differences in performance between the cleaning methods a t-test was used with $p < 0.05$ set as the significance level.

Results

Semi-synthetic data

The distinguishing characteristics that define the time-series of the observed simulated EOG, EMG, ECG and the combined signals remained constant for all the channels and participants, with only the amplitude, frequency distribution and length changing within the valid ranges.

Fig. 2 shows the simulated EOG signal for a randomly chosen participant. Fig. 2(a) shows the EOG only signal at the Cz position. Fig. 2(b) shows the "clean" EEG data and the EOG contaminated EEG data. Fig. 2(c) is a close-up view of Fig. 2(b), showing the contaminated data in black and "clean" data in blue between the zero- and four-second range.

The simulated EOG signal in Fig. 2(a) shows similar slow frequency, high amplitude, and brief patterns to the EOG signal simulated by Zeng et al. [83]. Furthermore, the EOG signal is non-stationary, varying in amplitude (10 to 100 μV) and frequency (0 to 10 Hz), which is characteristic of real EOG signals [65]. The short duration of the EOG signal is also similar to what is expected from real EOG signals [65, 78]. Considering the SNR, it can be seen in Fig. 7 that the EOG has a lowest and highest SNR of -18 dB and 15 dB, respectively. This is comparable to other studies [56, 59, 62]. Looking at the EOG data SNR distribution in Fig. 7, it is clear that the SNR of each region differs significantly relative to each other. The frontal/temporal SNR is also much lower than the rest as this is the region most affected by the VEOG signal. The second most affected region is the temporal area, influenced mainly by the HEOG activity.

Fig. 3 shows the results of the simulation of the temporalis and frontalis time series data, simulated for five seconds, using the temporalis and frontalis frequency from Goncharova et al. [28] as reference for a random channel and participant. Figs. 3(a) and 3(b) describes the temporalis data simulation and Figs. 3(c) and 3(d) the frontalis. The dashed blue line in Fig. 3(a) represents the temporalis frequency reference data. The solid blue line represents the Fourier Transform of the time-series simulated data which was simulated based on an Inverse Fourier Transform of the reference data. The reference data, the dashed lines on Figs. 3(a) and 3(c), were first adjusted according to the chosen percentage at the relevant frequencies by the appropriate amplitude using Eq. (4).

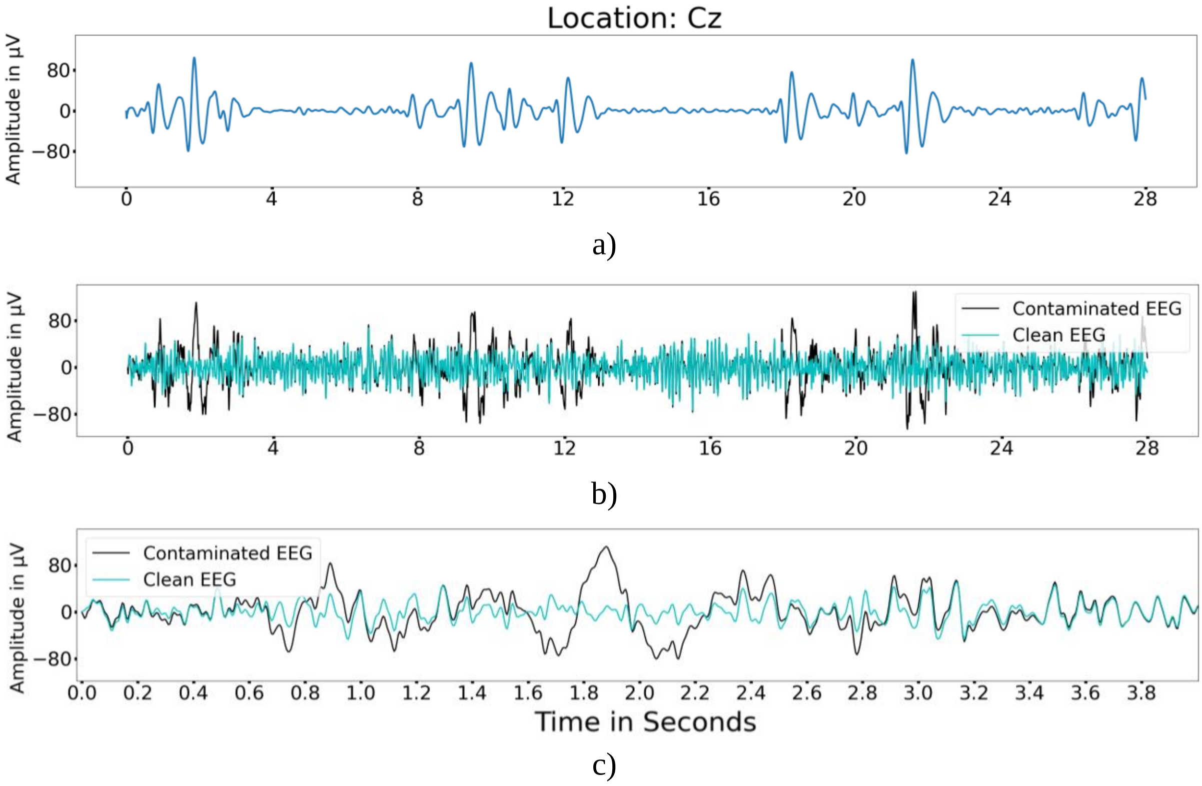


Fig. 2 (a) pure EOG data at location Cz after propagating and combining the HEOG and VEOG data; (b) the contaminated and "clean" EEG data; (c) an expanded view of (b).

