Applying Outlier Detection and Independent Component Analysis for Compressed Sensing EEG Measurement Framework

Shun KATSUMATA[1], Daisuke KANEMOTO[2*], and Makoto OHKI[1]
[1]*Graduate Faculty of Interdisciplinary Research, University of Yamanashi*
[2]*Graduate School of Engineering, Osaka University*
*E-mail: dkanemoto@eei.eng.osaka-u.ac.jp

***Abstract*—We previously proposed a framework that applies independent component analysis (ICA) after random undersampling to compress EEG signals with artifacts. In the previous framework, compressed signals are decomposed into several independent EEG signal components, and all components that include artifacts are set to zero in order to reconstruct the compressed EEG measurement signal without artifacts. In this paper, we propose a framework using outlier detection-ICA instead of ICA for efficiently removing only the artifacts. Herein, the performance of the proposed framework is evaluated using raw EEG signals with an ocular artifact pseudo-model. Compared with previous framework, the addition of outlier detection lowers the normalized mean square error and preserves the EEG signals with a high accuracy.**

Keywords—EEG, compressed sensing, independent component analysis, outlier detection, artifact

I. Introduction

Electroencephalograms (EEGs) provide biometric information that can reveal signs of brain inflammation, epilepsy [1], sleep disorders, and Alzheimer's disease [2]. Long-term wireless EEG telemonitoring is desirable for these diagnoses. EEG wireless monitoring systems require minimized energy consumption due to limited battery life; with a low energy consumption, smaller portable EEG devices could be designed that use little battery power. The compressed sensing (CS) theory is an attractive method for signal acquisition and compression that can reduce power dissipation [3]. CS theory allows for signal recovery from fewer measurements than the conventional Nyquist sampling technique if the signal is sparse in some transform domains. However, other biological signals and artifacts caused by measurement mechanisms degrade the EEG signal quality during monitoring. Ocular artifacts (OAs), or eye-blink artifacts, are typical high-amplitude artifacts that disturb the low-frequency region of EEG signals and worsen the signal sparsity.

Independent component analysis (ICA) has been demonstrated as an effective algorithm for removing artifacts from EEG signals [4], [5]. For example, we recently proposed a framework that applies ICA after random undersampling [6]. We found that compressed EEG signals with artifacts can be reconstructed efficiently using the framework to eliminate the artifact components. However, to obtain an improved reconstructed signal, we focused on how to more accurately detect and remove the artifact components. In a previous work, a combination of outlier detection and ICA (OD-ICA) methods was proposed [7] to remove OAs automatically from EEG signals. The research results are shown for only uncompressed signals.

In this paper, we propose a new framework that can remove artifacts from the inserted region only before reconstruction by utilizing OD-ICA in a compressed domain. In the proposed framework, a high-precision signal can be achieved owing to the improved sparsity. The rest of this paper is organized as follows. Section II describes the previous EEG measurement framework and briefly discusses the theory of OD-ICA. Section III details the proposed framework, and Section IV presents quantitative reconstruction performance results. Section V concludes the paper.

II. Methods

A. Reported framework

The previous framework [6] applies ICA to remove artifact interference after random undersampling in a data processing unit. In the framework, we use a random sampling measurement matrix in CS, and N-size data are compressed into $M(<N)$-size data. After data compression, multichannel EEG signals are decomposed into independent components (ICs) according to their statistical independence conditions. EEG signals with artifacts are decomposed into several EEG signal components and artifact components by ICA. We then obtain EEG signals without artifacts by setting the artifact components to zero, then remixing the decomposed ICs. Although signal processing in this previously reported framework is simple, it is difficult to accurately separate only pure artifact components as single ICs. This means that some of the EEG signal remains in the artifact components, and thus zeroing the artifact components also removes meaningful EEG signals.

