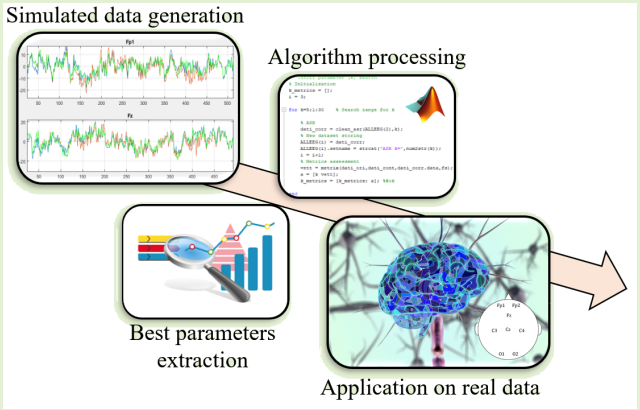
A Method for Optimizing the Artifact Subspace Reconstruction Performance in Low-Density EEG

Andrea Cataldo, *Senior Member, IEEE*, Sabatina Criscuolo, Egidio De Benedetto, *Senior Member, IEEE*, Antonio Masciullo, Marisa Pesola, Raissa Schiavoni, and Sara Invitto

***Abstract*—Electroencephalogram (EEG) plays a significant role in the analysis of cerebral activity, although the recorded electrical brain signals are always contaminated with artifacts. This represents the major issue limiting the use of the EEG in daily life applications, as the artifact removal process still remains a challenging task. Among the available methodologies, artifact subspace reconstruction (ASR) is a promising tool that can effectively remove transient or large-amplitude artifacts. However, the effectiveness of ASR and the optimal choice of its parameters have been validated only for high-density EEG acquisitions. In this regard, this study proposes an enhanced procedure for the optimal individuation of ASR parameters, in order to successfully remove artifacts in low-density EEG acquisitions (down to four channels). The proposed method starts from the analysis of real EEG data, to generate a large semisimulated dataset with similar characteristics. Through a fine-tuning procedure on this semisimulated data, the proposed method identifies the optimal parameters to be used for artifact removal on real data. The results show that the algorithm achieves an efficient removal of artifacts preserving brain signal information, also in low-density EEG signals, thus favoring the adoption of the EEG also for more portable and/or daily-life applications.**



***Index Terms*—Artifact removal, artifact subspace reconstruction (ASR), brain computer interface (BCI), electroencephalography, low-density system, measurement system.**

Manuscript received 30 August 2022; accepted 15 September 2022. Date of publication 28 September 2022; date of current version 31 October 2022. This work was supported in part by the Integrated Technologies and Enhanced Sensing for Cognition and Rehabilitation (INTENSE) Project—Italian Ministry of Economic Development Accordo Innovazione Ministerial Decree (DM) (31 December 2021) under Grant F/310148/01-05/X56. The associate editor coordinating the review of this article and approving it for publication was Prof. Mohammad Russel. *(Corresponding author: Egidio De Benedetto.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of Vito Fazzi Hospital, Lecce, Italy, and performed in line with the Declaration of Helsinki.

Andrea Cataldo, Antonio Masciullo, and Raissa Schiavoni are with the Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy (e-mail: andrea.cataldo@unisalento.it).

Sabatina Criscuolo and Egidio De Benedetto are with the Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, 80125 Naples, Italy (e-mail: egidio.debenedetto@unina.it).

Marisa Pesola is with the Center for Advanced Metrological and Technological Services (CeSMA), University of Naples Federico II, 80146 Naples, Italy.

Sara Invitto is with the Department of Biological and Environmental Sciences and Technologies, University of Salento, 73100 Lecce, Italy.

Digital Object Identifier 10.1109/JSEN.2022.3208768

I. Introduction

ELECTROENCEPHALOGRAPHY is a well-established neuroimaging technique largely used to analyze brain activity, mostly in clinics and laboratories [1], [2]. Indeed, new electroencephalogram (EEG)-based applications for more practical use are being investigated, thanks to their noninvasiveness, ease of use, and potential wearability and portability [3], [4], [5], [6], [7]. However, the use of the EEG outside of clinical and research settings is still limited because the cerebral signal is heavily influenced by noise and interference, leading to a variety of artifacts [8] that compromise the correct extraction of the features of interest [9]. Artifacts can be caused either by nonphysiological [10] or physiological sources [11]. The latter, which are more difficult to remove [12], are due, for example, to eye movements, blinks, and muscle activity [13], [14], [15]. As a result, EEG data is a nonstationary, nonlinear stochastic mixture of brain signals and artifacts. The removal of EEG artifacts and the identification of interfering signal outcomes are critical preprocessing steps for the correct measurement of neurophysiological phenomena

of interest related to brain activity [16], [17]. At the state of the art, there are many techniques and algorithms developed for *artifact removal*, which can be grouped into four major categories: *regression methods*, *filtering methods*, *blind source separation methods* (BSS), and *source decomposition methods* [11], [18], [19], [20], [21]. However, such techniques often require the use of additional reference channels (e.g., electrooculography, electromyography) or classifiers to identify and discard artifact-related components. Additionally, they carry a high computational burden [22]. Indeed, there is still no agreement on an optimal removal technique for all types of artifacts. Several factors, such as the total number of channels, the need for a reference channel, and the characteristics of the chosen algorithm (e.g., linearity, automation, and online applicability), need to be considered in practice. These factors may lead to increased system noise and complexity, but also increased user discomfort. The total number of channels, in particular, limits the ability to successfully remove artifacts; only a few algorithms are suitable for single-channel and multichannel applications [23].

To overcome these limitations, the *Artifact Subspace Reconstruction* (ASR) method has been proposed in recent years as an online, automatic, component-based artifact removal method for nonstationary large-amplitude or transient artifacts [24], [25]. As observed in the literature, the ASR method holds unexplored potential, although recent studies have reported promising results [26], [27], [28]. These works have mainly focused on using ASR to remove artifacts from high-density EEG acquisitions, typically using suboptimal default ASR parameters. However, it has recently been demonstrated that ASR performs better than other multichannel techniques as the number of channels decreases by down to four [29]. Moreover, excellent performance of ASR has been demonstrated in removing artifacts on eight-channel steady-state visual evoked potential signals [30]. Nevertheless, its application and optimization for low-density EEGs are still in question.

Based on these considerations, this study proposes a method for tuning ASR parameters in order to make artifact removal more efficient by considering eight, six, and four channels. In doing so, not only the user-defined *ASR rejection threshold parameter k* but also the *ASR sliding window length wl* were investigated. However, tuning these two parameters directly on real data is not possible, since the original pure signal is not available to calculate comparison metrics and quantify the efficiency of the correction. Thus, the basic idea of the proposed method lies in the generation of semisimulated data with characteristics similar to the real available data to be processed. Once the algorithm is tuned on the semisimulated data, it is possible to find the best ASR parameter values to apply to the real data.

