2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Treating Electrical and Biopotential Artifacts in an EEG Pilot Study Experiment*

Irina E. Nicolae, Alina E. Sultana, Ruxandra Aursulesei and Szabolcs Fulop

Abstract— **With the increase in life expectancy, as well as in the performance and complexity of healthcare systems, the need for fast and accurate information has also grown. EEG devices have become more accessible and necessary in clinical practice. In daily activity, artifacts are ubiquitous in EEG signals. They arise from: environmental, experimental and physiological factors, degrade signal quality and render the affected part of the signal useless. This paper proposes an artifact cleaning pipeline including filters and algorithms to streamline the analysis process. Moreover, to better characterize and discriminate artifacts from useful EEG data, additional physiological signals and video data are used, which are correlated with subject's behavior. We quantify the performance reached by Peak Signal-to-Noise Ratio and clinical visual inspection. The entire research and data collection took place in the laboratories of XPERI Corporation.**

***Clinical Relevance*—Since the occurrence of artifacts cannot be controlled, it is essential to have a precise process of recognition, identification and elimination of noise. Therefore, it is important to distinguish EEG artifacts from abnormal activity in order to minimize the chance of EEG misinterpretation, that can lead to false diagnosis, especially regarding the study of epileptiform activities or other neurologic or psychiatric disorders (e.g. degenerative diseases, dementia, depression, sleep disorders, Alzheimer's disease, schizophrenia, etc.).**

I. Introduction

The electroencephalogram (EEG) records the electrical activity of the brain measured on the scalp. An EEG experimental study consists in collecting EEG signals from several subjects in a laboratory controlled or free outdoor environment, while taking into account the significant inter- and intra-subject variability. The goal is to detect similarities and differences in brain activity responses for the working hypothesis in question. Although the EEG waveforms are designed to reveal the electrical activity of the cortex, they also record the electrical activities arising from sites other than the brain. Whereas, any activity that does not involve cortical origin is considered an artifact. Depending on the source that generates the artifacts, two classes can be defined: physiological artifacts and external artifacts [1]. Physiological artifacts are generated by the human body from sources other than the brain, such as: heart, muscle and eye artifacts. External artifacts come from outside of the body, as e.g. equipment and external sources.

The recognition of artifacts, the detection of the source that generates them and their elimination is a highly important process in the processing and analysis of EEG signals. The purpose of removing artifacts is to reduce the chances of misclassification and misinterpretation of brain activity. Since EEG activity is quite small, measured in microvolts (μV), one major challenge of EEG signal analysis is to detect and consistently remove these types of non-cerebral signals.

This paper aims to remove different types of artifacts across subjects and testing scenarios using a simple to use pipeline to improve the EEG data cleaning process in terms of processing time and data quality performance. This preliminary step of pre-processing the EEG data is envisioned to be used for the analysis of a large-scale collection of EEG data, being a useful tool for clinical applications and brain computer interfaces (BCI). Therefore, the main purpose of this study is to provide a standardized, well defined pre-processing methodology that minimizes the impact of different types of signal artifacts regardless the subject's EEG variability, scenarios and stimuli applied. The types of artifacts we address are both physiological (eye and eyelids movements, muscle, cardiac) and external artifacts (power line, electrode pop and physical subject's movement), which are the commonly observed in EEG studies.

II. Related works

In the literature, various methods applicable to EEG data cleaning are reported, but the artifact removal step remains an open problem for the EEG data processing. To our knowledge, there is no global standard used to remove EEG artifacts. Existing methods for the mitigation of EEG signal artifacts are divided in two main approaches: single artifacts removal approach and hybrid approach – where two or more methods are combined. Among the single artifact removal techniques, the following are noteworthy [2, 3]: 1) regression techniques and 2) Wavelet transform, used to filter the ocular artifacts; 3) Blind Source Separation (BSS): which uses a mixture of unsupervised learning methods with no prior information and extra reference channels; 4) Empirical mode decomposition (EMD) for removing muscle artifacts; and 5) Different filtering methods, e.g.: adaptive, Wiener and Bayes filtering. Among the hybrid approach, we note: 1) EMD – BSS, a good approach for muscle artifact removal under few-channel scenarios; 2) Wavelet – Independent Component Analysis (ICA) technique proposed to avoid shortcomings of ICA; 3) and BSS-SVM: when the recorded EEG data is decomposed