B. OD-ICA

In the OD-ICA method, OD detects the peak pattern position and range resulting from OAs. Importantly, the threshold value for detecting the range must be derived. The artifact ranges are determined as follows. First, all local maxima in the component

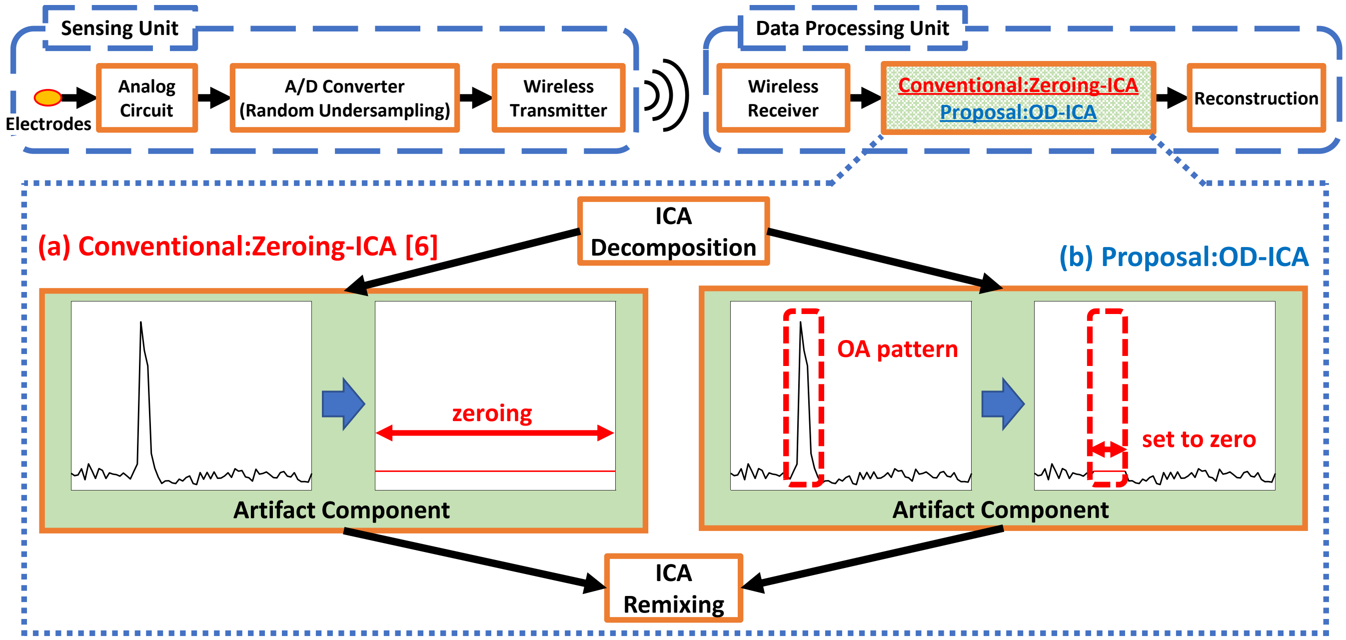


Fig. 1. Conventional and proposed EEG measurement frameworks.

are detected (N: number of local maxima in the component). Next, the differences in amplitude values between neighboring local maxima are calculated ($\Delta V = V_{n+1} - V_n, n = 1{:}\,N$). Here, V_n is the amplitude of the n-th local maximum in the measured EEG signal. Then, ΔV is compared with the threshold, and ranges exceeding the threshold are detected as an artifact in the inserted range.

There are several ways to determine a threshold, but in this framework, the threshold is determined using the same adjusted box plot as that used in [7], which is a modified version of the box plot. The signals are represented by quartiles sequenced from smaller to larger, and the data are divided in order of Q_1, Q_2, and Q_3. The signal is divided into former and latter sections based on Q_2, the signal median. Q_1 is the median of the former section, and Q_3 is the median of the latter section. The IQR (interquartile range) is calculated as the difference between Q_3 and Q_1 ($IQR = Q_3 - Q_1$), which is a robust measure of scale. The skewness estimator MC (medcouple) value is also calculated, which is a robust measure of skewness [8]. Using Q_1, Q_3, IQR, and MC, the upper and lower limits are calculated by the following equations.

If $MC \geq 0$:

$$\text{Lower Limit} = Q_1 - 1.5e^{-4MC} IQR$$

$$\text{Upper Limit} = Q_3 + 1.5e^{3MC} IQR \qquad (1)$$

If $MC < 0$:

$$\text{Lower Limit} = Q_1 - 1.5e^{-3MC} IQR$$

$$\text{Upper Limit} = Q_3 + 1.5e^{4MC} IQR. \qquad (2)$$

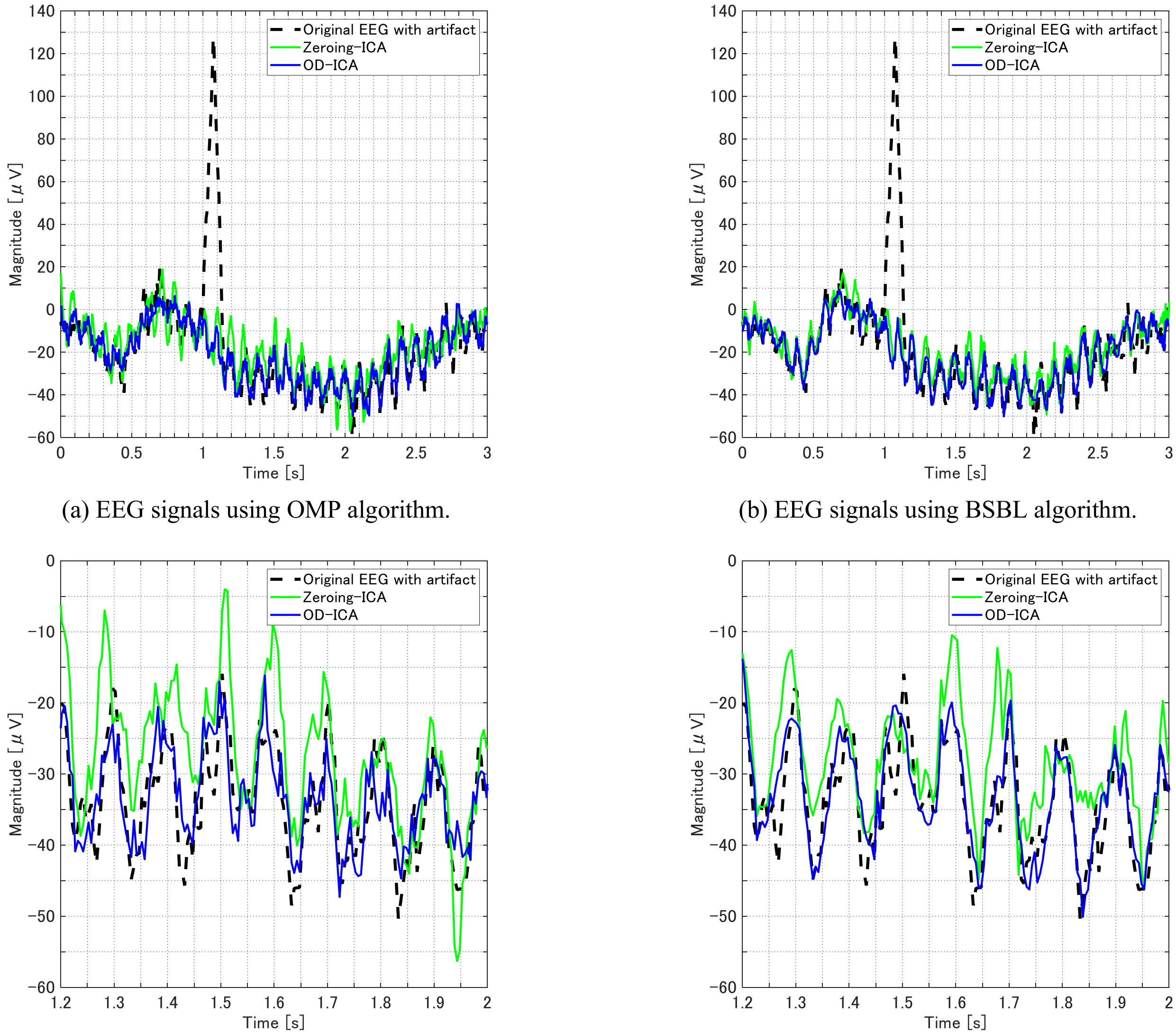
In this study, we adopted the threshold upper limit for upward-shaped artifacts and the lower limit for downward-shaped artifacts.

III. Proposed Framework

Fig.1 shows the frameworks of the previously reported and currently proposed EEG measurements. Both frameworks compress the EEG signals based on CS by random undersampling with an A/D converter. The compressed EEG signals are sent to the data processing unit by a wireless transmitter in the sensing unit and decomposed into ICs (several EEG signal components and artifact components) by ICA. Fig.1 (a) shows the conventional method, in which the whole OA component is set to zero. Then, EEG signals without OA are obtained by remixing the ICs. The advantage of the framework in Fig.1 (a) is simplicity. However, since some EEG signals remains in the OA components, there is the problem of meaningful EEG signals also being removed. In contrast, as shown in Fig.1 (b), our newly proposed framework sets only the OA pattern within a rectangle to zero. This successfully preserves the meaningful EEG signals.