The outline of this article is as follows. Section II provides the background on the ASR algorithm. Section III outlines the proposed method and the metrics used to evaluate its performance. Section IV describes the data used in this study and the implementation of the proposed approach. Finally, Section V reports the results obtained by applying the proposed approach to semisimulated data and real data.

II. Background

The basic concept of the ASR process is the extraction of reference statistics from an artifact-free data segment to calibrate the correction of contaminated data [24]. High-amplitude nonstationary artifacts, such as muscle artifacts and eye blinks, are identified and rejected with automatic thresholding in the domain of the principal components (PCs).

The ASR process consists of three major steps [22].

1) *Extraction of reference data:* A portion of the signal without artifacts is identified by calculating the root-mean-square (rms) values on 1-s sliding windows for each channel. Then, the z-score is computed along the entire channel to assess the dispersion degree and discern clean reference data. A minimum 30-s/1-m length is usually recommended for reference data, but the duration can vary.
2) *Threshold definition to identify artifact components:* After an infinite impulse response (IIR) filtering, a mixing matrix is calculated as the square root of the covariance matrix of the filtered reference data. Furthermore, the eigenvectors are used to project onto the PC space. In the projected space, rms values with mean μ_i and standard deviation σ_i are calculated on sliding windows of the new data. The default *sliding window length* (wl) is set at 0.5 s. Then, a rejection criterion Γ_i is determined by a user-defined *cutoff threshold parameter* (k) multiplied by the standard deviation

$$\Gamma_i = \mu_i + k \cdot \sigma_i. \tag{1}$$

   The cutoff parameter k establishes how aggressively faulty data are removed. Smaller values of k are associated with higher aggressiveness.
3) *Artifact component rejection and signal reconstruction:* Finally, the transformation procedures of the second step are applied to the uncleaned EEG data. For each window, the algorithm identifies which PC exceeds the rejection threshold Γ_i in the projection space. Artifact components that fulfill the criterion are set to zero before reconstructing the cleaned signal.

Hence, the ASR algorithm performance is heavily influenced by user-selected parameters, in particular, the already mentioned *cutoff threshold parameter* (k) and *sliding window length* (wl). However, as reported in Section I, the majority of ASR-based works use standard parameters [24], [30]. Studies focusing on the optimal value of k, in particular, revealed that this value could be between 20 and 30, which is small enough to remove artifacts and preserve most of brain information [27], [22].

III. Proposed Method

The proposed method allows for the customization of ASR parameters to improve artifact removal. Starting from semisimulated data, the two considered parameters k and wl are optimized, and then their best values are applied to the real data. Fig. 1 describes this three-step procedure in full depth.

1) *Preliminary analysis of real EEG data and generation of semisimulated dataset:* At this phase, the real EEG

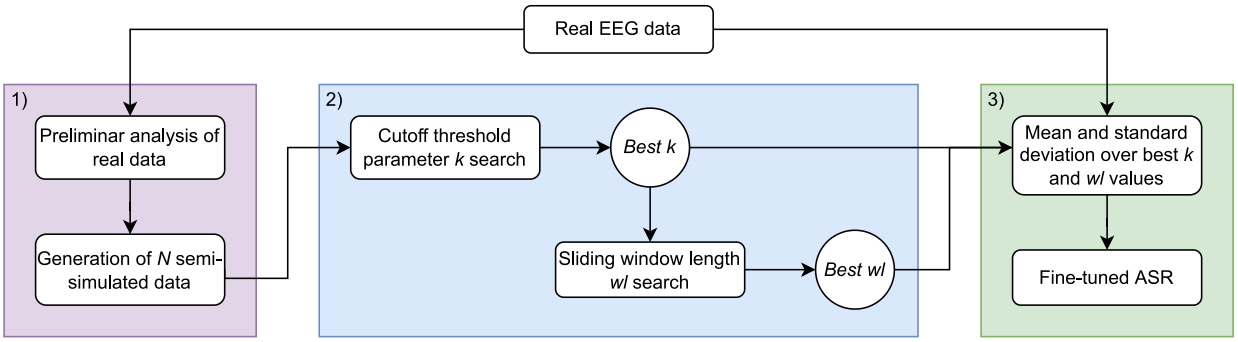


Fig. 1. Pipeline of the proposed method for fine-tuning ASR parameters.

data corrupted by artifacts are preliminarily analyzed. The data are analyzed both in the time and frequency domains in order to identify some characteristics, such as artifact type and duration, amplitude, and sampling rate. In particular, semisimulated EEG data are generated with the same channels and sampling rate as the real data. A further check was then made in terms of comparing the amplitudes of the two EEGs through the evaluation of the mean and the standard deviation of the signal in the time domain and, finally, the spectral similarity of the semisimulated and real EEG signals was quantified by comparing the two power spectral density (PSD). Based on these considerations, semisimulated data are generated to reflect the available real data characteristics.

2) *ASR parameters customization:* This step provides the first search to investigate aggressive and nonaggressive values of k at a fixed value of wl. Then, once the best value for k was calculated, the wl value was parameterized.
Regarding the automatic choice of the best k and wl values, it was carried out considering three metrics between pure and corrected signals: rms error (RMSE), Gamma value (γ), and correlation coefficient (ρ). Each of these metrics measures quantitatively how well the EEG dataset has been corrected and it can be calculated for the entire EEG trace and for specific signal conditions, that is, with muscular or ocular artifacts. Then, an autoselect function determines the k and wl that optimize the greatest number of metrics on the different segments of the signals. In addition to this quantitative evaluation, the results were visually inspected.
The procedure was iterated by determining the best k and wl parameters on the whole semisimulated dataset.

3) *Application of k and wl values to real data:* The mean and standard deviation of the best k and wl values obtained in the preceding step are computed. On real data, extreme and average values of the obtained range are used in the ASR algorithm. The effectiveness of artifact removal is evaluated through visual inspection.

A. Metrics for Automatic Choice of k and wl

On semisimulated data, several metrics can be found in the literature to evaluate artifact removal performance. Some of them rely on the evaluation of the distortion in each specific band of the EEG signal [31]. However, this kind of approach is not adequate for efficient numerical computation and does not preserve the overall power of the EEG signal. Starting from these considerations, in this work, the following three metrics were chosen [32].