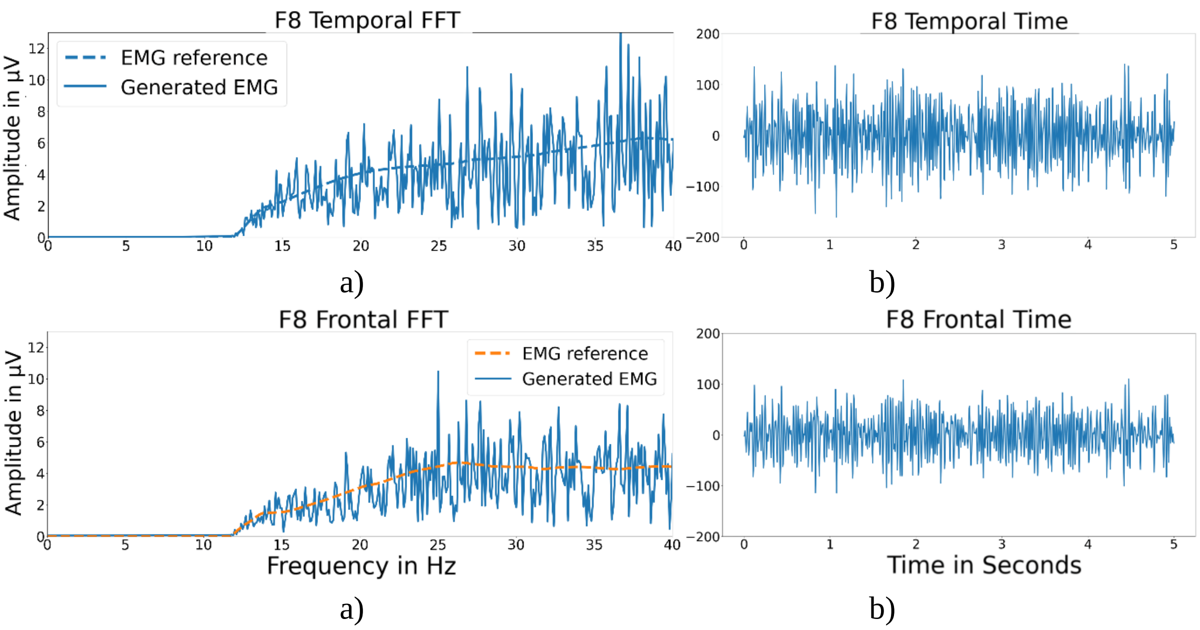


Fig. 3 (a) and (c): the solid line represents the amplitude spectra of simulated time-series temporalis and frontalis contamination and the dashed line represents the amplitude spectra from the reference temporalis and frontalis data, adjusted according to the chosen percentage; (b) and (d): the simulated time-series data of the temporalis and frontalis contamination.

Fig. 4 shows the result of simulating the EMG data for a random channel and participant. Fig. 4(a) shows the simulated temporalis data in the first half and the simulated frontalis data in the second half of the time series simulation. Fig. 4(b) shows the "clean" and contaminated EEG data. Fig. 4(c) is an expanded view of the EMG contamination time-series data.

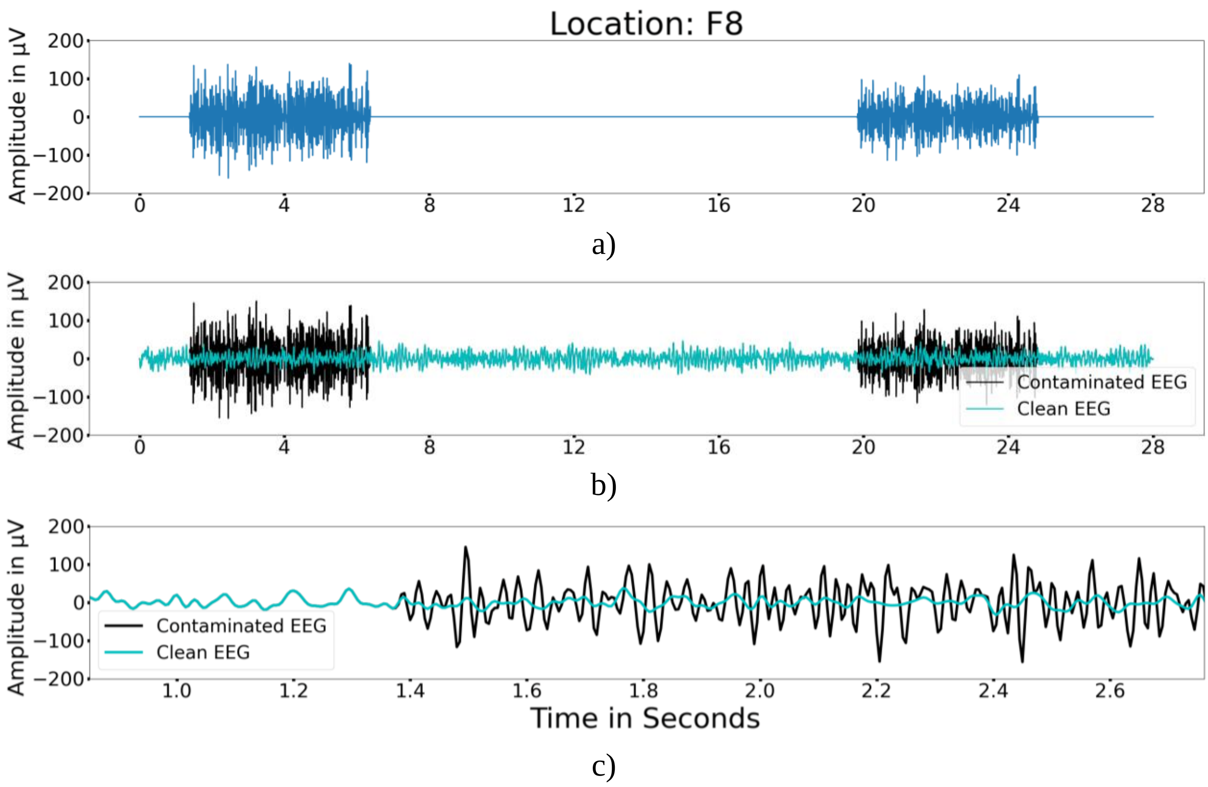


Fig. 4 (a) pure EMG data at location F8 after combining temporalis and frontalis data; (b) the contaminated and "clean" EEG data; (c) an expanded view of (b).

Considering the EMG data, Figs. 3(a) and 3(b) indicates that the amplitude spectra of the simulated data closely follow the amplitude spectra of the reference data (which is based on the data from Goncharova et al. [28]) but with high variance. A limitation of the simulation is the low number of data points available in the short time-series simulation. As a result, it is difficult to accurately capture the 0 to 40 Hz range of frequency data when one second only contains two hundred data points. As seen in Fig. 4(b), the EMG activity can produce magnitudes much higher than the EEG signals. This is in line with observations in literature [52, 55, 76]. Referring to Fig. 4(c), one can see that the original rhythm of the "clean" EEG data is entirely obscured by the EMG artefacts, making the analysis and interpretation of the EEG signals difficult. This is in line with observations by previous work [52, 76]. The EMG signal displays a spontaneous bursting behaviour of Gaussian noise, also characteristic of EMG signals [8, 51, 55]. Fig. 7 shows that the EMG contamination ranges between the lowest and highest SNR values of -30 dB and 15 dB. This EMG SNR range is similar to other studies [12, 64].

