



# Independent Low-Rank Matrix Analysis-Based Automatic Artifact Reduction Technique Applied to Three BCI Paradigms

Suguru Kanoga<sup>1\*</sup>, Takayuki Hoshino<sup>1,2</sup> and Hideki Asoh<sup>1</sup>

<sup>1</sup> Artificial Intelligence Research Center, Department of Information Technology and Human Factors, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan, <sup>2</sup> Graduate School of Media and Governance, Keio University, Kanagawa, Japan

## OPEN ACCESS

### Edited by:

Tzzy-Ping Jung,  
University of California, San Diego,  
United States

### Reviewed by:

Erwei Yin,  
China Astronaut Research and  
Training Center, China  
Dong Ming,  
Tianjin University, China

### \*Correspondence:

Suguru Kanoga  
s.kanouga@aist.go.jp

### Specialty section:

This article was submitted to  
Brain-Computer Interfaces,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 25 February 2020

**Accepted:** 20 April 2020

**Published:** 09 June 2020

### Citation:

Kanoga S, Hoshino T and Asoh H  
(2020) Independent Low-Rank Matrix  
Analysis-Based Automatic Artifact  
Reduction Technique Applied to Three  
BCI Paradigms.  
Front. Hum. Neurosci. 14:173.  
doi: 10.3389/fnhum.2020.00173

Electroencephalogram (EEG)-based brain-computer interfaces (BCIs) can potentially enable people to non-invasively and directly communicate with others using brain activities. Artifacts generated from body activities (e.g., eyeblinks and teeth clenches) often contaminate EEGs and make EEG-based classification/identification hard. Although independent component analysis (ICA) is the gold-standard technique for attenuating the effects of such contamination, the estimated independent components are still mixed with artifactual and neuronal information because ICA relies only on the independence assumption. The same problem occurs when using independent vector analysis (IVA), an extended ICA method. To solve this problem, we designed an independent low-rank matrix analysis (ILRMA)-based automatic artifact reduction technique that clearly models sources from observations under the independence assumption and a low-rank nature in the frequency domain. For automatic artifact reduction, we combined the signal separation technique with an independent component classifier for EEGs named ICLabel. To assess the comparative efficiency of the proposed method, the discriminabilities of artifact-reduced EEGs using ICA, IVA, and ILRMA were determined using an open-access EEG dataset named OpenBMI, which contains EEG data obtained through three BCI paradigms [motor-imagery (MI), event-related potential (ERP), and steady-state visual evoked potential (SSVEP)]. BCI performances were obtained using these three paradigms after applying artifact reduction techniques, and the results suggested that our proposed method has the potential to achieve higher discriminability than ICA and IVA for BCIs. In addition, artifact reduction using the ILRMA approach clearly improved (by over 70%) the averaged BCI performances using artifact-reduced data sufficiently for most needs of the BCI community. The extension of ICA families to supervised separation that leaves the discriminative ability would further improve the usability of BCIs for real-life environments in which artifacts frequently contaminate EEGs.

**Keywords:** electroencephalogram, brain-computer interface, independent component analysis, artifact reduction, independent low-rank matrix analysis

## 1. INTRODUCTION

An electroencephalogram (EEG)-based brain–computer interface (BCI) is a well-established technology that enables communicating with others without performing actual body movements by finding specific brain activity patterns from EEGs and converting these into predefined commands (Wolpaw et al., 2002). Several paradigms are used for eliciting robust time-independent or -dependent potentials in EEGs, such as motor imagery (MI) (Pfurtscheller and Da Silva, 1999), event-related potential (ERP) (Squires et al., 1976), and steady-state visual evoked potential (SSVEP) (Regan, 1966). Along with these paradigms, developments in machine-learning-based classifiers/identifiers have contributed to improvements in finding specific (elicited) patterns. These paradigms, in combination with signal processing modules and emerging technologies (e.g., wearable sensing, mobile computing, and virtual/augmented reality), have attracted increasing attention in many domains, such as medicine and robotics for real-world applications (Wang et al., 2018; Ogino et al., 2019; Vourvopoulos et al., 2019).