1) rms Error: This is an absolute error measure in which deviations are squared to prevent positive and negative values from canceling each other out. With this measure, larger value errors are also amplified, a feature that can facilitate the elimination of methods with the most significant errors. The RMSE formula is

$$\sqrt{\sum_{i=1}^{n} \frac{(\mathrm{EEG}_{\mathrm{corr},i} - \mathrm{EEG}_{\mathrm{true},i})^2}{N}} \tag{2}$$

where $\mathrm{EEG}_{\mathrm{true}}$ is the original simulated EEG dataset, $\mathrm{EEG}_{\mathrm{corr}}$ is the corrected dataset after ASR, and N is the number of samples of data. In an ideal case of the corrected signal being perfectly equal to the true signal, the RMSE would be equal to 0.

2) Gamma Value γ: An efficient parameter for evaluating artifact removal enhancements is artifact removal gain γ. It is defined as the ratio between two different signal-to-artifact ratios (SARs)

$$\gamma = 10 \cdot \log\left(\frac{\mathrm{SAR}_A}{\mathrm{SAR}_B}\right) \tag{3}$$

where SAR_B is the SAR between $\mathrm{EEG}_{\mathrm{true}}$ and contaminated EEG signal ($\mathrm{EEG}_{\mathrm{cont}}$), while SAR_A is the SAR between $\mathrm{EEG}_{\mathrm{true}}$ and $\mathrm{EEG}_{\mathrm{corr}}$. Therefore, the γ value also takes into account the contribution of the corrupted EEG signal, which is not the case with RMSE, and this makes the parameter particularly useful. It is clear that positive gamma values identify an improved signal-to-noise ratio, while negative values indicate a decrease, and zero is no improvement at all.

3) Cross Correlation: Cross correlation is a measure of the similarity of two signals as a function of a time shift or translation applied to one of them. For $\mathrm{EEG}_{\mathrm{true}}$ and $\mathrm{EEG}_{\mathrm{corr}}$ discrete functions, the cross correlation is defined as

$$(\mathrm{EEG}_{\mathrm{true}} * \mathrm{EEG}_{\mathrm{corr}})[n] = \sum_{m=-\infty}^{\infty} \mathrm{EEG}_{\mathrm{true}}[m]\mathrm{EEG}_{\mathrm{corr}}[m+n] \tag{4}$$

TABLE I
SUMMARIZED CHARACTERISTICS OF REAL DATA AND SEMISIMULATED

	Real data	Semi-simulated data
Baseline	60 s	60 s
Artifacts	60 s	60 s
Sampling frequency	1000 Hz	256 Hz
Re-sampling frequency	200 Hz	200 Hz

where n is called displacement or lag and the complex conjugate of the signal does not appear since the EEG is a real signal. Since there is no interest in the translation of the signals, we considered the cross correlation value for $n = 0$, normalized between -1 and 1.

The automatic choice of k and wl parameters is made by selecting the parameters in order that:

1) the RMSE between EEG_{true} and EEG_{corr} was minimum and
2) the Gamma value and correlation coefficient between EEG_{true} and EEG_{corr} were maximum.

IV. IMPLEMENTATION

In this section, the implementation of the proposed method is described in detail.

A. Description of the Datasets

1) Real Data: Real EEG data were collected from a publicly available dataset in order to test artifact removal techniques [33]. Each trace was recorded with a Brain Products helmet with 27 EEG channels and three electrooculographic channels [34] at a sampling rate of 1000 Hz and then made available at a resampling rate of 200 Hz. This dataset contains clean and contaminated data from one recording session for each of the 13 subjects. Data were collected over the course of two experimental sessions. The first phase required participants to focus their attention on a fixation cross on a screen while avoiding movement. In this way, 30 s of clean signal (baseline) was acquired for each subject. Participants in the second phase performed muscular and ocular artifacts guided by cues on the screen. In random order, ten repetitions of nine different types of artifacts were performed, for a total length of 40–50 min. To reduce computational costs, the entire EEG trace was trimmed. The two baseline segments, in particular, were preserved, as were nine subsequent contaminated segments from the artifact conditions. Furthermore, eight channels were extracted to assume a few-channel acquisition. In the end, the raw EEG traces were base-normalized and filtered. As well known, base normalization allows for a signal normally distributed, whereas filtering is a preprocessing step fundamental to improving the signal-to-noise ratio, by attenuating noisy frequencies. As a matter of fact, EEG signals can often be exposed to strong power line interference at 50 or 60 Hz or can be influenced by the presence of a dc offset. For this reason, filtering is a good practice before the subsequent processing. Table I summarizes real data characteristics.

2) Semisimulated Data: Pure EEG signals were simulated with the generation function of the *MRC EEG data simulator* available online [35]. To simulate a pure signal, a duration of 120 s and a sampling frequency of 256 Hz were chosen.

TABLE II
CORRELATION COEFFICIENTS TO WEIGH OCULAR ARTIFACTS

Channel	Fp1	Fp2	Fz	C3	Cz	C4	O1	O2
Weight	1	1	0.73	0.24	0.01	0.12	0.31	0.28

The first 60 s of the trace were fully preserved to represent clean calibration data (baseline) for ASR applications, while the remaining 60 s were contaminated with real ocular and muscle artifacts. The used real artifact segments were extracted from the online *DenoiseNet database* [36] and combined with the pure semisimulated data. In order to obtain 30 s of muscular artifacts and 30 s of ocular artifacts, 152-s-long segments for each artifact kind were randomly extracted and combined. To properly combine these signals, the artifact amplitude was scale-adapted to obtain a signal-to-noise ratio from -20 to 5 dB [37]. Furthermore, the inclusion of ocular artifacts was weighed channel-wise due to different propagation of eye components over the scalp [38]. The weights for eight selected channels were determined by calculating the correlation coefficients between genuine electrooculographic data and the matching EEG real data [39]. Ocular artifacts were more visible in the frontal and occipital areas [40], whereas the central area was thought to better appreciate muscle movements [41]. Table II summarizes the selected channels and the aforementioned correlation coefficients.

The whole simulation process was repeated 20 times to obtain different virtual subjects with a random choice of artifacts. Finally, all the semisimulated signals were filtered and resampled to match real data characteristics, as shown in Table I. The spectral similarity of two EEG sets was quantified by comparing the PSD of each semisimulated data to the PSD of one of the real sources.

B. Algorithm Implementation

As described in Section III, the method proposed in this study entails first fine-tuning the parameters on a semisimulated dataset, based on the real data, followed by applying the mean of best results obtained on synthetic data to real data.

