Online Automatic Artifact Rejection using the Real-time EEG Source-mapping Toolbox (REST)

Luca Pion-Tonachini, *Student Member, IEEE,* Sheng-Hsiou Hsu, *Student Member, IEEE,* Chi-Yuan Chang, *Student Member, IEEE,* Tzyy-Ping Jung, *Fellow, IEEE* and Scott Makeig

***Abstract*— Non-brain contributions to electroencephalographic (EEG) signals, often referred to as artifacts, can hamper the analysis of scalp EEG recordings. This is especially true when artifacts have large amplitudes (e.g., movement artifacts), or occur continuously (like eye-movement artifacts). Offline automated pipelines can detect and reduce artifact in EEG data, but no good solution exists for online processing of EEG data in near real time. Here, we propose the combined use of online artifact subspace reconstruction (ASR) to remove large amplitude transients, and online recursive independent component analysis (ORICA) combined with an independent component (IC) classifier to compute, classify, and remove artifact ICs. We demonstrate the efficacy of the proposed pipeline using 2 EEG recordings containing series of (1) movement and muscle artifacts, and (2) cued blinks and saccades. This pipeline is freely available in the Real-time EEG Source-mapping Toolbox (REST) for MATLAB (The Mathworks, Inc.).**

I. Introduction

Brain computer interface (BCI) and other real-time EEG applications often suffer when artifacts (unwanted signals included in a recording) are present [1] [2]. In electroencephalographic (EEG) recordings, typical examples include perturbations induced by the retinal electrical dipoles during eye movements, high-frequency signals from scalp muscle activity, and large signal spikes and shifts from electrode impedance changes during subject movements (Figure 2). While the definition of 'artifact' is largely context dependent, artifacts are typically detrimental to signal analyses as the amplitudes of these artifacts can easily eclipse the brain-generated EEG activity.

Previous methods for cleaning EEG data have taken several approaches: spectral-intensity thresholds [3], filtering based on simultaneous electrooculographic (EOG) recordings [4], complexly stacked wavelets, blind source separation (BSS), and wavelet-based classifiers [5]. Each method comes with its own assumptions and may therefore fail in some situations. For example, none of the above-mentioned methods support real-time processing with the exception of EOG filtering – and that requires a dedicated EOG recording in addition to the EEG electrode montage. To the authors' knowledge, no existing EEG artifact cleaning method effectively cleans the data in near real time without a need for recording EOG or other artifact reference channels.

L. Pion-Tonachini is in the Dept. of Electrical and Computer Engineering (ECE) and the Swartz Center for Computational Neuroscience (SCCN) of University of California San Diego (UCSD).

S.-H. Hsu, C.-Y. Chang, and T.-P. Jung are in the Dept. of Bioengineering (BIOE) and SCCN of UCSD.

S. Makeig directs the SCCN

Independent component analysis (ICA) has been widely used for separating spatially stereotyped artifacts such as saccades and eye-blink activities from EEG data [6]. Online recursive ICA (ORICA) [7] has successfully converted the computationally-expensive ICA algorithm into incremental, recursive update rules that enable online, near real-time ICA decomposition [8]. However, ICA decomposition is sensitive to unique, large-amplitude artifacts which can severely degrade the returned independent components (IC) and ICA-based methods typically require visual inspection to manually identify the artifact-related ICs.

This study proposes to apply artifact subspace reconstruction (ASR) [9], an automated, near real-time-capable algorithm, prior to ORICA. We demonstrate that ASR can effectively remove transient, large-amplitude artifacts from EEG data and thus stabilize the ICA decomposition and improve artifact separation. Next, we implement a real-time capable IC classifier, here EyeCatch [10], to automate recognition and removal of artifact-related ICs. The full online, real-time, automatic artifact rejection (AR) pipeline, featuring ASR, ORICA, and EyeCatch, is available in the open-source Real-time EEG Source-mapping Toolbox (REST) [11] illustrated in Figure 1B. REST can be downloaded from the url: `https://github.com/goodshawn12/REST`.

II. Materials and Methods

A. Review of methods

ASR is an automated, variance-based EEG cleaning algorithm. It uses a short initial recording of artifact-free EEG data from which it learns a statistical model of the EEG data. Then for each incoming data window, ASR applies a principal component analysis (PCA) like linear transformation to the data using a transform matrix learned from the initial calibration data. If any principal component (PC) of the new data window is much larger than in the calibration data, that PC is removed from the window. An inverse transform then projects the data window back into the original channel coordinates.