Regarding the EMG data SNR distribution in Fig. 7, one can see that the temporal regions have the largest SNR values, followed by the frontal/temporal region, the central region, and finally the occipital/parietal region. It makes sense that the central SNR is higher than the occipital/parietal region because it is closer to the frontal/temporal and temporal region, where the EMG originated.

Fig. 5(a) shows the concatenated and clipped ECG data to represent a longer time span. Fig. 5(b) represents the ECG contaminated EEG data after the ECG data has been added to a random participant and random location, in this case, F8, and its amplitude decreased by a random percentage. Fig. 5(c) shows an expanded view of the ECG contamination.

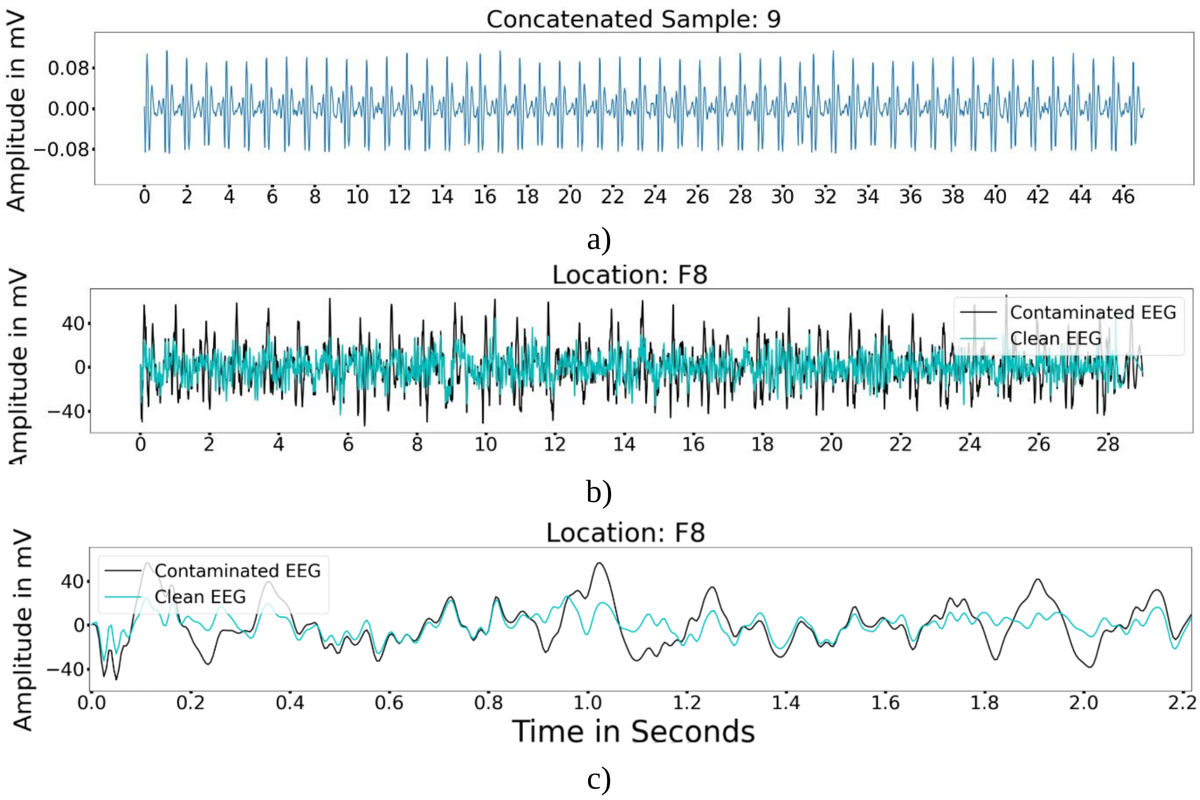
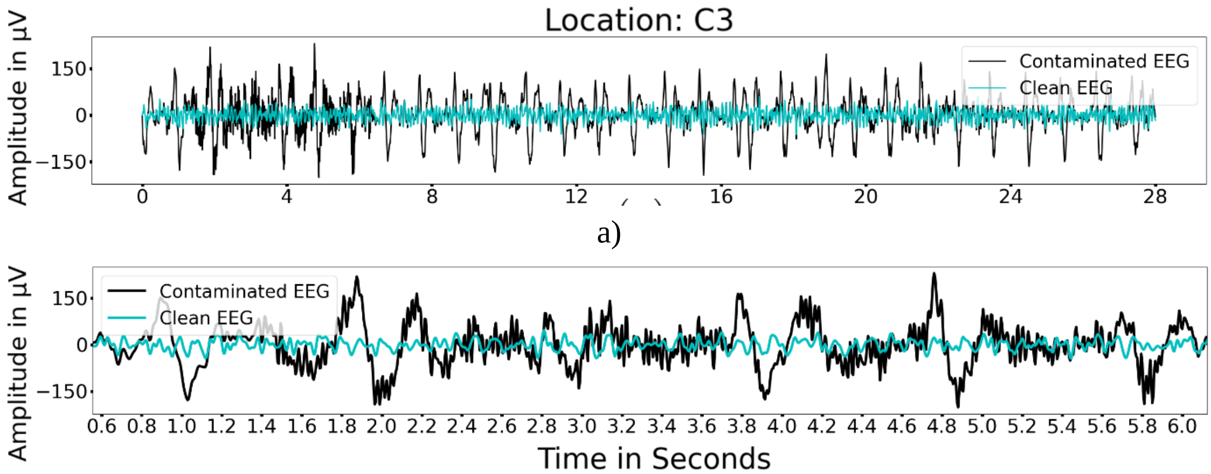


Fig. 5 (a) the ECG data after it has been bandpass filtered, clipped and concatenated to represent a longer time span; (b) the ECG contaminated EEG data and EEG data after the ECG artefact has been added to a random channel and its amplitude decreased by a random amount; (c) an expanded view of (b).