Moreover, the generated semisimulated signal was processed in *MATLAB* through *EEGLAB*, an open-source toolbox for EEG analysis [42]. With a focus on low-density EEG, the number of acquisition channels was reduced from eight (see Table II), associated with the brain areas mainly considered in brain computer interface (BCI) applications, to six channels (excluding Fp1 and $C4$) and then to four channels (excluding Cz and $O1$). The channels were reduced randomly, with at least one related to the brain areas initially considered remaining.

The ASR algorithm was implemented with *clean_rawdata()* EEGLAB plug-in, by using the *clean_asr.m* function. The basic principle is to find a baseline and perform statistics on it. With a sliding window on the data, the function searches the subspaces where activity deviates from the baseline. Then, the bad subspaces are treated as missing data, and their contents are reconstructed by using statistics calculated on clean data, ensuring the data does not contain unusually strong power.

As mentioned in Section III, this study focused on adjusting k and wl user-selected parameters to improve artifact removal with the ASR algorithm. In particular.

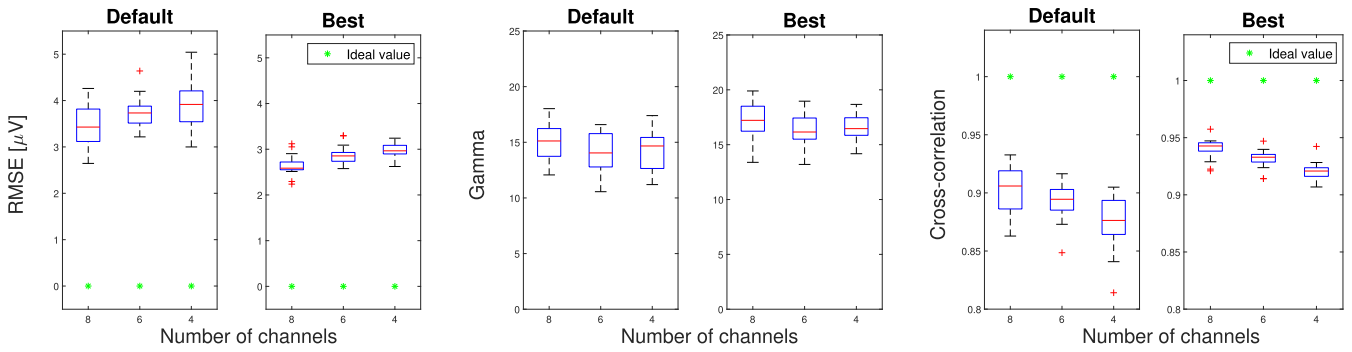


Fig. 2. Boxplot for all the semisimulated data when the number of channels decreases. The three considered metrics are calculated on the entire EEG trace. Red line: median value, blue box: interquartile range, and red cross: outlier. (a) RMSE value. (b) Gamma value. (c) Cross-correlation value.

1) *k search* has been made by varying k in a range from 5 to 30 with a 1-s step. Quite conservative value is 20, which is default *clean_asr.m* function parameter.
2) *wl [s] search*, once the best value for k, has been performed by changing its value in a range from 0.2 to 2 s, with a 0.1-s step. The default *clean_asr.m* function value is 0.5 s.

Before applying ASR, real raw signals were base-normalized and filtered with the EEGLAB functions *pop_rmbase.m* and *pop_eegfiltnew.m*. In particular, the EEG signals were filtered with a high-pass filter to filter out slow frequencies less than 0.5 Hz and with a notch filter (48–52 Hz) to eliminate the line interference at 50 Hz.

The described algorithm to optimize ASR parameters was made available at https://github.com/anthonyesp/low_density_eeg_asr.git

V. Experimental Results

The procedure described in Sections III and IV was first tested on 20 randomly generated semisimulated EEG traces and, subsequently, was applied to the real EEG data. Moreover, the results obtained on the semisimulated data allow to highlight the effectiveness of the automatic optimization algorithm in choosing the best ASR parameters. Finally, the best k and wl parameters in terms of the mean of all values obtained from the 20 iterations of the procedure were used on the real dataset, showing a significant improvement in ASR correction over that made with the default parameters. The procedure shows the same encouraging results at both eight, six, and four channels.

A. Results on Semisimulated Data

As mentioned in Section III, for each configuration of the 20 semisimulated datasets, the optimal pair of parameters in terms of k and wl was found through the use of the automatic choice procedure described in the second step of the proposed method 2) (see Fig. 1). The use of such a function based on different types of metrics allows identifying the k and wl that ensure the best correction. As a matter of fact, to assess artifact removal on each semisimulated EEG trace, RMSE (2), Gamma value (3) and correlation coefficient (4) were chosen (see Section III-A). These metrics highlight the difference between the contaminated signal EEG_{cont} and the corrected signal EEG_{corr} after artifact removal.

TABLE III
Metrics Values for Default Parameter and Best Parameters of a Single Subject

	Muscolar artifacts							
	Default parameters				Best parameters			
# Ch	RMSE	Gamma	Corr	MAE (PSD)	RMSE	Gamma	Corr	MAE (PSD)
8	4.01	22.49	0.83	14.92	3.09	24.77	0.89	11.88
6	4.61	21.27	0.79	16.47	3.45	23.70	0.87	11.76
4	4.20	22.1	0.80	13.89	3.59	23.44	0.85	11.28
	Ocular artifacts							
	Default parameters				Best parameters			
# Ch	RMSE	Gamma	Corr	MAE (PSD)	RMSE	Gamma	Corr	MAE (PSD)
8	3.55	16.54	0.86	8.37	2.93	18.2	0.89	7.66
6	3.99	14.79	0.84	10.93	3.15	16.87	0.88	8.39
4	4.04	16.21	0.81	9.86	3.58	17.26	0.83	8.98
	Total EEG trace							
	Default parameters				Best parameters			
# Ch	RMSE	Gamma	Corr	MAE (PSD)	RMSE	Gamma	Corr	MAE (PSD)
8	2.68	20.76	0.92	7.68	2.13	22.76	0.95	6.51
6	3.05	19.52	0.90	8.99	2.36	21.75	0.94	6.82
4	2.91	20.18	0.90	7.98	2.54	21.38	0.92	6.98

Table III shows the metric values both in the case of the corrected dataset with the default parameters ($k = 20$ and $\text{wl} = 0.5$ s) and in the case of the best k and wl chosen by the automatic function. To be brief, the table only includes one of the semisimulated EEG traces. A clear improvement in terms of metrics can be noted, since the RMSE decreases ever more to 0, the correlation coefficient increases reaching nearly 1, and also the γ value increases. The efficiency of correction with the best parameters is also demonstrated by calculating mean absolute error (MAE) in PSD, which is a useful metric for analyzing information loss after signal correction [43]. Moreover, this metric measures the average magnitude of the errors in the original signal PSD and the corrected signal PSD. Therefore, in the current work, MAE-PSD values between EEG_{true} and the EEG_{corr} with the two parameter sets, default and best, were calculated. As shown in Table III, MAE-PSD values are lower for the correct signal with the best parameters.