I.E. Nicolae is with Xperi Corporation (external collaborator) and with University Transilvania of Brasov, Romania; R. Aursulesei is a neurologist clinician associated with Hospital of Psychiatry, "Prof. Dr. Al. Obregia", Bucharest, Romania and external collaborator with XPERI. A.E. Sultana is with Xperi Corporation and with University Politehnica of Bucharest, Romania; and S. Fulop is with XPERI.

*Research supported by XPERI and National Council of Scientific Research (CNCS) (in part, via project PN-III-P1-1.1-PD-2019-0971 as with author I.E. Nicolae). CNCS funder had no role in the design of the study, collection, analysis, interpretation of the data, writing of the manuscript or in the decision to publish the results.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

into multiple components and several of them are used as input to a set of SVM classifiers.

The remainder of this paper is structured as follows: section III presents the study methodology, section IV describes the experiment, section V the pre-processing methodology, while section VI shows the pipeline methodology applied on different artifacts with the performance evaluation given by the reached Peak Signal-to-Noise Ratio (PSNR) rates, complemented with discussions and conclusion in section VII.

III. Methodology

Artifact removal is a tedious task as the probability of their occurrence is difficult to model in practice. The current study aims to provide a sequence of standard methods to clean contaminated EEG data. We address the artifact removal using a BSS algorithm to clean the biopotential artifacts, to avoid the computational cost limits of hybrid methods and apply a filtering pipeline to remove the device related noise. Among the BSS algorithms, an ICA with Multiple Artifact Rejection Algorithm (MARA) approach [4] is applied. Signal sources are decomposed into linear fusion of cerebral and artifactual sources, as independent components (ICs) [5] and the clean signal is reconstructed by discarding the artifactual ICs. From the experience of previous EEG data analysis studies [6], we propose MARA as an efficient method for the removal of the above-mentioned artifacts and, in addition, we complement the ICs selection with the classification results of the ICLabel algorithm [7].

To control physiological artifacts, we added additional sensors to our study: electrooculogram, (EOG), electrocardiogram (ECG), electromyogram (EMG) and photoplethysmogram (PPG), since the physiological signals are synchronized with the EEG data, the source of the artifacts can be precisely identified. Their use brings several advantages to the study: 1) the number and quality of recording sensors is increased, which is a requirement in most BSS algorithms that there should be as many measurement sensors as underlying sources [8], and 2) the additional sensors complement the EEG signals to describe the activities investigated within the EEG studies. To assess the performance of the proposed approach we have used both PSNR metric and clinician validation using the visual inspection approach. The two approaches were performed independently and cross-validated.

IV. EEG Pilot Study Experiment

A. Equipment

The EEG data was acquired by a Neuron-Spectrum 5 EEG device with 21 electrodes (10-20 positioning), plus 5 additional physiological signals: vertical EOGv, horizontal EOGh, bilateral masseter maxillary muscles EMG, ECG and PPG. Subject's activity was recorded with a 4k resolution camera, synchronized with the EEG signals. The selection of EEG montages plays an important role in better capturing the neural activity. To generalize the analysis, we used referential and bipolar montages: Double Banana 21 (db21) and Monopolar 21 (m21, ear lobes references), the most common referential and bipolar montages referred in the literature [9].

To cover a wide range of artifacts, we included various visual and auditory stimuli in the scenarios, played with a 4k resolution LCD and standard speakers.