IV. Evaluation

We evaluated the performance of the proposed framework in MATLAB. The raw EEG signals were recorded at 200 Hz in 16 channels, and 100 epochs each containing 3 s were used. In this framework, we target the EEG contaminated with OA, and thus we generated an OA pseudo-model to evaluate recovery accuracy. Generally, the magnitude of OAs received near the eye electrodes is larger than that near other parts of the body. Therefore, the maximum OA voltage at electrodes FP1 and FP2 was set to 150 μV and to 75 μV at electrodes F3, F4, F7, and F8. In the other electrodes (C3, C4, P3, P4, T3, T4, T5, T6, O1, and O2), the OA voltage was set to 15 μV. The OAs were added to the measured raw EEG signals. In this evaluation, we used the



(a) EEG signals using OMP algorithm.

(b) EEG signals using BSBL algorithm.

(c) magnified view of (a).

(d) magnified view of (b).

Fig. 2. EEG signals (original with artifact, OD-ICA, zeroing-ICA).
(a) recovered using OMP algorithm. (b) recovered using BSBL algorithm. (c) magnified view of (a). (d) magnified view of (b).

principal component analysis-ICA algorithm [9] for ICA and the orthogonal matching pursuit (OMP) algorithm or block sparse Bayesian learning (BSBL) algorithm [10] for reconstruction. We used the discrete cosine transform (DCT) basis to represent EEG signal sparsely, and random undersampling technique [6] in A/D converter to compress EEG signal. The BSBL reconstruction method works considerably better than OMP but requires complicated calculations and long reconstruction times. Thus, BSBL implementations may not be suitable for real or quasi-real time reconstruction.

Fig.2 (a) and (b) show examples of EEG waveforms (uncompressed original data with artifacts, OD-ICA and zeroing-ICA [6]) using OMP and BSBL, respectively at FP1. The compression ratio ($M/N \times 100\%$) is 25%. A large pseudo-model OA is shown from 1.0 to 1.15 s in the original data. To see the reconstructed signal waveform in more detail, Fig.2 (c) and (d) show magnified views of Fig.2 (a) and (b), respectively, from 1.2 to 2.0 s. Compared with the previously reported framework, our proposed OD-ICA framework can cleanly reconstruct the EEG waveforms.

To evaluate the performance of the proposed framework, the normalized mean square error (NMSE) was calculated by

$$\mathrm{NMSE} = \frac{\|x - \hat{x}\|_2^2}{\|x\|_2^2}, \tag{3}$$

where x is the original EEG signal and $\hat{x}$ is the reconstructed EEG signal. Fig.3 and Fig.4 show the NMSE values for two different artifact-added parts using OMP and BSBL, respectively. In these results, we ignored the data from 1.0 to

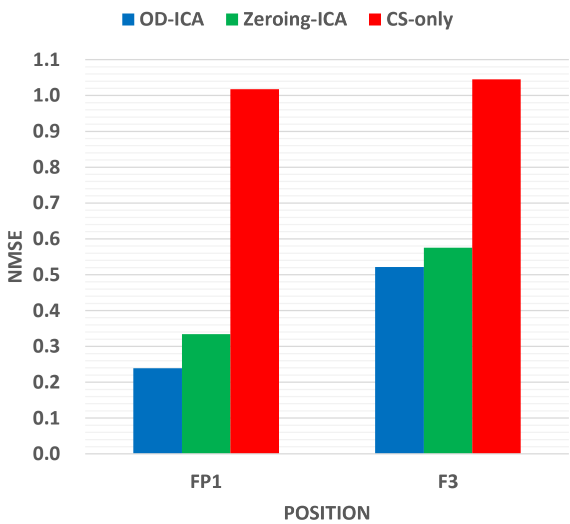


Fig. 3. NMSE results using OMP algorithm.