Instead, for an overview, the boxplots of the three metrics on the total of the semisimulated traces are shown as the number of channels decreases in Fig. 2. As can be seen, when the signal is corrected with the default parameters, the

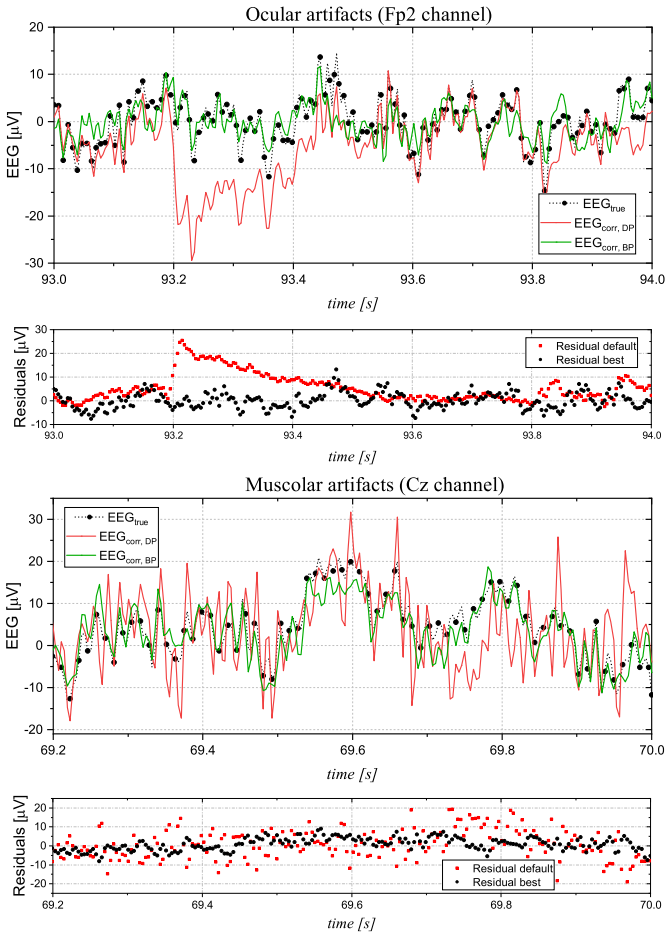


Fig. 3. Comparison between EEG_{true} (black), EEG_{corr} with default parameters (red), and EEG_{corr} with best parameter (green) traces at eight channels. Residual panels (the difference between EEG_{true} and EEG_{corr}) for each figure are reported.

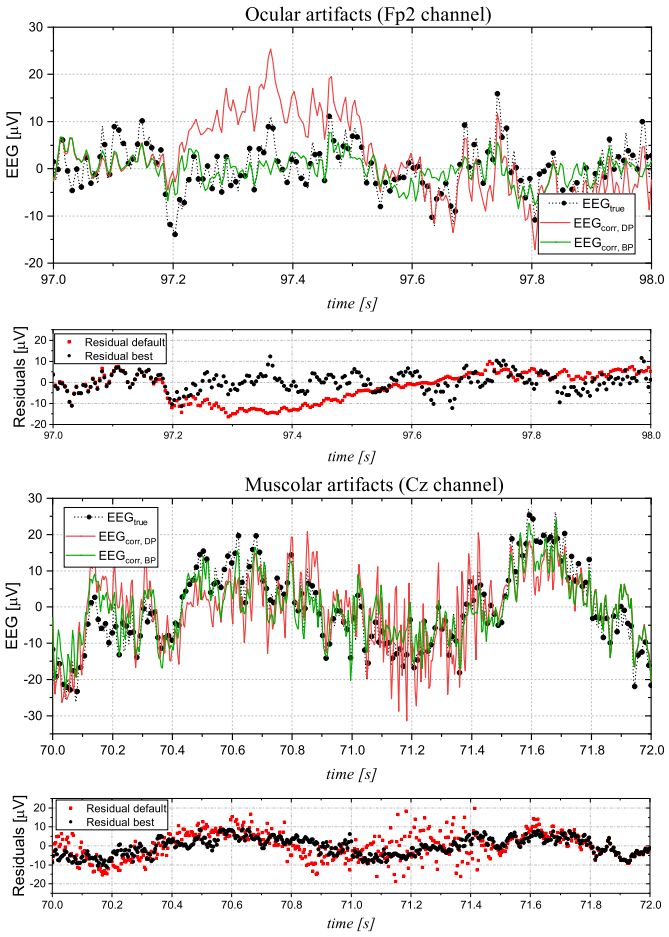


Fig. 4. Comparison between EEG_{true} (black), EEG_{corr} with default parameters (red), and EEG_{corr} with best parameter (green) traces at six channels. Residual panels (the difference between EEG_{true} and EEG_{corr}) for each figure are reported.

RMSE reaches a higher median value and a larger standard deviation. On the contrary, when the signal is corrected with the best parameters, the RMSE has lower median values and smaller standard deviations, approaching the ideal value of zero. Similar considerations can be made for gamma and cross correlation. Considering both metrics, higher median values and better standard deviation ranges are observed in the case of the signal corrected with the best parameters.

The results obtained in terms of metrics and automatic choice of the best k and wl are reflected directly in the EEG trace by comparing EEG_{true} with both the dataset corrected with the default parameters $EEG_{corr,DP}$ and the dataset corrected with the best parameters $EEG_{corr,\,BP}$. In particular, Fig. 3 represents the same simulated subject of Table III at eight channels. For the sake of clarity, out of the eight channels, only the most significant in terms of effective removal of artifacts are reported. In particular, the Fp2 channel was chosen for ocular artifacts and the Cz and $C3$ channel for muscle artifacts.

It is evident in Fig. 3 that $EEG_{corr,DP}$ is still compromised by artifacts. In particular, there are typical peaks of ocular artifacts for the Fp2 channel and the presence of muscle artifacts characterized by a high-frequency trend for the Cz channel. On the other hand, $EEG_{corr,BP}$ better follows the trend of the original signal preserving its shape. Finally, Figs. 4 and 5 show the same results for the six- and four-channel cases, respectively. It is clear that, even if the number of channels decreases, the algorithm works well. The results highlight how the use of the optimization algorithm allows for a much more efficient correction of the EEG data: this is evident from the analysis of the metrics and the graphs.