Regarding the ECG pulse data and Fig. 5(b), it is clear that the ECG signal shows a simple and periodic pattern, characteristic of ECG [70, 77]. The amplitude of the ECG data in Fig. 5(b) is relatively low but the amplitude greatly depends on the electrode position relative to the scalp artery and it differs for different participants [21, 77]. Referring to Fig. 5(e), one can see that the original rhythms of the "clean" EEG signal are not highly distorted, as noted by Taha and Raheem [75] regarding ECG contamination. Fig. 7 shows that the ECG contamination ranges between a lowest and highest SNR of -7 dB and 15 dB. These ranges are comparable to literature [11, 21, 22, 36, 57]. The ECG data in Fig. 7 shows similar SNR distributions for each region. Ideally, the SNR for each region would have been normally distributed but is skewed due to the limitation of total contaminated channels.

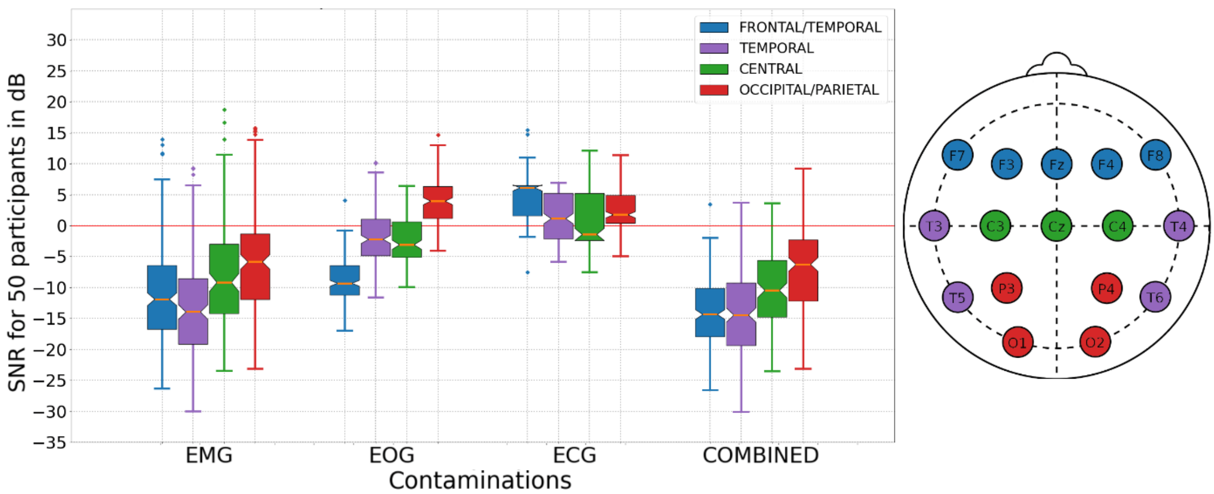
Fig. 6(a) shows the "clean" EEG data and the EEG signal contaminated by ECG, EMG, and EOG artefacts at C3. In Fig. 6(a), the original "clean" EEG data is blue/green, and the contaminated EEG data is black. Fig. 6(b) shows an expanded view and it is clear that the distortion is significant and detrimental. Fig. 7 shows the SNR distribution of all 50 datasets of the semi-synthetic artefact data.

Fig. 6(b) shows that combining the three physiological artefacts results in an evidently significant and detrimental distortion of the original "clean" EEG data, typical of measured contaminated EEG signals. Analysing the SNR of the combined-artefact data in Fig. 7, one can see that it has the lowest overall SNR. The combined-artefact data has a similar trend to that of the EMG contaminated data, with the temporal and frontal/temporal regions displaying the lowest SNR, with the SNR increasing slightly as one moves towards the central regions and finally the occipital/parietal regions. Ultimately the result is contaminated EEG data of which the noiseless data is available a priori. Such a dataset is extremely valuable when assessing different cleaning methods.



a) b)

Fig. 6 (a) results of combining the ECG, EMG and EOG data to the EEG data; (b) expanded view of (a).



a) b)

Fig. 7 (a) the SNR distribution of the different regions for the different types of contamination; (b) the legend depicting the positions of the data, with the colours of different locations marked, referring to the locations used to calculate the SNR on the left sub-figure.

BSS cleaning methods

Fig. 8 shows the distribution of the difference in SNR between the cleaned data and the original contaminated data for the entire head for all 50 participants. Each boxplot shows the SNR difference for 16 locations for each of the 50 participants, therefore representing 800 data points. Note that all differences were statistically significant at $p < 0.05$ unless otherwise indicated. The results of the four methods for each type of contamination are shown. All results are significantly different except for $*p = 0.06 > 0.05$. $**p = 0.17 > 0.05$, $***p = 0.57 > 0.05$.

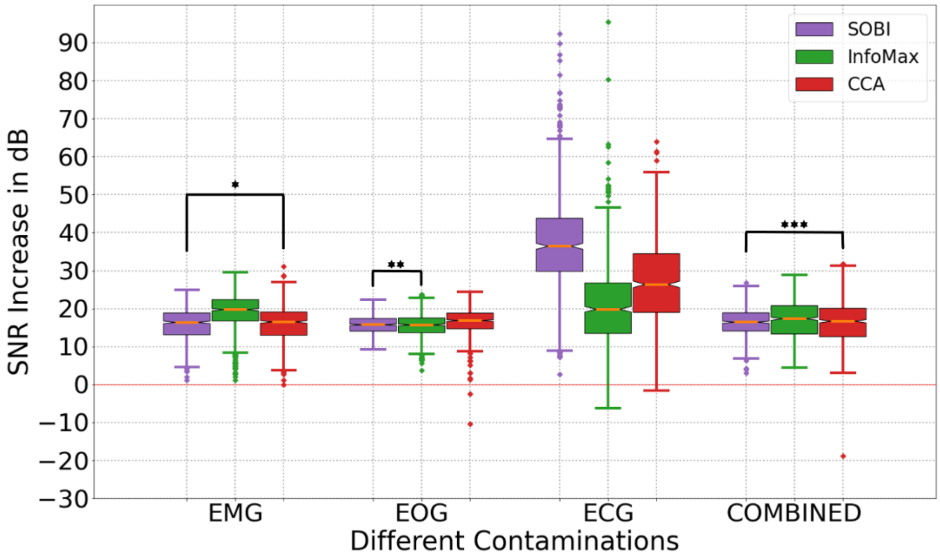


Fig. 8 The boxplots show the distribution of the difference in SNR between the cleaned data and the contaminated data for the whole head for all 50 participants

ICA methods such as SOBI and Extended Infomax have become the default choice for removing EOG artefacts from EEG data [11]. However, there is no consensus on which particular method works best for removing EOG artifacts as different studies support and promote either Extended Infomax [44, 77], SOBI [43, 66] or CCA [2]. The support for different methods could be due to the subjectivity and the expertise involved in manual identification of the artefact containing components [2].