Artifact potentials must be reduced in all EEGs in order to realize robust real-world BCI applications because strong artifact contamination effects can easily reduce BCI performances (Kanoga et al., 2019b). During EEG measurements in real environments, biological artifacts like muscular and ocular ones cannot be avoided because it is difficult for people to voluntarily control the number of their artifact-generating activities. For example, healthy adult males blink  $\sim 20$  times per minute (i.e., once every  $\sim 3$  s) to maintain the moisture of their eyes (Karson, 1983). Most BCI paradigms provide visual stimuli or cues for eliciting specific neuronal patterns, and the abovementioned artifacts could contaminate the resulting EEGs. Further, muscular and ocular artifacts will contaminate all EEGs as long as the scalp has some conductivity. These unavoidable artifacts have high-amplitude electrical potentials and overlapping frequency characteristics compared to EEGs (Halliday et al., 1998; Hagemann and Naumann, 2001); thus, contamination by such artifacts makes EEG-based classification/identification hard. These contamination effects can be attenuated by increasing the distance from the source (Kanoga et al., 2016).

For denoising the contamination effects in EEG analysis, the well-known and powerful blind source separation (BSS) technique based on independent component analysis (ICA) has been widely used for the last 20 years (Jung et al., 2000). Usually, artifact reduction involves three steps: (1) training a demixing matrix, (2) identifying the types of separated independent components (ICs), and (3) remixing EEGs by using only neuronal ICs and an inverse demixing matrix. To improve both the computational cost and the accuracy of training a demixing matrix, many ICA algorithms, such as fast ICA (Hyvärinen and Oja, 1997), second-order blind interference (SOBI) (Belouchrani et al., 1997), and information maximization (infomax) ICA (Bell and Sejnowski, 1995), have been proposed. The SOBI and infomax ICA algorithms are used most commonly for EEG signal processing (Choi et al., 2005;

Urigüen and Garcia-Zapirain, 2015). For example, EEGLAB, an enormous interactive toolbox for EEG analysis, implements the infomax ICA algorithm (Delorme and Makeig, 2004). Recently, ICLabel (Pion-Tonachini et al., 2019), an automatic IC classifier, has been integrated into the EEGLAB toolbox for online streaming EEG data. Overall, an ICA-based approach remains the gold-standard for artifact reduction.

While real-world EEG-based BCI applications are being developed steadily, ICA-based source estimation still poses some problems. ICA algorithms comprehensively minimize the reconstruction error with a linear combination for an entire sequence of trials. However, this approach overestimates sources for representing the latent waveform of an observation; thus, the estimation leads to oversubtraction or spectral distortion of the EEG power (Wallstrom et al., 2004; Castellanos and Makarov, 2006). Proposing a more rigorous representation of estimated sources than ICA is a major challenge in EEG signal processing for constructing effective classifiers/identifiers.

To represent meaningful waveforms from an observation, we focused on the recurrent properties of an artifactual waveform over trials. Biological artifacts are based on a person's organ structure. An organ system reproducibly and unconsciously activates a non-cerebral source (e.g., eyeball) in the same manner and generates similar electrical potentials (e.g., electrooculogram signals); therefore, person-specific artifacts share a few basic functions for representing the waveforms and can be considered low-rank matrices comprising multiple short time segments. This study represents and removes such waveforms using an independent low-rank matrix analysis (ILRMA) that finds a low-rank non-negative matrix based on statistical independence (Kitamura et al., 2016). To improve its usability for EEG analysis, we used ICLabel for artifact reduction. We investigated the discriminabilities of artifact-reduced EEGs obtained by different methods by using OpenBMI, an open-access EEG dataset that contains EEG data, through the three abovementioned BCI paradigms (MI, ERP, and SSVEP). The BCI performances with these three paradigms after applying artifact reduction techniques were obtained. The results suggest that the proposed method can potentially achieve higher discriminability than ICA for BCIs.