TABLE IV
BEST k AND WL PARAMETERS FOR EACH CHANNEL IN TERMS OF MEAN AND STANDARD DEVIATION FOR REAL DATA

	8 channels	6 channels	4 channels
k [adim]	11.75 (4.29)	10.20 (3.59)	9.75 (2.57)
wl [s]	1.58 (0.45)	1.47 (0.55)	1.61 (0.37)

B. Application on Real Data

Finally, in order to improve the process of artifact correction and removal on real data, the optimal range for the ASR k and wl parameters was calculated based on the 20 semisimulated datasets. As described in the third step of proposed method 3) (see Fig. 1), the optimal range was found in terms of mean and standard deviation for each channel, as reported in Table IV. However, on real data, it is not possible to use a metric that quantitatively defines the improvement in artifact correction, since in this case, EEG_{true} is unknown. Therefore, it is not

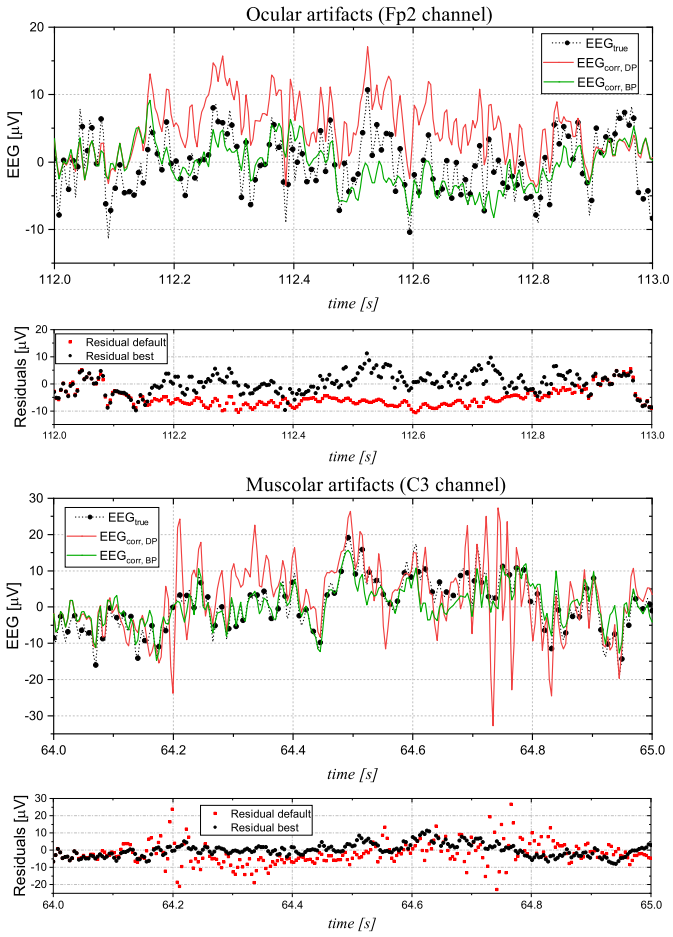


Fig. 5. Comparison between EEG_{true} (black), EEG_{corr} with default parameters (red), and EEG_{corr} with best parameter (green) traces at four channels. Residual panels (the difference between EEG_{true} and EEG_{corr}) for each figure are reported.

possible to calculate the differences between the true clean signal and the corrected signal. As a matter of fact, there is a lack of consensus between researchers on the evaluation criterion for the applied artifact removal techniques. In the recent literature, some assessment processes have been proposed but none of them still outperforms visual inspection [17]. In fact, visual inspection by an expert operator is still considered the best criterion for evaluating artifact removal techniques on real EEG data.

On such basis, Fig. 6 shows the comparison between the original contaminated signal EEG_{cont} and the two corrected signals with the default parameters $EEG_{corr,DP}$ and with the mean of best parameters $EEG_{corr,BP}$ of Table IV. In particular, for the sake of brevity, the results obtained on only one trace of real EEG signal and only the four-channel low-density case are shown, since it represents the most critical situation. As a matter of fact, in Fig. 6, $EEG_{corr,DP}$, in some time intervals, follows the trend of EEG_{cont}, not performing artifact removal. Indeed, $EEG_{corr,BP}$ shows a clear removal of the remaining artifacts. This signal has amplitude values between -10 and 10 μV, an optimum range for EEG signals amplitudes. However, this is not always evident in Fig. 6 for scaling reasons, having to visualize on the same graph the very pronounced artifact. The clear removal of artifacts can

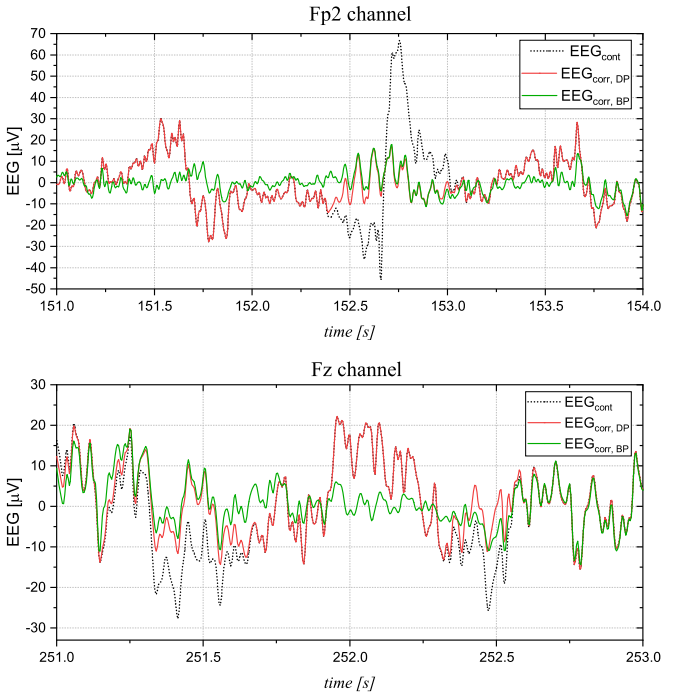


Fig. 6. Comparison between EEG_{cont} (black), EEG_{corr} with default parameters (red), and EEG_{corr} with best parameter (green) traces at four channels for real data EEG.

be seen especially in the range 151–152 and 152.5–153 s of the Fp2 channel and in a small range centered on 251.5 and 252 s of the Fz channel. In addition, it is worth noting how $EEG_{corr,BP}$ follows the same trend in terms of shape as the original signal: this means that there was no overcorrection or loss of information. Finally, artifact correction was also tested using the upper and lower extremes of the optimal range of Table IV for k and wl. As a matter of fact, the lower value and the upper value lead to an overcorrection and an undercorrection of some artifacts, respectively. For this reason, mean values of k and wl allow for better and more balanced results. However, in practical cases, the operator can change the parameters within the proposed range according to specific needs.