ORICA consists of 2 stages. The first, online whitening, can be thought of as an online, recursive form of PCA. This is done to facilitate learning in the subsequent stage – online recursive ICA. This optimizes the same objective as offline Infomax ICA. Finally, the ICA solution from the second stage is projected to the nearest orthogonal matrix which further facilitates model convergence.

EyeCatch classifies ICs as either eye movement-related or not by first calculating the maximum correlation value of

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

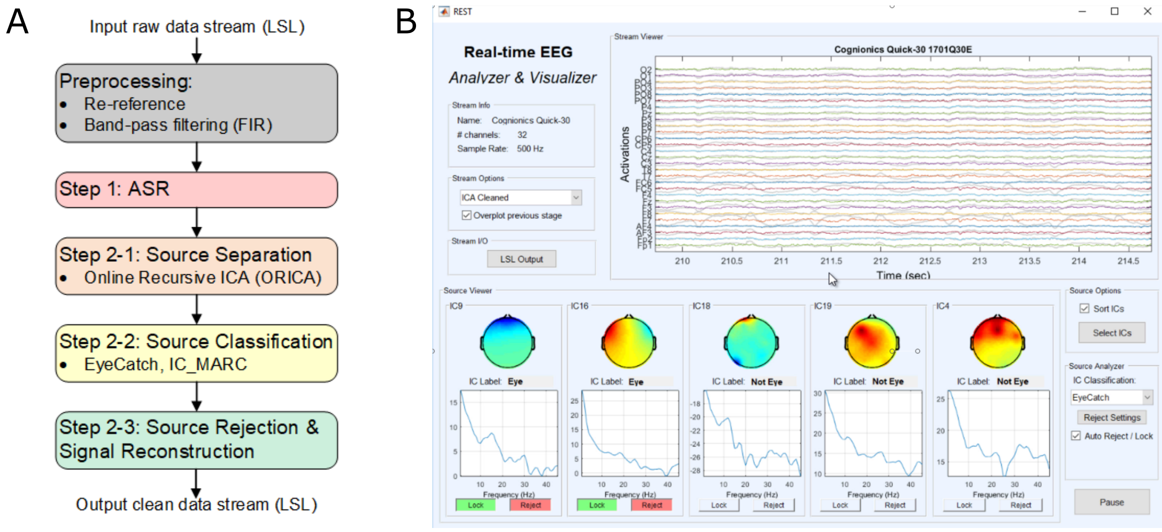

Fig. 1. (A) An overview of the REST data-cleaning pipeline combining ASR, ORICA and an IC classifier. (B) The REST graphic user interface during a period containing repeated eye movements. In the lower left, two eye-related components have been found and removed by EyeCatch. In the upper right, the reconstructed channel signals are shown in color against uncleaned (gray) data traces in which the eye-movement artifacts are still visible.

each IC scalp map to thousands of IC scalp maps in the method's library which account for eye-movement activities. An IC scalp map represents the relative contributions of a given IC to the scalp channels. Any IC for which the maximum correlation value, called the similarity score, is greater than a preset threshold is marked for removal. REST uses a modified version of EyeCatch that is computationally lighter and has a lower rejection threshold to support its use in a real-time setting.

B. Experimental design

To evaluate the proposed AR pipeline in REST, we first collected an eyes-closed EEG dataset in which a healthy subject performed a series of cued artifact-inducing actions comprising jaw clenching, scalp electrode tapping, head turning, and jumping. The subject rested for 2 minutes before performing each type of action for 10 seconds, with 5-second inter-action intervals. Our goal was to evaluate the effects of the concomitant artifacts on ORICA convergence and to determine whether the proposed pipeline could mitigate those effects.

Next, we recorded 2 minutes during which the subject blinked at 1-sec intervals, followed by 2 minutes in which the subject performed lateral eye-movements using voluntary saccades at 1-sec intervals. Before and after each artifact period the subject rested with eyes closed for 2 minutes. We wanted to characterize the performance of the proposed pipeline in automatically identifying and rejecting eye-related ICs, thus clean the recording of eye-movement artifacts.

C. Dataset recording and analysis

The EEG data was recorded using a Cognionics Quick30 headset with a 500-Hz sampling rate using dry electrodes for which electrode impedances are in the range of hundreds of Ohms. Electrode P07 was excluded from the analysis because of a known cap hardware issue.