B. Participants and experimental scenario

We collected EEG and video data from 10 participants (7 males and 3 females) with an average age of 24 years old. All participants are employees of Xperi Corporation and received a priori information on the experiment. They expressed their consent to take part in the non-invasive experiment and their permission for brain signals recording. The data was completely anonymized. To qualitatively ensure the acquisition process, the neurologist expert validated the entire acquisition protocol. The possibility of any neurological disorder has been excluded and the Helsinki Declarations principles have been considered. The study took place in a laboratory environment, with normal environmental noise, as 30 dB and low electro-magnetic interferences rates (E = 20 V/m and M = 50 nT). To mimic EEG studies, participants achieved the experiments seated and were asked to relax, focus and reduce as possible extra body and eye movements. Each subject followed five scenarios with visual and/or auditory stimuli (1-3 min each): 1) relaxed mode (closed eyes); 2) relaxed mode (open eyes); 3) focused mode – cognitive matching exercise, with two images side by side (open eyes); 4) auditory cognitive task providing answers out loud (closed eyes); and 5) a game on the phone (open eyes).

V. EEG Pre-processing Methodology

In practice, the artifacts are differentiated from EEG signals based on physiological activity principles as the neural activity has a logical topographic field of distribution with an expected fall-off of voltage potentials, while artifacts have no logical distribution that defies the principles of localization.

A. Filtering and artifacts removal

The EEG pre-processing was performed by a chain of filters: low-pass filters for anti-aliasing, spatial filters for artifacts removal and high-pass filters for electrical current drifts removal. We are interested in keeping a wider range of frequencies in order to envision the possibility to capture different neural activities, so we keep all the usual rhythms from delta to gamma (0.5 - 75 Hz). For capturing even lower rhythms or higher gamma activity, the frequency interval limits can be easily extended within filtering. The proposed pipeline includes, as follows:

1) **high-pass filtering:** 0.5 Hz FIR filter, using least-squares error minimization and reverse with zero-phase effect digital filtering, as in [8].

2) **50 Hz Power Line cleaning** with Cleanline algorithm [10]; Even though ICA has the capacity to clean power line noise, we applied this step before, to have a cleaner signal from electrical interferences and therefore aiming to improve sources detection. This step was applied three times to efficiently reduce even strong power line interferences (up to 50dB), repetition which was set after thorough investigation.

3**) ICA-MARA algorithms**: ICA Infomax Algorithm has been applied twice to improve components detection, since ICA uses random weights on first iteration for components detection, while in the second iteration, it better estimates the data after what it has learned on the first step. The automatic

Multiple Artifact Rejection Algorithm (MARA) [4] was used to investigate and select the components, in conjunction with the ICLabel algorithm [7]. We have chosen MARA selection as it is not limited to a specific type of artifact, and as section VI shows, it is able to handle various types of artifacts equally well. It classifies artifactual components ("reject vs. accept") by extracting six EEG signal features from the spatial, spectral and temporal domains. The ICLabel algorithm categorizes the components via crowd labeling latent Dirichlet allocation (CL-LDA), considering composition over classes rather than indicating a discrete class and uses variance measures on fractions of the compositions as confidence. The final artifactual components were selected with clinician supervision, considering both methods.

4) **low-pass filtering:** 75Hz Chebyshev type II, order 10 (75Hz pass-band edge frequency, 3 dB ripple and 82 Hz stopband, 50 dB attenuation).

The external physiological signals were filtered via acquisition software: ECG: 0.5-75Hz band-pass with 50Hz notch, EMG 10-100Hz with 50Hz notch and EOG: 0.5-15Hz.