Referring to Fig. 8, one can see that the three methods evaluated (SOBI, Extended Infomax and CCA) performed similarly on the EOG artefacts. The CCA showed a slightly higher SNR compared to the other two which was found to be statistically significant. These results support literature that shows the three BSS methods are comparable in their ability to remove EOG artefacts [2, 43, 44, 66, 77].

CCA is often proposed as a more reliable method for removing EMG artefacts than ICA methods [9, 74, 77]. However, results from literature show that SOBI and Extended Infomax are as good as CCA at removing EMG artefacts from EEG data [68, 77]. Our results, seen in Fig. 8, shows that Extended Infomax performed statistically better overall with the removal of the EMG artefacts. The results of SOBI and CCA in removing the EMG artefacts were similar. These results show that the three BSS methods are comparable, supporting similar findings in the literature [9, 51, 68, 74, 77].

ICA methods are typically the preferred methods when it comes to removing ECG artefacts [39, 77], with SOBI reported to generally perform better than other methods in removing ECG artefacts [15, 77]. From Fig. 8, we can see that SOBI outperformed the other methods with statistically significant results, followed by CCA and Extended Infomax.

Most research involving cleaning methods usually focuses on only one type of artefact. As observed in the literature, Extended Infomax methods are primarily applied to EOG artefact removal [44, 77], SOBI methods to removing EOG and ECG artefacts [15, 77] and CCA methods to removing EMG artefacts [9, 74, 77]. Testing the methods on the combination of these artefacts is relatively unexplored [16, 25, 82]. Referring to Fig. 8, the three BSS methods show very similar results. Extended Infomax performed slightly better than CCA and SOBI.

Discussion

The results suggest that it is indeed possible to create realistic semi-synthetic EEG data contaminated with several physiological artefacts. Furthermore, the artefacts can be adjusted to simulate a wide range of variability as commonly seen in practice, and it can be added individually or in combination.

In this study, we presented a semi-synthetic dataset to simulate various types of artefacts commonly observed in EEG recordings, including EOG, EMG, and ECG artefacts. The simulated EOG signal displayed slow frequency, high amplitude, and brief patterns, which are similar to what is expected from real EOG signals, with comparable location specific SNRs to those reported in other studies [56, 59, 62, 65]. The simulated EMG signal also displayed a spontaneous bursting behavior of Gaussian noise, which is characteristic of EMG signals, with comparable location specific SNR ranges to those in the literature [12, 51, 55, 64]. Furthermore, the ECG signal displayed a simple and periodic pattern, similar to what is expected from real ECG signals, with SNR ranges that are also comparable to literature [11, 21, 22, 36, 57]. The characteristics of these simulated artefacts make them realistic and representative of the types of artefacts that can be encountered in real-world EEG data.

By evaluating the performance of different artefact removal methods on these realistic artefacts, this study provides valuable insights into the effectiveness of different methods for removing artefacts from EEG data. Our results showed that the simulated artefacts were able to contaminate the EEG signal and introduce significant distortion, making the analysis and interpretation of the EEG signals difficult.

To evaluate the effectiveness of different artefact removal methods, we tested three popular BSS methods, including SOBI, Extended Infomax, and CCA. The results showed that all three methods performed similarly in removing EOG artefacts, with CCA having a slightly higher SNR compared to the other two methods, comparable to what was found in literature [2, 43, 44, 66, 77]. In removing EMG artefacts, Extended Infomax performed statistically better overall, while the results of SOBI and CCA were similar, consistent with previous findings in the literature [9, 51, 68, 74, 77]. When it came to removing ECG artefacts, SOBI outperformed the other methods with statistically significant results, followed by CCA and Extended Infomax, as expected from literature [15, 77].

It is important to note that most research on cleaning methods usually focuses on only one type of artefact, and the combination of artefacts is relatively unexplored. Our study highlights the value of a semi-synthetic dataset for assessing different cleaning methods and suggests that different BSS methods may be preferred for different types of artefacts. Overall, these findings contribute to the assessment and development of more effective and accurate artefact removal methods for EEG data analysis.

Conclusion

This work addresses the need for a verifiable ground truth to evaluate EEG cleaning and artefact removal methods. The paper proposes a direct measurement method of the quality of artefact removal resulting from a given cleaning method by starting with a known “clean” EEG data set. The properties of actual artefacts are analysed and used to generate semi-synthetic data with verifiably similar properties. Semi-synthetic artefacts are added to the “clean” data allowing the generation of a wide range of data impractical with conventional measurement processes. In addition, it provides measurable parametric profiles for the degree of artefact inclusion.

Cleaning methods can be applied to the semi-synthetic dataset and the resulting cleaned data can be compared to the original dataset producing a quantifiable, repeatable comparison. Therefore, the semi-synthetic dataset developed can be used in future studies to test and compare the robustness of different cleaning methods. This study developed a simple, accurate, and diverse semi-synthetic dataset that effectively compares artefact removal methods. This method, using semi-synthetic datasets, is believed to provide a straightforward yet stable datum for comparing alternative cleaning methods.