## 2. ARTIFACT REDUCTION TECHNIQUES

### 2.1. Mixing and Demixing of EEGs

A typical approach to artifact reduction in EEG observations is based on the following assumption:  $P$ -channel EEG observations are overdetermined/determined and linear combinations of unknown cerebral  $Q$  sources including artifactual and neuronal ones and white noises. Neuronal cells have limited propagation because the cortical connectivity is highly weighted toward short ( $< 500 \mu\text{m}$ ) connections (Budd and Kisvárdy, 2001). Thus, the electrical potentials of neuronal activities spread through a contiguous cortical region with a high attenuation penalty in proportion with the distance from the sources (Arieli et al., 1996; Onton and Makeig, 2006). By matching the underlying dynamics of the generation and propagation of EEG potentials,

the aforementioned assumption can be represented as

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{d}(n), \tag{1}$$

where  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_P(n)]^T$  is the EEG observation at the  $n$ th sampling point ( $1 \leq n \leq N$ ),  $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_Q(n)]^T$  is the unknown source,  $\mathbf{A}$  is the  $P \times Q$  full-rank unknown mixing matrix, and  $\mathbf{d}(n) = [d_1(n), d_2(n), \dots, d_P(n)]^T$  is the additive zero-mean noise ( $^T$  indicates the transpose).

Here, we have only an artifact-(un)contaminated EEG observation matrix  $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(N)] \in \mathbb{R}^{P \times N}$ . Because source estimation through the inverse process is evidently intractable, four additional assumptions are made (James and Hesse, 2004; Vigario and Oja, 2008): (1) the noise is spatially uncorrelated with the observed data ( $\mathbb{E}[\mathbf{A}\mathbf{s}(n)\mathbf{d}(n)] = \mathbf{0}$ , where  $\mathbb{E}[\cdot]$  is the expectation operator); (2) the noise is temporally uncorrelated ( $\mathbb{E}[\mathbf{d}(n)\mathbf{d}(n + \tau)] = \mathbf{0}$ , where  $\tau$  is the lag time ( $\forall \tau > 0$ )); (3) the number of sources is equal to or lower than the number of observations ( $P \geq Q$ ); and (4) the mixing matrix  $\mathbf{A}$  does not change over time. Under these assumptions, we can simultaneously estimate both the source matrix  $\hat{\mathbf{S}} = [\hat{\mathbf{s}}(1), \dots, \hat{\mathbf{s}}(N)] \in \mathbb{R}^{Q \times N}$  and the demixing matrix  $\mathbf{W}(= \mathbf{A}^{-1}) \in \mathbb{R}^{Q \times P}$  to blindly separate the observations into artifactual/neuronal sources:

$$\hat{\mathbf{s}}(n) = \mathbf{W}\mathbf{x}(n). \tag{2}$$

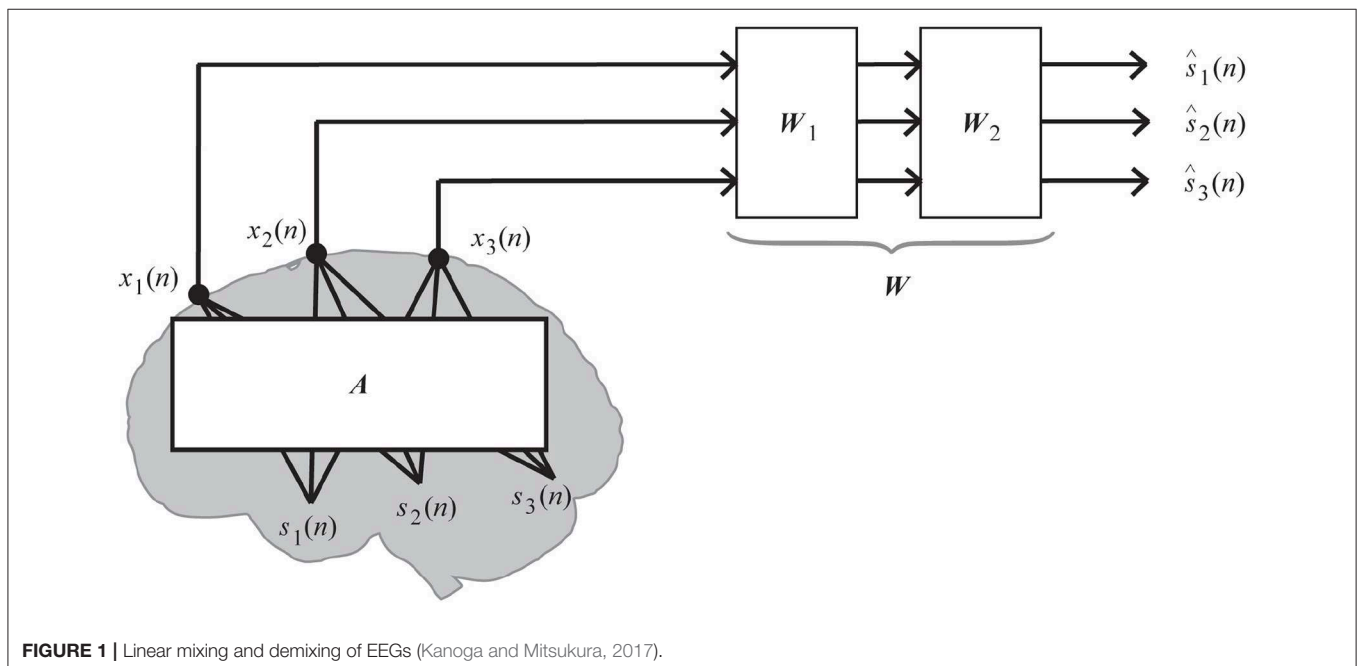
The linear mixing and demixing of EEGs shown in **Figure 1** accounts for the comprehensive demixing matrix  $\mathbf{W}(= \mathbf{W}_1\mathbf{W}_2)$  because signal separation algorithms sometimes first decorrelate the data by  $\mathbf{W}_1$  and then demix them by  $\mathbf{W}_2$ , which is originally learned from the algorithm. In this study, we decorrelated the