VI. Pre-processing Results

For signal processing and clinical investigations of the artifacts, we present the results of four representative subjects, chosen to have a variety number of artifact types with frequent occurrence. The pre-processing pipeline successfully detected and cleaned various types of artifacts as can be seen in the followings. Fig. 1 shows an artifactual Independent Component (Artif. IC) example as a mix of ECG and PPG activity, as seen in the time evolutions of the signals (Fig. 1.B and Fig. 1.C). In addition, the scalp map topography shows pronounced activity outside the scalp indicating an external source of noise (Fig. 1A), complemented by power spectral density (Fig. 1D) showing strong power in the frequencies above 10Hz (specifically, the curve does not follow the expected 1/f shape). The component was classified with 72% as ‘other’ by ICLabel and 31% by MARA. This example indicates a case when MARA classification can be supplemented by ICLabel. Between MARA and ICLabel, MARA obtained about 2-3 false negatives and ICLabel around 5-7. This happened when the components were a mix of different sources and for these cases, we decided by comparing the components characteristics. While ICLabel tends to detect better brain and heart artifacts better, MARA is more successful at detecting localized artifacts, such as channel noise, as observed in our selection – but a thorough investigation should be performed to draw accurate conclusions.

The data quality performance after the cleaning pipeline can be seen in Table 1, in relation to the mean PSNR values over all channels (using as ground truth a cleaned sample, filtered within 0.5-75 Hz with removed power line interferences, having no artifacts as detected by clinician validation). The signal noise is improved with about 60% on average as detected by PSNR. More than 9 Artif. ICs are detected; note that at least 3 ICs should be detected as artifacts: subtle horizontal and vertical eye movements and heart activity, even with closed eyes in the relaxed case. Subject S10 was anxious and moved additionally; hence, more artifactual components were found, even in the relaxed scenarios (1, 2).

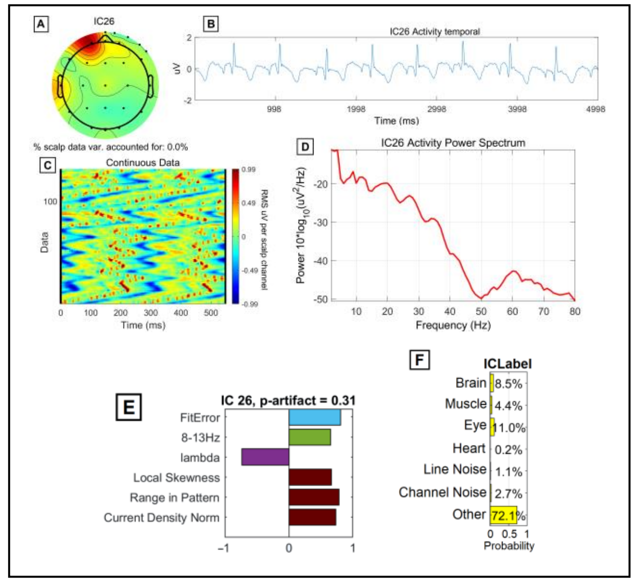

Figure 1. Heartbeat mixed with pulse ICA component and probabilities as determined by MARA and ICLabel algorithms.

TABLE I. Artifactual Cleaning - PSNR Improvement

Montage: db21	No. artif. ICs			PSNR (%)		
Subjects	**S1**	**S4**		**S1**	**S4**	
scenario1	9	-		54.2	-	
scenario2	11	-		64.8	-	
scenario3	17	-		72.6	-	
scenario4	-	21		-	75.6	
scenario5	22	-		74.5	-	
Montage: m21	**No. artif. ICs**			**PSNR (%)**		
Subjects	**S1**	**S9**	**S10**	**S1**	**S9**	**S10**
scenario1	11	12	17	40.8	27.8	59.1
scenario2	16	15	16	62.2	47.2	56.2
scenario3	18	19	20	62.7	48.8	67.4
scenario5	19	19	24	56.9	43.2	64.3

The precise detection of the artifactual external sources, like ocular, muscle, heart artifacts, as perfectly shown in Fig. 1 for heart influences, was obtained by aiding from the external physiological channels in the composition of ICA, as can be seen in Fig. 2, when comparing between the information included within ICA. The ocular movement is cleaned when using ICA-MARA+ (all extra channels) and still present when using ICA-MARA only, the heart activity such as pulse and heartbeat is likewise suppressed when adding ECG. As for the muscle noise, at t = 49-52s, influences remain when using ICA-MARA without EMG.