References

1. Acar Z. A., S. Makeig (2013). Effects of Forward Model Errors on EEG Source Localization, Brain Topography, 26, 378-396.
2. Al-Nuaimy F. N. H. (2020). A New Eliminating EOG Artifacts Technique Using Combined Decomposition Methods with CCA and HPF Techniques, Telkomnika, 18, 2580-2586.
3. Anderer P., S. Roberts, A. Schlogl, G. Gruber, G. Klosch, W. Hermann, P. Rappelsberger, O. Filz, M. J. Barbanoj, G. Dorffner, B. Saletu (1999). Artifact Processing in Computerized Analysis of Sleep EEG – A Review, Neuropsychobiology, 40, 150-157.
4. Barthelemy Q., L. Mayaud, D. Ojeda, M. Congedo (2019). The Riemannian Potato Field: A Tool for Online Signal Quality Index of EEG, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27, 244-255.
5. Barzegaran E., S. Bosse, J. Kohler, A. M. Norcia (2019). EEGSourceSim: A Framework for Realistic Simulation of EEG Scalp Data Using MRI-based Forward Models and Biologically Plausible Signals and Noise, Journal of Neuroscience Methods, 328, 108377, https://doi.org/10.1016/j.jneumeth.2019.108377.
6. Burger C., D. J. van den Heever (2015). Removal of EOG Artefacts by Combining Wavelet Neural Network and Independent Component Analysis, Biomedical Signal Processing and Control, 15, 67-79.
7. Cassani R., T. H. Falk, F. J. Fraga, A. M. Kanda, R. Anghinah (2014). The Effects of Automated Artifact Removal Algorithms on Electroencephalography-based Alzheimer's Disease Diagnosis, Frontiers in Aging Neuroscience, 6(55), https://doi.org/10.3389/fnagi.2014.00055.
8. Chavez M., F. Grosselin, A. Bussalb, F. De Vico Fallani, X. Navarro-Sune (2018). Surrogate-based Artifact Removal from Single-channel EEG, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(3), 540-550.
9. Chen H., W. Chen, Y. Song, L. Sun, X. Li (2019). EEG Characteristics of Children with Attention-deficit/hyperactivity Disorder, Neuroscience, 406, 444-456.
10. Chen X., A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, R. K. Ward (2016). Removing Muscle Artifacts from EEG Data: Multichannel or Single-Channel Techniques?, IEEE Sensors Journal, 16, 1986-1997.
11. Cho S., M. H. Song, Y. C. Park, H. S. Choi, K. J. Lee (2007). Adaptive Noise Cancelling of Electrocardiogram Artifacts in Single Channel Electroencephalogram, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology, Lyon, France, 3278-3281.
12. Choudhry M. S., R. Kapoor, Abhishek, A. Gupta, B. Bharat (2016). A Survey on Different Discrete Wavelet Transforms and Thresholding Techniques for EEG Denoising, Proceedings of the International Conference on Computing, Communication and Automation, Greater Noida, India, 1048-1053.

13. Craig J. K., E. Tenke (2005). Reference-free Quantification of EEG Spectra: Combining Current Source Density (CSD) and Frequency Principal Components Analysis (fPCA), Clinical Neurophysiology, 116(12), 2826-2846.
14. Croft R. J., R. J. Barry (1998). EOG Correction: A New Perspective, Electroencephalography and Clinical Neurophysiology, 107, 387-394.
15. Daly I., M. Billinger, R. Scherer, G. Muller-Putz (2013). On the Automated Removal of Artifacts Related to Head Movement from the EEG, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21, 427-434.
16. Daly I., N. Nicolaou, S. J. Nasuto, K. Warwick (2013). Automated Artifact Removal from the Electroencephalogram: A Comparative Study, Clinical EEG and Neuroscience, 44, 291-306.
17. Daly I., R. Scherer, M. Billinger, G. Muller-Putz (2015). FORCe: Fully Online and Automated Artifact Removal for Brain-computer Interfacing, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23, 725-736.
18. Delorme A., T. Sejnowski, S. Makeig (2007). Enhanced Detection of Artifacts in EEG Data Using Higher-order Statistics and Independent Component Analysis, NeuroImage, 34(4), 1443-1449.
19. Demos J. N. (2018). Getting Started with EEG Neurofeedback, 2nd Ed., W. W. Norton & Company.
20. Devuyst S., T. Dutoit, P. Stenuit, M. Kerkhofs, E. Stanus (2008). Cancelling ECG Artifacts in EEG Using a Modified Independent Component Analysis Approach, EURASIP Journal on Advances in Signal Processing, 747325, https://doi.org/10.1155/2008/747325.
21. Dora C., K. Biswal (2019). Efficient Detection and Correction of Variable Strength ECG Artifact from Single Channel EEG, Biomedical Signal Processing and Control, 50, 168-177.
22. Dora C., K. Biswal (2020). Correlation-based ECG Artifact Correction from Single Channel EEG Using Modified Variational Mode Decomposition, Computer Methods and Programs in Biomedicine, 183, 105092, https://doi.org/10.1016/j.cmpb.2019.105092.
23. Ermer J. J., J. C. Mosher, S. Baillet, R. M. Leahy (2001). Rapidly Recomputable EEG Forward Models for Realistic Head Shapes, Physics in Medicine and Biology, 46, 1265-1281.
24. Fitzgibbon S., D. M. W. Powers, K. J. Pope, C. R. Clark (2007). Removal of EEG Noise and Artifact Using Blind Source Separation, Journal of Clinical Neurophysiology, 24(3), 232-243.
25. Frolich L., T. S. Andersen, M. Morup (2015). Classification of Independent Components of EEG into Multiple Artifact Classes, Psychophysiology, 52, 32-45.
26. Gao J., C. Zheng, P. Wang (2010). Online Removal of Muscle Artifact from Electroencephalogram Signals Based on Canonical Correlation Analysis, Clinical EEG and Neuroscience, 41, 53-59.
27. Goh S. K., H. A. Abbass, K. C. Tan, A. Al-Mamun, C. Wang, C. Guan (2017). Automatic EEG Artifact Removal Techniques by Detecting Influential Independent Components, IEEE Transactions on Emerging Topics in Computational Intelligence, 1, 270-279.
28. Goncharova I. I., D. J. McFarland, T. M. Vaughan, J. R. Wolpaw (2003). EMG Contamination of EEG: Spectral and Topographical Characteristics, Clinical Neurophysiology, 114(9), 1580-1593.