data before applying a matrix factorization technique; thus, the following representation of  $\mathbf{W}$  has the same meaning as  $\mathbf{W}_2$  in **Figure 1**.

In practice, artifact reduction requires three stages of processing: (1) decomposing the input matrix; (2) identifying whether the decomposed component is artifactual or neuronal; and (3) reconstructing the artifact-reduced signals using only neuronal components. In this case, we assumed that the EEG observations are labeled signals; the dimensionality of the sources and observations is the same (the value of  $P$  and  $Q$  is 20, 32, or 10 for MI, ERP, or SSVEP paradigm. The value depends on the number of selected channels in the BCI paradigm. More detailed information is described in sections 3.1, 3.2, and 3.3). In addition, we decomposed datasets using three BSS methods for ICA families: ICA, independent vector analysis (IVA), and ILRMA. Then, the decomposed components were automatically identified using the ICLabel function in the EEGLAB toolbox proposed by Pion-Tonachini et al. (2019). Based on the labels, artifact-reduced signals were linearly reconstructed.

Although ICA algorithms handle time-series data, IVA and ILRMA algorithms approximate a bin-wise instantaneous mixture model in a short-time Fourier transform (STFT) domain. An EEG time series is transformed into a sequence of complex-valued signals by using STFT with a 50% overlapped 1-s Hamming window. Thus, the observations and sources in each time-frequency slot are described as  $\mathbf{x}_{ij} = [x_{ij1}, x_{ij2}, \dots, x_{ijP}]^T \in \mathbb{C}^P$  and  $\hat{\mathbf{s}}_{ij} = [\hat{s}_{ij1}, \hat{s}_{ij2}, \dots, \hat{s}_{ijQ}]^T \in \mathbb{C}^Q$ , where a couple of  $(i, j)$  defines the  $i$ th frequency bin and  $j$ th time frame over STFT ( $1 \leq i \leq I$  and  $1 \leq j \leq J$ ). IVA and ILRMA algorithms assume the following mixing system:

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij}, \tag{3}$$



**FIGURE 1** | Linear mixing and demixing of EEGs (Kanoga and Mitsukura, 2017).

where  $\mathbf{A}_i = [\mathbf{a}_{i1}, \mathbf{a}_{i2}, \dots, \mathbf{a}_{iQ}]^H \in \mathbb{C}^{P \times Q}$  is a frequency-wise mixing matrix ( $\mathbf{a}_{iq}$  is the steering vector for the  $q$ th source, and  $^H$  indicates the Hermitian transpose). In this paper, we set the value of time length to 1 s because some frequency-domain artifact reduction techniques translate EEG data into STFT domain based on 1-s windows (Kanoga and Mitsukura, 2014; Mohammadpour and Rahmani, 2017) and ILRMA showed its high separation accuracy for 1-s time length data (Kitamura et al., 2017). This mixing system is the rank-1 spatial model (Duong et al., 2010); thus, the relationships between observations and sources can be represented:

$$\hat{\mathbf{s}}_{ij} \approx \mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij}, \tag{4}$$

where  $\mathbf{y}_{ij} = [y_{ij1}, y_{ij2}, \dots, y_{ijQ}]^T \in \mathbb{C}^Q$  is the STFT of the estimated signals and  $\mathbf{W}_i = [\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots, \mathbf{w}_{iQ}]^H = \mathbf{A}_i^{-1} \in \mathbb{C}^{Q \times P}$  is the demixing matrix. Note that the demixing matrices for IVA, ILRMA, and ICA have different dimensionalities because of the differences in the domain used (IVA and ILRMA:  $\mathbf{W} \in \mathbb{R}^{Q \times P \times I}$ , ICA:  $\mathbf{W} \in \mathbb{R}^{Q \times P}$ ).

## 2.2. Matrix Factorization Techniques

### 2.2.1. Independent Component Analysis

ICA is the most famous classical method for separating multichannel EEG observations  $\mathbf{x}(n)$  into statistically independent sources  $\hat{\mathbf{s}}(n)$  based on an estimated demixing matrix  $\mathbf{W}$  (Jung et al., 2000; Delorme et al., 2007). The sources can be said to be statistically independent when the following relationship holds:

$$p(\hat{\mathbf{s}}) = \prod_{q=1}^Q p(\hat{\mathbf{s}}_q), \tag{5}$$

where  $p(\hat{\mathbf{s}})$  and  $p(\hat{\mathbf{s}}_q)$  are the joint and the marginal probability distribution of the sources, respectively. Thus, ICA algorithms optimize the demixing matrix  $\mathbf{W}$  by minimizing the dependence between these distributions. This study applied the extended infomax ICA algorithm implemented by Lee et al. (1999) using the `runica` function in EEGLAB to the observations. In this algorithm, the dependence in the distributions is represented as the mutual information (Kullback–Leibler distribution) between the estimated sources and observations  $\mathcal{I}(\hat{\mathbf{s}}; \mathbf{x})$ :

$$\mathcal{I}(\hat{\mathbf{s}}; \mathbf{x}) = \mathcal{H}(\hat{\mathbf{s}}) - \sum_{q=1}^Q \mathcal{H}(\hat{\mathbf{s}}_q), \tag{6}$$

where

$$\mathcal{H}(\hat{\mathbf{s}}) = - \int p(\hat{\mathbf{s}}) \log p(\hat{\mathbf{s}}) d\hat{\mathbf{s}}, \tag{7}$$

$$\mathcal{H}(\hat{\mathbf{s}}_q) = - \int p(\hat{\mathbf{s}}_q) \log p(\hat{\mathbf{s}}_q) d\hat{\mathbf{s}}_q. \tag{8}$$

By applying the relationship  $p(\mathbf{x}) = p(\hat{\mathbf{s}})/|\det \mathbf{W}|$  to Equation (7), Equation (6) can be rewritten as a cost function for optimizing

the demixing matrix:

$$\mathcal{I}(\mathbf{W}) = \text{const.} - \sum_{q=1}^Q \mathcal{H}(\hat{\mathbf{s}}_q) - \log |\det \mathbf{W}|. \tag{9}$$

The entropy of given observations  $\mathcal{H}(\mathbf{x})$  is a constant. In addition, a gradient update rule based on the natural gradient (Amari, 1998) with learning rate  $\eta$  is used to solve the optimization problem:

$$\mathbf{W} \leftarrow \mathbf{W} + \eta \Delta \mathbf{W}, \tag{10}$$

where

$$\Delta \mathbf{W} = (\mathbf{I} - \mathbb{E}[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^T])\mathbf{W}. \tag{11}$$

In every iteration, the distribution of the estimated source for the score function  $\varphi(\hat{\mathbf{s}}_q)$  is chosen from the super-Gaussian or sub-Gaussian based on the sign of the fourth cumulant of each source  $c_4 = M_4 - 3M_2^2$ , where  $M_k$  is the  $k$ th moment ( $M_k = \mathbb{E}[\hat{\mathbf{s}}_q^k]$ ).

$$\varphi(\hat{\mathbf{s}}_q) = -(\hat{\mathbf{s}}_q + \text{sgn}(c_4)\tanh(\hat{\mathbf{s}}_q)). \tag{12}$$

In a real environment, the expectation operator  $\mathbb{E}[\cdot]$  is the expected value of the empirical distribution (the sample average of the variable).