We processed both datasets using REST in a simulated real-time setting by rebroadcasting the data through the lab-streaming-layer [12]. REST applied the proposed pipeline consisting of common-average re-referencing, FIR bandpass-filtering (1-50 Hz), ASR, ORICA, IC classification and rejection using EyeCatch, and channel data reconstruction. This pipeline is shown in Fig. 1A. We processed the data from the first experiment, both with and without ASR; we processed the second dataset both with and without applying ORICA and EyeCatch.

To quantify the results of this analysis, we calculated the signal-to-noise ratio (SNR) after applying each different pipeline (e.g. with or without ASR). For zero-mean signals, SNR is the ratio of signal variance to noise variance, σ_s^2/σ_a^2. As we cannot claim to know those values exactly, we approximated the SNR by dividing the average channel variance during the rest periods, σ_s^2, by the average channel variance during artifact periods, σ_n^2. In reality, $\sigma_n^2 \approx (\sigma_s + \sigma_a)^2$ since it reflects variability in the summed artifact and brain EEG activity. Therefore this SNR approximation is conservative and forms a lower bound on true SNR.

We computed the correlations of scalp maps from the ORICA decomposition to those from the offline Infomax ICA decomposition of the same data. We used offline Infomax ICA applied to the common-average referenced and FIR bandpass-filtered data with artifact periods manually removed as a gold-standard solution to compare against. Excluding the artifacts, these recordings are relatively stationary and so the offline solution is effectively the best-case solution possible for ORICA.

Finally, to get a sense of how ORICA learns under these different conditions, we also look at the dynamics of the non-stationary index (NSI) which quantifies the magnitude

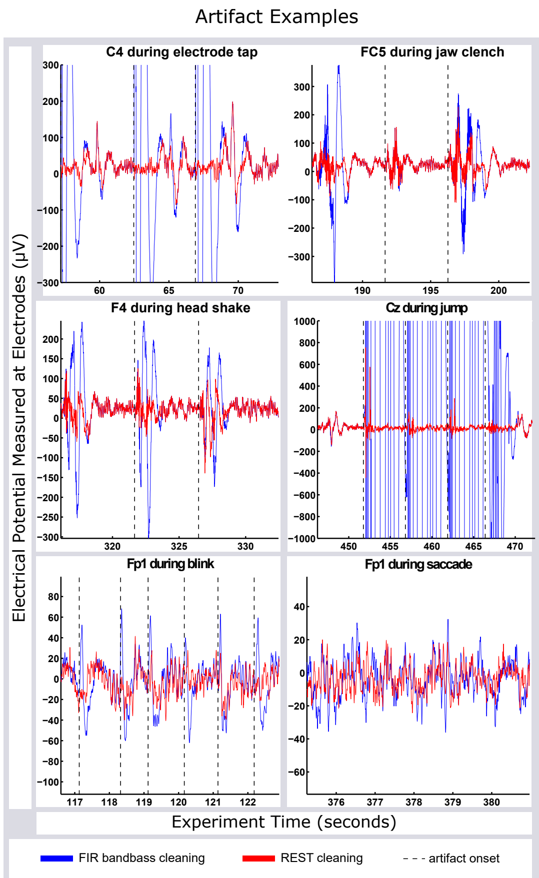

Fig. 2. Examples of common EEG artifacts. From top to bottom and left to right: electrode tapping, jaw clenching, head shaking, jumping, blinking, and eye movements. The blue traces are cleaned with an FIR bandpass filter, while the red traces are further processed with ASR and ICA-based cleaning using ORICA and EyeCatch. Artifact onset times are indicated by black dotted lines.

of the ORICA gradient over time [8]. When data is very improbable under the current ORICA model, e.g. during transient artifacts, the NSI will have a large value relative to baseline.

III. Results and Discussion

TABLE I

SNR at distinct stages of processing.

Artifact	FIR	ASR	ASR-ICA	ICA
Electrode Tap	0.174	1.09	1.05	0.176
Jaw Clench	0.586	0.723	0.693	0.580
Head Shake	3.78e-5	0.349	0.339	3.78e-5
Jump	1.05e-4	0.529	0.511	1.05e-4
Blink	0.465	0.465	0.791	-
Saccade	0.4979	0.810	1.01	-

A. Motion and muscle artifacts

ASR successfully removed a significant portion of the artifact-induced signal features in the first experiment. As shown in the top four frames of Fig. 2, while not all of the artifact signals were removed by ASR, the exceptionally strong artifact periods were consistently reprojected to a