29. Gramfort A., M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. S. Hamalainen (2013). MEG and EEG Data Analysis with MNE-Python, Frontiers in Neuroscience, 7, https://doi.org/10.3389/fnins.2013.00267.
30. Gratton G. (1998). Dealing with Artifacts: The EOG Contamination of the Event-related Brain Potential, Behavior Research Methods, Instruments, and Computers, 30, 44-53.
31. Gratton G., M. G. H. Coles, E. Donchin (1983). A New Method for Off-line Removal of Ocular Artifact, Electroencephalography and Clinical Neurophysiology, 468-484.
32. Guarnieri R., M. Marino, F. Barban, M. Ganzetti, D. Mantini (2018). Online EEG Artifact Removal for BCI Applications by Adaptive Spatial Filtering, Journal of Neural Engineering, 15, 056009, https://doi.org/10.1088/1741-2552/aacfdf.
33. Harris C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernandez del Rio, M. Wiebe, Peterson, Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant (2020). Array Programming with NumPy, Nature, 585, 357-362.
34. Haufe S., A. Ewald (2019). Simulation Framework for Benchmarking EEG-based Brain Connectivity Estimation Methodologies, Brain Topography, 32, 625-642.
35. Haufe S., V. V. Nikulin, K. Muller, G. Nolte (2013). A Critical Assessment of Connectivity Measures for EEG Data: A Simulation Study, NeuroImage, 64, 120-133.
36. Hou Z., Y. Dong, X. Wu (2020). A Template Addition Method for Eigentriple Rearrangement in Singular Spectrum Analysis for Processing Biopotential Signals with Extremely Lower SNRs, IEEE Sensors Journal, 20, 3142-3150.
37. Hu J., C. Wang, M. Wu, Y. Du, Y. He, J. She (2015). Removal of EOG and EMG Artifacts from EEG Using Combination of Functional Link Neural Network and Adaptive Neural Fuzzy Inference System, Neurocomputing, 151, 278-287.
38. Hunter J. D. (2007). Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95.
39. Iriarte J., E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, J. Artieda (2003). Independent Component Analysis as a Tool to Eliminate Artifacts in EEG: A Quantitative Study, Journal of Clinical Neurophysiology, 20, 249-257.
40. Islam M. K., A. Rastegarnia, Z. Yang (2016). Methods for Artifact Detection and Removal from Scalp EEG: A Review, Neurophysiologie Clinique/Clinical Neurophysiology, 46(4-5), 287-305.
41. Jiang X., G. B. Bian, Z. Tian (2019). Removal of Artifacts from EEG Signals: A Review, Sensors, 19(5), 987, https://doi.org/10.3390/s19050987.
42. Jose A. D., D. Collison (1970). The Normal Range and Determinants of the Intrinsic Heart Rate in Man, Cardiovascular Research, 4(2), 160-167.
43. Joyce C. A., I. F. Gorodnitsky, M. Kutas (2004). Automatic Removal of Eye Movement and Blink Artifacts from EEG Data Using Blind Component Separation, Psychophysiology, 41, 313-325.
44. Jung T., S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T. J. Sejnowski (2000). Removal of Eye Activity Artifacts from Visual Event-related Potentials in Normal and Clinical Subjects, Clinical Neurophysiology, 111, 1745-1758.
45. Khamis H., R. Weiss, Y. Xie, C. W. Chang, N. H. Lovell, S. J. Redmond (2016). 250 Telehealth ECG Records (Collected Using Dry Metal Electrodes) with Annotated QRS and Artifact Masks, and MATLAB Code for the UNSW Artifact Detection and

UNSW QRS Detection Algorithms, Research Domain, Software and Database, TELE ECG Database, https://doi.org/10.7910/DVN/QTG0E.

46. Kierkels J. J., G. J. van Boxtel, L. L. Vogten (2006). A Model-based Objective Evaluation of Eye Movement Correction in EEG Recordings, IEEE Transactions on Biomedical Engineering, 53, 246-253.
47. Klados M. A., C. Papadelis, C. Braun, D. Bamidis (2011). REG-ICA: A Hybrid Methodology Combining Blind Source Separation and Regression Techniques for the Rejection of Ocular Artifacts, Biomedical Signal Processing and Control, 6, 291-300.
48. Klados M. A., D. Bamidis (2016). A Semi-simulated EEG/EOG Dataset for the Comparison of EOG Artifact Rejection Techniques, Data in Brief, 8, 1004-1006.
49. Lagerlund T. D., F. W. Sharbrough, N. E. Busacker (1997). Spatial Filtering of Multichannel Electroencephalographic Recordings through Principal Component Analysis by Singular Value Decomposition, Journal of Clinical Neurophysiology, 14(1), 73-82.
50. Lins O. G., T. W. Picton, P. Berg, M. Scherg (1993). Ocular Artifacts in EEG and Event-related Potentials I: Scalp Topography, Brain Topography, 6, 51-63.
51. Liu Q., A. Liu, X. Zhang, X. Chen, R. Qian, X. Chen (2019). Removal of EMG Artifacts from Multichannel EEG Signals Using Combined Singular Spectrum Analysis and Canonical Correlation Analysis, Journal of Healthcare Engineering, Article ID 4159676.
52. Lugaresi E., G. Coccagna, M. Mantovani, R. Lebrun (1972). Some Periodic Phenomena Arising during Drowsiness and Sleep in Man, Electroencephalography and Clinical Neurophysiology, 32, 701-705.
53. Makeig S., T. Jung, A. J. Bell, D. Ghagreman, T. J. Sejnowski (1997). Blind Separation of Auditory Event-related Brain Responses into Independent Components, Proceedings of the National Academy of Sciences of the United States of America, 94(20), 10979-10984.
54. Minguillon J., M. A. Lopez-Gordo, F. Pelayo (2017). Trends in EEG-BCI for Dailylife: Requirements for Artifact Removal, Biomedical Signal Processing and Control, 31, 407-418.
55. Mucarquer J. A., P. Prado, M. J. Escobar, W. El-Deredy, M. Zanartu (2020). Improving EEG Muscle Artifact Removal with an EMG Array, IEEE Transactions on Instrumentation and Measurement, 69, 815-824.
56. Naga R., S. Chandralingam, T. Anjaneyulu, K. Satyanarayana (2012). Denoising EOG Signal Using Stationary Wavelet Transform, Measurement Science Review, 12, 46-51.
57. Navarro X., F. Poree, A. Beuchee, G. Carrault (2015). Denoising Preterm EEG by Signal Decomposition and Adaptive Filtering: A Comparative Study, Medical Engineering and Physics, 37, 315-320.
58. Patil S. S., M. K. Pawar (2012). Quality Advancement of EEG by Wavelet Denoising for Biomedical Analysis, Proceedings of the 2012 International Conference on Communication, Information and Computing Technology (ICCICT), Mumbai, 1-6.
59. Paulson K., O. Alfahad (2018). Identification of Multi-channel Simulated Auditory Event-related Potentials Using a Combination of Principal Component Analysis and Kalman Filtering, Proceedings of the 3rd International Conference on Biomedical Imaging, Signal Processing, Seoul, South Korea, 18-22.
60. Pedregosa F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay (2011). Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830.