Fig. 3 shows the effective pipeline on eye saccades (horizontal and vertical movements with fixations), lateral eye movements, blinks and heartbeat influences. Using the EMG to detect the talking movements helped clearly separating the artifactual source, as seen in Fig. 4.

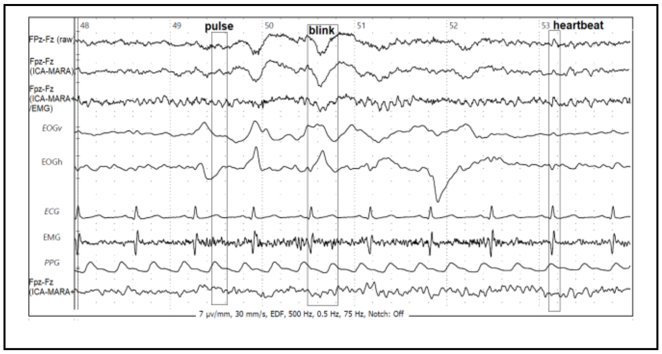

Figure 2. Blink and heart artifacts cleaning. From top to bottom: raw data; ICA-MARA, ICA-MARA+ (no EMG), EOGv, EOGh, ECG, EMG, PPG and ICA–MARA+; Subject S4, scenario 4, Fpz-Fz.

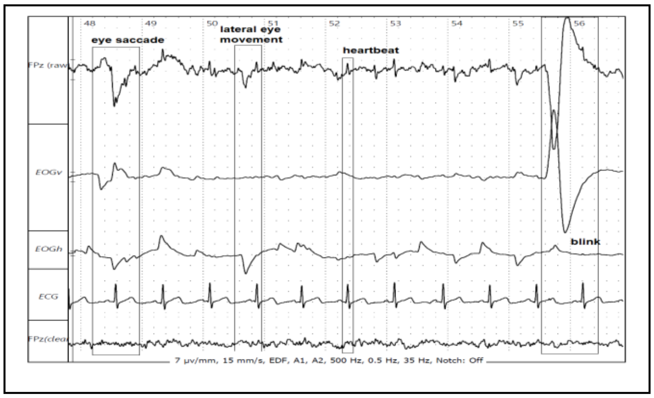

Figure 3. Pipeline cleaning for eye movements and heartbeat artifacts. From top to bottom: FPz raw data, EOGv, EOGh, ECG, FPz cleaned using ICA-MARA+. Subject S1, scenario 3, FPz.

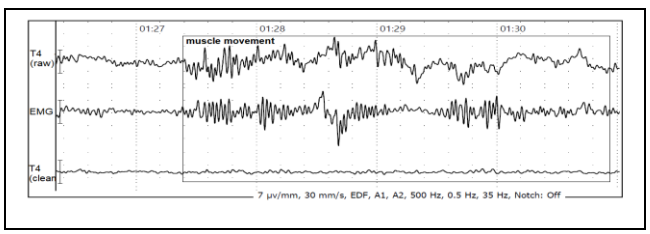

Figure 4. Masseter muscle artifact cleaning. From top to bottom: raw data (T4), EMG, ICA-MARA+ (T4). Subject S9, in scenario 3.

Non-physiological electronic noise, as e.g. produced by channel pop in Fig. 5 are also cleaned successfully. For double banana montage, the artifacts tend to propagate more from the frontal EEG channels towards the back channels in the scalp due to the derivations' computation, as seen here for T4-T6, T6-O2 derivations. Considering powerline, there are cases when CleanLine does not suppress the line noise sufficiently, and some power line noise of about 5-10 dB remain, but ICA additionally picks up and removes the remaining interferences. Further, we mention an example of pipeline steps PSNR improvement for subject S4: 1Hz 2%, 50Hz 1.6%, ICA-MARA 66%, ICA-MARA+ (no EMG) 23.5%, ICA-MARA+ 2.3%, ICA-MARA+ 75Hz 2%. For this case, adding EMG does not improve much (only 2.3%) since the task does not involve serious body muscle movements, but for example for S1, scenario5, adding EMG improves PSNR with 19.5%.