VI. Conclusion

In this work, the performance of the ASR for the removal of artifacts in a low-density scenario was investigated. This is a promising artifact-removal technique that has been used in a variety of applications but has received little attention in the literature for low-density EEGs. The goal of this research was to develop an automatic algorithm to optimize the choice of the ASR's two main user-selected parameters, namely the cutoff threshold parameter k and the sliding window length wl. Starting from semisimulated EEG data based on the actual data at hand, the ASR was applied for a number of channels from eight to four. A range from 5 to 30 was explored for k, while wl was varied between 0.2 and 2.0 s. Three different metrics, RMSE, γ, and cross correlation, were calculated on different signal segments to assess artifact removal in order to identify the best values of these parameters. For statistical significance, the algorithm was repeated on several semisimulated traces.

In each iteration, the value of k and the value of wl most voted by the considered metrics were selected. Visual inspection of signals confirmed the better performance of the ASR with optimized parameter values compared to default values.

Therefore, the results of this optimization were applied to real EEG data using the average values among the best k and wl. The resulting signal was then compared to the signal corrected using the default ASR configuration. Also, in this case, an improvement in the removal of the ocular and muscle artifacts was observed through visual inspection. In conclusion, ASR proved to be a powerful and automatic method for removing artifacts in low-density EEG signals, which can favor the successful employment of the EEG also for portable, practical applications. As a result, the proposed technique has the great advantage of identifying the fundamental parameters of ASR to be used for a good artifact removal process on real data. In this way, artifacts can be removed more effectively than with the use of the default parameters. However, it is necessary to simulate data that is as close to the real signals as possible, in order to find the best parameters for the algorithm applied to them. As mentioned above, this is related to a lack of knowledge of the original pure real signal, making calculating comparison metrics between pure and corrected signals not feasible. In this regard, future work will be dedicated to overcoming this limitation by developing a quantitative metric for selecting the best parameters for removal based on real data. This, in fact, in addition to visual inspection, could enhance parameters optimization, with the purpose of embedding ASR as an online processing technique in portable systems.

ACKNOWLEDGMENT

The authors would like to thank Prof. Pasquale Arpaia for his insightful suggestion and theoretical support in the work.

The EEG data used for the generation of the simulated data were acquired by the DREAM Neuroscience Laboratory, Vito Fazzi Hospital, Lecce, Italy.

REFERENCES

[1] J. C. Henry, "Electroencephalography: Basic principles, clinical applications, and related fields," *Neurology*, vol. 67, no. 11, p. 2092, 2006.

[2] P. Arpaia, A. Cataldo, S. Criscuolo, E. De Benedetto, A. Masciullo, and R. Schiavoni, "Assessment and scientific progresses in the analysis of olfactory evoked potentials," *Bioengineering*, vol. 9, no. 6, p. 252, 2022.

[3] S. Saha *et al.*, "Progress in brain computer interface: Challenges and opportunities," *Frontiers Syst. Neurosci.*, vol. 15, Feb. 2021, Art. no. 578875.

[4] Y. Zhang, S. Q. Xie, H. Wang, and Z. Zhang, "Data analytics in steady-state visual evoked potential-based brain–computer interface: A review," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1124–1138, Jan. 2020.

[5] T. Shi, L. Ren, and W. Cui, "Feature extraction of brain–computer interface electroencephalogram based on motor imagery," *IEEE Sensors J.*, vol. 20, no. 20, pp. 11787–11794, Oct. 2020.

[6] P. Arpaia, E. De Benedetto, L. De Paolis, G. D'Errico, N. Donato, and L. Duraccio, "Performance enhancement of wearable instrumentation for AR-based SSVEP BCI," *Measurement*, vol. 196, Jun. 2022, Art. no. 111188.

[7] A. Apicella *et al.*, "Enhancement of SSVEPs classification in BCI-based wearable instrumentation through machine learning techniques," *IEEE Sensors J.*, vol. 22, no. 9, pp. 9087–9094, May 2022.

[8] G. V. S. Karthik, S. Y. Fathima, M. Z. U. Rahman, S. R. Ahamed, and A. Lay-Ekuakille, "Efficient signal conditioning techniques for brain activity in remote health monitoring network," *IEEE Sensors J.*, vol. 13, no. 9, pp. 3276–3283, Sep. 2013.

[9] M. Z. I. Ahmed, N. Sinha, S. Phadikar, and E. Ghaderpour, "Automated feature extraction on AsMap for emotion classification using EEG," *Sensors*, vol. 22, no. 6, p. 2346, 2022.

[10] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals—Practices and possibilities," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 488–500, May 2012.

[11] J. Urigüen and B. Garcia-Zapirain, "EEG artifact removal-state-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, 2015, Art. no. 031001.

[12] W. Tatum, A. Husai, S. Benbadis, and P. Kaplan, "Normal EEG," in *Handbook of EEG Interpretation*. New York, NY, USA: Demos Medical Publishing, 2008, pp. 1–50.

[13] R. Ranjan, B. C. Sahana, and A. K. Bhandari, "Ocular artifact elimination from electroencephalography signals: A systematic review," *Biocybern. Biomed. Eng.*, vol. 41, no. 3, pp. 960–996, 2021.

[14] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–494, 2007.

[15] X. Chen *et al.*, "Removal of muscle artifacts from the EEG: A review and recommendations," *IEEE Sensors J.*, vol. 19, no. 14, pp. 5353–5368, Jul. 2019.

[16] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, "Trends in EEG-BCI for daily-life: Requirements for artifact removal," *Biomed. Signal Process. Control*, vol. 31, pp. 407–418, Jan. 2017.

[17] W. Mumtaz, S. Rasheed, and A. Irfan, "Review of challenges associated with the EEG artifact removal methods," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102741.

[18] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: A review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[19] M. K. Islam, A. Rastegarnia, and Z. Yang, "Methods for artifact detection and removal from scalp EEG: A review," *Clin. Neurophysiol.*, vol. 46, nos. 4–5, pp. 287–305, 2016.

[20] T. P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.

[21] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task," *Sensors*, vol. 15, no. 11, pp. 29015–29035, 2015.

[22] C. Chang, S. Hsu, L. Pion-Tonachini, and T. Jung, "Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1114–1121, Apr. 2020.

[23] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward, "Removing muscle artifacts from EEG data: Multichannel or single-channel techniques?" *IEEE Sensors J.*, vol. 16, no. 7, pp. 1986–1997, Apr. 2016.