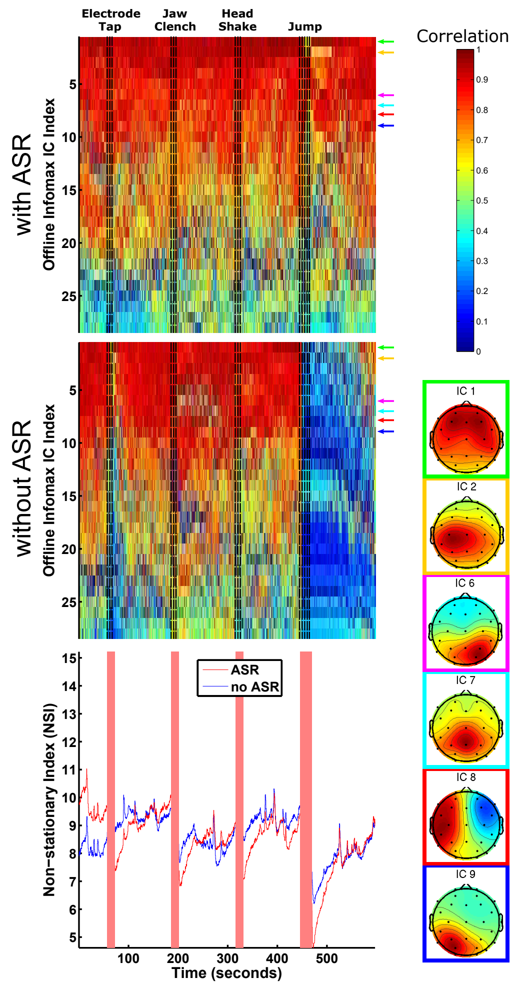

Fig. 3. The two top images indicate how well, at different times during the transient artifact recording, the ORICA decomposition results match results of an offline Infomax ICA decomposition of the whole dataset. The top plot shows results of applying ASR before ORICA while the middle one does not use ASR. Example scalp maps for selected rows are shown, matched by border color to arrows pointing to the corresponding row in the plots. The bottom plot traces the non-stationary index values throughout the recording, both with and without use of ASR. In the top plots, artifact onsets are shown as black dashed lines while in the bottom plot they are shown as red lines.

more normal range for brain-dominated EEG. This effect can also be seen in Table I by the consistent rise in approximate SNR in the ASR and ASR-ICA columns as compared to the other two, which forgo ASR. Furthermore, Fig. 3 indicates that ASR stabilized the ORICA decomposition in the presence of large-amplitude artifacts. This is indicated by the lower correlation values following the artifact events when ASR was omitted. In such cases, the ORICA model rapidly changed to try to better fit the artifact activity. When ASR preprocessing was added, there was little change in the ORICA model before and after artifact occurrences, demonstrating increased robustness to non-brain EEG 'noise.' This is particularly evident after subject jumps. Without ASR preprocessing, the entire model was lost after the jumps, i.e., not a single IC from the ORICA decomposition correlated highly with any ICs in the offline Infomax ICA solution. This is pivotal because, even though the ICA portion of

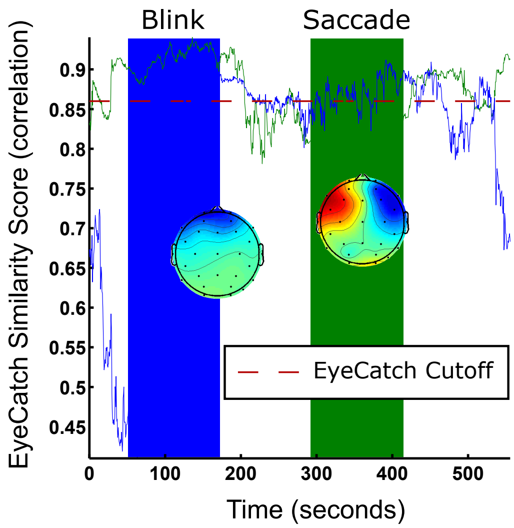

Fig. 4. The above traces show how the eye-related independent components (IC) learned by ORICA were rated by EyeCatch. The threshold for removal, 0.86, is indicated as a red dashed line. The period where the subject was blinking is indicated by the blue shaded region while the period where the subject looked back and forth is indicated by the green shaded region. The subject kept his eyes closed during all other time periods (white background). The blink IC was found and removed quickly during blinking while the saccade IC did not remain suprathreshold during lateral eye-movement.

the pipeline was not used in this experiment, typical EEG recording will have both eye movement-related artifacts as well as body motion artifacts. If the ORICA decomposition is lost every time a body motion artifact occurs (as it does without ASR preprocessing), then any eye movement-related ICs may not be found and could not then be cleanly removed from the data.