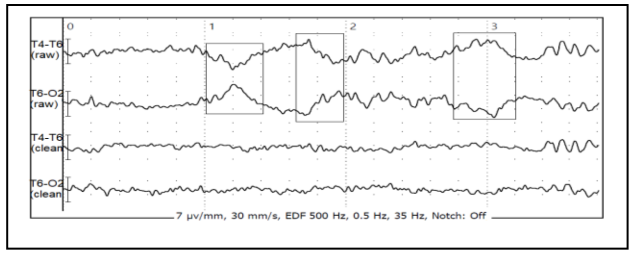

Figure 5. Channel artifacts cleaning. From top to bottom: raw data (T4-T6), raw data (T6-O2), ICA-MARA+ (T4-T6), ICA-MARA+ (T6-O2). Subject S1, scenario 3, derivations T4-T6, T6-O2.

To ensure the validity of artifactual types, the 4K camera – synchronized with EEG helped validated part of the artifacts, such as significant motion, facial and eye movements, by visual inspection with expert clinician.

VII. Discussion and Conclusion

This paper proposed a pre-processing methodology to minimize the EEG signals artifacts. The study novelty consists in: 1) the inclusion of new additional physiological channels such as EMG, PPG to capture the subject talking movements and its pulse; and 2) the parallel use of two different complementary methods to refine the selection of the artifactual components, benefiting from the advantages of both MARA and ICLabel algorithms. The analysis has been cross validated by two approaches used independently, which showed that: 1) the signal noise is significantly decreased as given by the PSNR metric after applying the cleaning pipeline; and 2) the proposed methodology is efficient without impacting the EEG data relevance, as the clinical inspection of cleaned signals vs. originals proved. The camera inclusion helped precisely characterizing the motion artifacts of subjects' activity. Future work consists in the automation of the proposed sequence without any parameter setting step and manual visual selection, an extended comparison of other removal artifacts methods used in the literature.

References

[1] V. Frederik, F. Luca, et al. "Critical comments on EEG sensor space dynamical connectivity analysis". *Brain Topogr*. 2016:1–12.

[2] K.T. Sweeney, T. E. Ward and S.F. McLoone, "Artifact removal in physiological signals practices and possibilities". *IEEE Trans. Inf. Technol. Biomed.* 2012; 16:488–500.

[3] J.A., B. Urigüen. Garciazapirain EEG artifact removal—State-of-the-art and guidelines. *J. Neural Eng.* 2015;12.

[4] I. Winkler, S. Haufe and M. Tangermann. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behav Brain Funct* 7, 30, 2011.

[5] B. Somers, A. Bertrand. Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis. *J. Neural Eng.* 2016; 13:066008.

[6] I.E. Nicolae, et al. "Assessing the Depth of Cognitive Processing as the Basis for Potential User-State Adaptation." *Front. In Neurosc.* v.11, 548, 2017.

[7] Pion-Tonachini L, Makeig S, Kreutz-Delgado K. Crowd labeling latent Dirichlet allocation. *Knowl Inf Syst.* 2017, 53(3):749-765.

[8] K. T. Sweeney, T. E. Ward and S. F. McLoone, "Artifact Removal in Physiological Signals—Practices and Possibilities," in *IEEE Trans. on Inf. Tech. in Biomedicine*, vol. 16, no. 3, pp. 488-500, 2012.

[9] Sazgar M., Young M.G. (2019) Overview of EEG, Electrode Placement, and Montages. In: Absolute Epilepsy and EEG Rotation Review. *Springer*, Cham.

[10] Mullen T, NTRC CleanLine: 2012 www.nitrc.org/projects/cleanline/