[24] T. R. Mullen *et al.*, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov. 2015.

[25] C. A. E. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," U.S. Patent 14 895 440, Apr. 28, 2016.

[26] M. Plechawska-Wojcik, M. Kaczorowska, and D. Zapala, "The artifact subspace reconstruction (ASR) for EEG signal correction. A comparative study," in *Proc. Int. Conf. Inf. Syst. Archit. Technol.* Cham, Switzerland: Springer, 2018, pp. 125–135.

[27] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung, "Evaluation of artifact subspace reconstruction for automatic EEG artifact removal," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1242–1245.

[28] B.-Y. Tsai, S. V. S. Diddi, L.-W. Ko, S.-J. Wang, C.-Y. Chang, and T.-P. Jung, "Development of an adaptive artifact subspace reconstruction based on Hebbian/anti-Hebbian learning networks for enhancing BCI performance," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 17, 2022, doi: 10.1109/TNNLS.2022.3174528.

[29] P. Arpaia, E. De Bendetto, A. Esposito, A. Natalizio, M. Parvis, and M. Pesola, "Comparing artifact removal techniques for daily-life electroencephalography with few channels," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Taormina, Italy, Jun. 2022, pp. 1–6.

[30] V. P. Kumaravel, V. Kartsch, S. Benatti, G. Vallortigara, E. Farella, and M. Buiatti, "Efficient artifact removal from low-density wearable EEG using artifacts subspace reconstruction," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 333–336.

[31] M. Saini, U. Satija, and M. D. Upadhayay, "Wavelet based waveform distortion measures for assessment of denoised EEG quality with reference to noise-free EEG signal," *IEEE Signal Process. Lett.*, vol. 27, pp. 1260–1264, 2020.

[32] M. M. N. Mannan, M. A. Kamran, and M. Y. Jeong, "Identification and removal of physiological artifacts from electroencephalogram signals: A review," *IEEE Access*, vol. 6, pp. 30630–30652, 2018.
[33] S. Ehrlich. *Dataset Automatic Artifact Removal*. Accessed: Sep. 30, 2022. [Online]. Available: https://github.com/stefan-ehrlich/dataset-automaticArtifactRemoval
[34] *Introducing the ActiCHamp Plus—Offering Active and Passive Electrode Recordings and More!* Accessed: Sep. 30, 2022. [Online]. Available: https://pressrelease.brainproducts.com/actichamp_plus/
[35] N. Yeung, R. Bogacz, C. B. Holroyd, and J. D. Cohen, "Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods," *Psychophysiology*, vol. 41, no. 6, pp. 822–832, Nov. 2004.
[36] *Ncclabsustech/EEGdenoiseNet: EEGdenoiseNet, a Benchmark Dataset, That Is Suited for Training and Testing Deep Learning-Based EEG Denoising Models, as Well as for Comparing the Performance Across Different Models*, 2020.
[37] M. Dora and D. Holcman, "Adaptive single-channel EEG artifact removal with applications to clinical monitoring," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 286–295, 2022.
[38] A. Anzolin, J. Toppi, M. Petti, F. Cincotti, and L. Astolfi, "SEED-G: Simulated EEG data generator for testing connectivity algorithms," *Sensors*, vol. 21, no. 11, p. 3632, 2021.
[39] S. Invitto *et al.*, "Obstructive sleep apnea syndrome and olfactory perception: An OERP study," *Respiratory Physiol. Neurobiol.*, vol. 259, pp. 37–44, Jan. 2019.
[40] J. Górecka and A. Biedka, "Determination of ocular artifacts in the clinical EEG software by a peripheral device," *Electronics*, vol. 10, no. 2, p. 108, 2021.
[41] L. Frølich and I. Dowding, "Removal of muscular artifacts in EEG signals: A comparison of linear decomposition methods," *Brain Informat.*, vol. 5, no. 1, pp. 13–22, 2018.
[42] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[43] A. K. Maddirala and R. A. Shaik, "Removal of EOG artifacts from single channel EEG signals using combined singular spectrum analysis and adaptive noise canceler," *IEEE Sensors J.*, vol. 16, no. 23, pp. 8279–8287, Dec. 2016.

Andrea Cataldo (Senior Member, IEEE) is an Associate Professor of Electric and Electronic Measurements with the University of Salento, Lecce, Italy. Prof. Cataldo is also the Co-Founder of the University spin-off company Monitoring Technologies, Lecce, a spin-off of the University of Salento. He has coauthored more than 200 publications. His research interests include reflectometry and microwave measurement techniques, uncertainty evaluation, characterization and optimization of sensors, and biomedical sensors.

Prof. Cataldo is a member of the IEEE Instrumentation and Measurement (I&M) Society and the Italian Group of Electrical and Electronic Measurements (GMEE).

Sabatina Criscuolo received the M.S. degree in biomedical engineering from the University of Naples Federico II, Naples, Italy, in 2021, where she is currently pursuing the Ph.D. degree in information and communication technology for health.

Her research interests focus on artificial intelligence techniques and systems to support precision medicine.

Egidio De Benedetto (Senior Member, IEEE) received the M.S. degree in materials engineering and the Ph.D. degree in information engineering from the University of Salento, Lecce, Italy, in 2006 and 2010, respectively.

He was with the Institute for Microelectronics and Microsystems, National Research Council, Lecce, from 2010 to 2012. Since 2019, he has been an Associate Professor with the Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy.

Antonio Masciullo received the M.S. degree in electronic engineering from the Polytechnic University of Bari, Bari, Italy, in 1995, and the Ph.D. degree in information engineering from the University of Salento, Lecce, Italy, in 2015.

He has been involved in the design and realization of electronic devices and equipment since 1997. Since 2003, he has been with the Department of Innovation Engineering, University of Salento, where he is currently a Laboratory Technician with Telecommunications, Automatic Control, and Electronic Measurement Laboratories.

Marisa Pesola received the M.S. degree in biomedical engineering from the University of Naples Federico II, Naples, Italy, in 2022.

She is currently a Research Fellow with the Center for Advanced Metrological and Technological Services (CeSMA), University of Naples Federico II. Her research interests focus on the processing of biomedical signals and artifact removal.

Raissa Schiavoni received the M.S. degree in communication engineering and electronic technologies from the University of Salento, Lecce, Italy, in 2020, where she is currently pursuing the Ph.D. degree in complex systems engineering.

Her research interests focus on microwave reflectometry systems for monitoring applications.

Sara Invitto is an Associate Professor of General and Experimental Psychology with the University of Salento, Lecce, Italy.

Open Access funding provided by 'Università degli Studi di Napoli "Federico II"' within the CRUI CARE Agreement