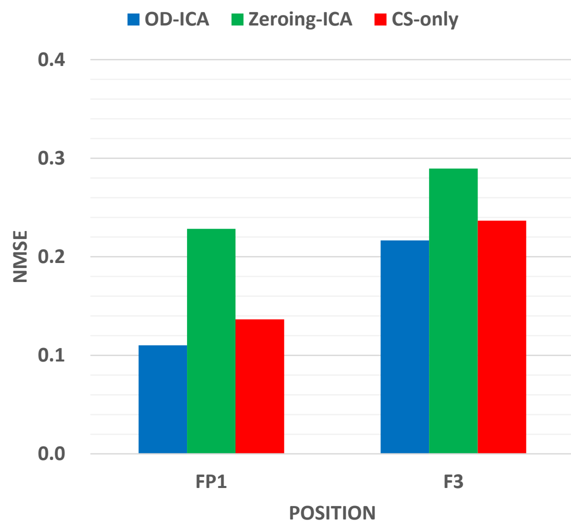


Fig. 4. NMSE results using BSBL algorithm.

1.15 s, in which the CS-only result is strongly affected by the artifact, to perform a fair comparison between OD-ICA, zeroing-ICA, and CS-only. The NMSE values of the proposed framework are lower than those of the CS-only and previous frameworks. Notably, the NMSEs of the previous framework using the BSBL algorithm are higher than those of the CS-only framework in this simulation environment (Fig.4). However, the proposed OD-ICA framework does lower the NMSE.

V. Conclusion

We proposed a new framework to apply OD-ICA after CS to remove artifacts from recorded EEG signals. This framework compresses the EEG signal and then removes artifacts using the OD-ICA method before reconstruction. In this paper, we show how the EEG signal is improved using OD-ICA instead of zeroing-ICA. Herein, the proposed framework was built and evaluated in MATLAB using 100 epochs of 3 s and 16 channel raw EEG signals with an OA pseudo-model. The results show that the NMSE is lower and the EEG signal can be reconstructed more cleanly compared with conventional methods using zeroing-ICA or CS-only.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP18K18023 and by JKA.

References

[1] H.R.Mohseni, A.Maghsoudi, and M.B.Shamsollahi, "Seizure Detection in EEG Signals: A Comparison of Different Approaches", in *Procs. IEEE Engineering in Medicine and Biology Society*, pp. 6724-6727, Sep. 2006.

[2] T.Musha, H.Matsuzaki, Y.Kobayashi, Y.Okamoto, M.Tanaka, and T.Asada, "EEG Markers for Characterizing Anomalous Activities of Cerebral Neurons in NAT (Neuronal Activity Topography) Method", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2332-2338, Aug. 2013.

[3] Z.Zhang, T.Jung, S.Makeig, and B.D.Rao, "Compressed Sensing of EEG for Wireless Telemonitoring with Low Energy Consumption and Inexpensive Hardware", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 221-224, Jan. 2013.

[4] N.P.Castellanos, and V.A.Makarov, "Recovering EEG Brain Signals: Artifact Suppression with Wavelet Enhanced Independent Component Analysis", *Journal of Neuroscience Methods*, vol. 158, no. 2, pp. 300-312, Dec. 2006.

[5] S.Makeig, A.J.Bell, T.Jung, and T.J.Sejnowski, "Independent Component Analysis of Electroencephalographic Data", *Advances in Neural Information Processing Systems*, pp. 145-151, Aug. 1996.

[6] D.Kanemoto, S.Katsumata, M.Aihara, and M.Ohki, "Framework of Applying Independent Component Analysis After Compressed Sensing for Electroencephalogram Signals", *IEEE Biomedical Circuits and Systems Conference*, pp. 145-148, Oct. 2018.

[7] S.Çınar, N.Acır, "A Novel System for Automatic Removal of Ocular Artefacts in EEG by Using Outlier Detection Methods and Independent Component Analysis", *Expert Systems with Applications*, vol. 68, pp. 36-44, Feb. 2017.

[8] M.Hubert and E.Vandervieren, "An Adjusted Boxplot for Skewed Distributions", *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5186-5201, Aug. 2008.

[9] K.Zhang and L.Chan, "ICA by PCA Approach: Relating Higher-Order Statistics to Second-Order Moments", in *Procs. International Conference on Independent Component Analysis and Signal Separation*, pp. 311-318, Mar. 2006.

[10] Z.Zhang and B.D.Rao, "Recovery of Block Sparse Signals using the Framework of Block Sparse Bayesian Learning", in *ICASSP 2012*, pp. 3345-3348, Mar. 2012.