For transient artifacts, Table I shows the addition of ICA-based cleaning produced a minute decrease in SNR; likely because there were no stereotyped artifacts for ORICA to learn and remove. All ICA-based cleaning could do is find occasional false-positives, which in this case increased the power of the signal negligibly. The NSI traces in the bottom panel of Fig. 3 show that ORICA follows the same general learning patterns with and without ASR, but exhibits more extreme NSI values without ASR preprocessing, as ORICA may be more directly exposed to effects of high-amplitude artifacts. It is worth noting that ORICA was able to find many brain-related ICs including those shown in the bottom-right of Fig. 3. This suggests possible further uses of REST beyond data cleaning, in particular as a tool for real-time monitoring and source analysis, given the added robustness provided by ASR.

B. Eye-induced artifacts

For the eye movement-related artifacts in the second experiment, Table I indicates ASR had a negligible effect on eye-blink artifacts and a more significant effect on saccade artifacts. However, SNRs during both types of eye activity were further improved by the addition of ICA-based cleaning. The speed with which ORICA found the eye movement-related ICs is shown in Fig. 4. Once the subject opened his eyes and began blinking, it took ORICA twenty-six seconds to converge well enough on the blink-related IC for EyeCatch to remove it. Even after two more minutes of eyes-closed resting, the maximum blink IC scalp map correlation remained near the EyeCatch threshold level and subsequently increased again when the subject reopened his eyes. However, the saccade-related IC EyeCatch score fluctuated across the rejection threshold as the subject performed lateral eye-movements resulting in incomplete saccade artifact rejection. It appears the altered version of EyeCatch used here was not ideal, as the changes seem to have introduced some instability in the correlations found, though no better option is currently available.

IV. Conclusion

We have introduced a new pipeline for real-time EEG artifact removal that combines the use of ASR, ORICA, and an IC classifier (here EyeCatch) using the Real-time EEG Source-mapping Toolbox (REST). We studied how the pipeline performed in the presence of six different types of artifacts common in EEG recordings and found it removed the majority of the artifact-induced signal features. We also compared the performance of the pipeline with and without an initial application of ASR and found that the presence of ASR stabilized the ORICA decomposition, which is desirable for cleaning the data of eye movement-related artifact. The pipeline is available as part of REST, which is freely available at the url: `https://github.com/goodshawn12/REST`

References

[1] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, "Trends in eeg-bci for daily-life: Requirements for artifact removal," *Biomedical Signal Processing and Control*, vol. 31, pp. 407–418, 2017.

[2] J. A. Urigüen and B. Garcia-Zapirain, "Eeg artifact removalstate-of-the-art and guidelines," *J Neural Eng*, vol. 12, no. 3, p. 031001, 2015.

[3] A. Gevins, C. Yeager, G. Zeitlin, S. Ancoli, and M. Dedon, "On-line computer rejection of eeg artifact," *Electroencephalography and clinical Neurophysiology*, vol. 42, no. 2, pp. 267–274, 1977.

[4] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from eeg data using blind component separation," *Psychophysiology*, vol. 41, no. 2, pp. 313–325, 2004.

[5] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic artifact rejection from multichannel scalp eeg by wavelet ica," *IEEE Sensors Journal*, vol. 12, no. 3, pp. 533–542, 2012.

[6] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Extended ica removes artifacts from electroencephalographic recordings," in *NIPS*, 1998, pp. 894–900.

[7] M. T. Akhtar, T.-P. Jung, S. Makeig, and G. Cauwenberghs, "Recursive independent component analysis for online blind source separation," in *ISCAS*. IEEE, 2012, pp. 2813–2816.

[8] S.-H. Hsu, T. R. Mullen, T.-P. Jung, and G. Cauwenberghs, "Real-time adaptive eeg source separation using online recursive independent component analysis," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 24, no. 3, pp. 309–319, 2016.

[9] C. A. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," U.S. Patent Application No. 14/895,440, 2014.

[10] N. Bigdely-Shamlo, K. Kreutz-Delgado, C. Kothe, and S. Makeig, "Eyecatch: Data-mining over half a million eeg independent components to construct a fully-automated eye-component detector," in *IEEE EMBC*. IEEE, 2013, pp. 5845–5848.

[11] L. Pion-Tonachini, S.-H. Hsu, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time eeg source-mapping toolbox (rest): Online ica and source localization," in *IEEE EMBC*. IEEE, 2015, pp. 4114–4117.

[12] C. Kothe, "Lab streaming layer (lsl)," *https://github.com/sccn/labstreaminglayer*, 2014.