61. Ponomarev V. A., A. Mueller, G. Candrian, V. A. Grin-Yatsenko, J. D. Kropotov (2014). Group Independent Component Analysis (gICA) and Current Source Density (CSD) in the Study of EEG in ADHD Adults, Clinical Neurophysiology, 125(1), 83-97.
62. Puthusserypady S., T. Ratnarajah (2006). Robust Adaptive Techniques for Minimization of EOG Artefacts from EEG Signals, Signal Processing, 86, 2351-2363.
63. Qayoom A., W. A. Rehman (2013). Artifact Processing of Epileptic EEG Signals: An Overview of Different Types of Artifacts, International Conference on Advanced Computer Science Applications and Technologies, Kuching, Malaysia, 358-361.
64. Queiroz C., G. M. da Silva, S. Walter, L. B. Peres, L. M. Luiz, S. C. Costa, K. C. de Faria, A. A. Pereira, M. F. Vieira, A. M. Cabral, A. O. Andrade (2022). Single Channel Approach for Filtering Electroencephalographic Signals Strongly Contaminated with Facial Electromyography, Frontiers in Computational Neuroscience, 16, 822987, https://doi.org/10.3389/fncom.2022.822987.
65. Reddy M. S., B. Narasimha, E. Suresh, K. S. Rao (2010). Analysis of EOG Signals Using Wavelet Transform for Detecting Eye Blinks, Proceedings of the International Conference on Wireless Communications and Signal Processing, Suzhou, China, 1-4.
66. Romero S., M. A. Mananas, M. J. Barbanoj (2008). A Comparative Study of Automatic Techniques for Ocular Artifact Reduction in Spontaneous EEG Signals Based on Clinical Target Variables: A Simulation Case, Computers in Biology and Medicine, 38, 348-360.
67. Sadasivan K., D. N. Dutt (1994). A Non-linear Estimation Model for Adaptive Minimization of EOG Artefacts from EEG Signals, International Journal of Bio-medical Computing, 36(3), 199-207.
68. Safieddine D., A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, I. Merlet (2012). Removal of Muscle Artifact from EEG Data: Comparison between Stochastic (ICA and CCA) and Deterministic (EMD and Wavelet-based) Approaches, EURASIP Journal on Advances in Signal Processing, 2012(1), 1-15.
69. Sai C. Y., N. Mokhtar, H. A. Cumming, M. Iwahashi (2018). Automated Classification and Removal of EEG Artifacts with SVM and Wavelet-ICA, IEEE Journal of Biomedical and Health Informatics, 22, 664-670.
70. Sakai M., D. Wei (2007). Detection of Electrocardiogram Mixed in Electroencephalogram by Stationarization, International Journal of Bioelectromagnetism, 9(2), 61-62.
71. Shou G., L. Ding (2013). Detection of EEG Spatial-spectral-temporal Signatures of Errors: A Comparative Study of ICA-based and Channel-based Methods, Brain Topography, 28, 47-61.
72. Suresh H. N., C. Puttamadappa (2008). Removal of EMG and ECG Artifacts from EEG Based on Real Time Recurrent Learning Algorithm, International Journal of Physical Sciences, 3, 120-125.
73. Sweeney K. T., H. Ayaz, T. E. Ward, M. Izzetoglu, S. F. McLoone, B. Onaral (2012). A Methodology for Validating Artifact Removal Techniques for Physiological Signals, IEEE Transactions on Information Technology in Biomedicine, 16, 918-926.
74. Sweeny K. T., T. E. Ward, S. F. McLoone (2012). Artifact Removal in Physiological Signals-practices and Possibilities, IEEE Transactions on Information Technology in Biomedicine, 16, 488-500.
75. Taha L. Y., E. Abdel-Raheem (2019). EEG signal Extraction Utilizing Null Space Approach, Proceedings of the IEEE 19th International Symposium on Signal Processing and Information Technology, Ajman, United Arab Emirates, 1-5.

76. Teng C., Y. Zhang, G. Wang (2014). The Removal of EMG Artifact from EEG Signals by the Multivariate Empirical Mode Decomposition, Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing, Guilin, China, 873-876.
77. Urigüen J. A., B. Garcia-Zapirain (2015). EEG Artifact Removal – State-of-the-art and Guidelines, Journal of Neural Engineering, 12, 031001, https://doi.org/10.1088/1741-2560/12/3/031001.
78. Venkataramanan S., P. Prabhat, S. R. Choudhury, H. B. Nemade, J. S. Sahambi (2005). Biomedical Instrumentation Based on Electrooculogram (EOG) Signal Processing and Application to a Hospital Alarm System, Proceedings of the International Conference on Intelligent Sensing and Information Processing, Chennai, India, 535-540.
79. Verleger R., T. Gasser, J. Möcks (1982). Correction of EOG Artifacts in Event-related Potentials of the EEG: Aspects of Reliability and Validity, Psychophysiology, 19(4), 472-480.
80. Vigario R., E. Oja (2008). BSS and ICA in Neuroinformatics: From Current Practices to Open Challenges, IEEE Reviews in Biomedical Engineering, 1, 50-61.
81. Wallstrom G. L., R. E. Kass, A. Miller, J. F. Cohn, N. A. Fox (2002). Correction of Ocular Artifacts in the EEG Using Bayesian Adaptive Regression Splines, Lecture Notes in Statistics, 167, 351-365.
82. Winkler I., S. Haufe, M. Tangermann (2011). Automatic Classification of Artifactual ICAComponents for Artifact Removal in EEG Signals, Behavioral and Brain Functions, 32, 701-705.
83. Zeng K., D. Chen, G. Ouyang, L. Wang, X. Liu, X. Li (2016). An EEMD-ICA Approach to Enhancing Artifact Rejection for Noisy Multivariate Neural Data, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24, 630-638.
84. Zou Y., V. Nathan, R. Jafari (2016). Automatic Identification of Artifact-related Independent Components for Artifact Removal in EEG Recordings, IEEE Journal of Biomedical and Health Informatics, 20, 73-81.

Wadda Benjamin du Toit, M.Sc.
E-mail: waddabenjamin@gmail.com

Wadda du Toit was a graduate student at the Department of Mechanical and Mechatronic Engineering at Stellenbosch University, South Africa.

Martin Philip Venter, M.Sc.
E-mail: mpventer@sun.ac.za

Martin Philip Venter is a senior lecturer at the Department of Mechanical and Mechatronic Engineering at Stellenbosch University, South Africa.

Assoc. Prof. David Vandenheever, M.Sc.
E-mail: davidvdh@abe.msstate.edu

David Vandenheever is an Associate Professor of Biomedical Engineering at Mississippi State University, USA. His main research focus is neural engineering, the processing and analysis of EEG data.