### 2.2.2. Independent Vector Analysis

IVA is an extension of the ICA algorithm to multivariate components (vectorized signals) (Hiroe, 2006; Kim et al., 2006b). Like ICA algorithms, IVA defines the dependence between joint probability distributions and marginal probability products using the Kullback–Leibler divergence; however, it introduces a vector density model that has a variance dependency within a source vector. This study applied the natural-gradient-based IVA algorithm implemented by Kim et al. (2006b) based on the `ivabss` function from an open-access toolbox available on Github (<https://github.com/teradepth/iva>).

Two conditions are assumed: (1) elements of a source vector are mutually independent of those of other source vectors; and (2) within a source vector, the elements are highly dependent on each other. Based on these assumptions, the cost function for multivariate random variables to separate the components from the observations can be written as

$$\mathcal{I}(\mathbf{W}) = \text{const.} - \sum_{q=1}^Q \mathcal{H}(\hat{\mathbf{s}}_q) - \sum_{i=1}^I \log |\det \mathbf{W}_i|. \tag{13}$$

The cost function preserves the inherent dependency within each source vector, though it removes the dependency between different source vectors.

By differentiating the object function with respect to the coefficients of demixing matrices  $\mathbf{W}_i$  and using the natural gradient, we can derive a gradient update rule as

$$\Delta \mathbf{W}_i \leftarrow \mathbf{W}_i + \eta \Delta \mathbf{W}_i, \tag{14}$$



where

$$\Delta \mathbf{W}_i = (\mathbf{I} - \mathbb{E}[\varphi_i(\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_I)\hat{\mathbf{s}}_i^*])\mathbf{W}_i, \quad (15)$$

where  $\mathbf{a}^*$  denotes the complex conjugate of  $\mathbf{a}$ . Among a number of possible function forms, one of the simplest and most effective score functions is given as follows:

$$\varphi_i(\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_I) = \frac{\hat{\mathbf{s}}_i}{\sqrt{\sum_{i=1}^I |\hat{\mathbf{s}}_i|^2}} \quad (16)$$

To define an optimal form of the function  $p(\hat{\mathbf{s}}_q)$ , which has dependency within a source vector, IVA algorithms introduce a vector density model as a scale mixture of a Gaussian distribution with a fixed mean and a variable variance (Kim et al., 2006a):

$$p(\hat{\mathbf{s}}_q) = \prod_{i=1}^I p(\hat{\mathbf{s}}_{iq}) = \alpha \prod_{i=1}^I \exp\left(-\sqrt{(\hat{\mathbf{s}}_{iq} - \mu_{iq})^\dagger \Sigma_{iq}^{-1} (\hat{\mathbf{s}}_{iq} - \mu_{iq})}\right) \quad (17)$$

where  $\alpha$  is a normalization term,  $\mu_{iq}$  and  $\Sigma_{iq}$  are respectively the mean vector and covariance matrix of the  $q$ th source signal in the  $i$ th frequency bin, and  $\mathbf{a}^\dagger$  is the conjugate transpose of  $\mathbf{a}$ .

### 2.2.3. Independent Low-Rank Matrix Analysis

The ILRMA method unifies IVA and non-negative matrix factorization (Lee and Seung, 1999; Sawada et al., 2013) by considering the determined situation ( $P = Q$ ) and a linear time-invariant mixing system (Kitamura et al., 2016). This study applied the ILRMA algorithm implemented by Kitamura et al. (2016) using the ILRMA function from an open-access toolbox available on Github (<https://github.com/d-kitamura/ILRMA>). The algorithm estimates both the demixing matrix  $\mathbf{W}_i$  and the STFT of the estimated signals  $\mathbf{y}_{ij}$  by approximately decomposing  $|y_{ijq}|^2$  into the non-negative elements  $t_{ik}$  and  $v_{kj}$  of the basis matrix  $\mathbf{T}_q \in \mathbb{R}^{I \times K}$  and the activation matrix  $\mathbf{V}_q \in \mathbb{R}^{K \times J}$  with a latent variable  $z_{qk}$  of the partitioning function  $\mathbf{Z}$ , which indicates whether or not the  $k$ th basis ( $1 \leq k \leq K$ ) belongs to the  $q$ th source. For the decomposition, the ILRMA algorithm has the following cost function:

$$\mathcal{I}(\mathbf{W}) = \sum_{ij} \left\{ \sum_q \log \sum_k z_{qk} t_{ik} v_{kj} + \sum_q \frac{|y_{ijq}|^2}{\sum_k z_{qk} t_{ik} v_{kj}} - 2 \log |\det \mathbf{W}_i| \right\}, \quad (18)$$

where  $y_{ijq} = \mathbf{w}_{iq}^H \mathbf{x}_{ij}$ . The cost function finds a low-rank time-frequency structure for sources using the first and second terms in Equation (18) and maximizes the statistical independence between sources using the second and third terms in Equation (18).

In this algorithm, the demixing matrix  $\mathbf{W}_i$  can be efficiently updated through iterative projection based on the auxiliary function technique (Ono, 2011):

$$\mathbf{V}_{iq} = \frac{1}{J} \sum_j \frac{1}{r_{ijq}} \mathbf{x}_{ij} \mathbf{x}_{ij}^H, \quad (19)$$

$$\mathbf{w}_{iq} \leftarrow (\mathbf{W}_i \mathbf{V}_{iq})^{-1} \mathbf{e}_q, \quad (20)$$

$$\mathbf{w}_{iq} \leftarrow \mathbf{w}_{iq} (\mathbf{w}_{iq}^H \mathbf{V}_{iq} \mathbf{w}_{iq})^{-\frac{1}{2}}, \quad (21)$$

where  $r_{ijq}$  is the estimated variance of each source under the complex Gaussian distribution and  $\mathbf{e}_q$ , a unit vector in which the  $q$ th element is equal to unity. These update rules have been reported to be faster and more stable than conventional update rules (e.g., natural gradient). After the update, the separated signal  $\mathbf{y}_{ij}$  is also updated:

$$\mathbf{y}_{ijq} \leftarrow \mathbf{w}_{iq}^H \mathbf{x}_{ij}. \quad (22)$$

In addition, the basis matrix  $\mathbf{T}_q$ , activation matrix  $\mathbf{V}_q$ , and partitioning function  $\mathbf{Z}$  can be updated by the majorization-minimization algorithm (Hunter and Lange, 2000):

$$z_{qk} \leftarrow z_{qk} \sqrt{\frac{\sum_{ij} |y_{ijq}|^2 t_{ik} v_{kj} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-2}}{\sum_{ij} t_{ik} v_{kj} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-1}}}, \quad (23)$$

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{jq} |y_{ijq}|^2 z_{qk} v_{kj} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-2}}{\sum_{jq} z_{qk} v_{kj} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-1}}}, \quad (24)$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{iq} |y_{ijq}|^2 z_{qk} t_{ik} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-2}}{\sum_{iq} z_{qk} t_{ik} (\sum_{k'} z_{qk'} t_{ik'} v_{k'j})^{-1}}}. \quad (25)$$

Finally, the estimated source model is represented as

$$r_{ijq} = \sum_k z_{qk} t_{ik} v_{kj}. \quad (26)$$

Note that the demixing matrix  $\mathbf{W}_i$  and the estimated variance  $r_{ijq}$  are normalized at each iteration to avoid the risk of diverging as follows:

$$\lambda_q = \sqrt{\frac{1}{IJ} \sum_{ij} |y_{ijq}|^2}, \quad (27)$$

$$\mathbf{w}_{iq} \leftarrow \mathbf{w}_{iq} \lambda_q^{-1}, \quad (28)$$

$$y_{ijq} \leftarrow y_{ijq} \lambda_q^{-1}, \quad (29)$$

$$r_{ijq} \leftarrow r_{ijq} \lambda_q^{-2}. \quad (30)$$

The number of bases for all sources  $K$  and number of iterations were set to  $J/10$  and 200, respectively.

### 2.3. Component Identification

For identifying the estimated ICs obtained from ICA, IVA, and ILRMA, we used the ICLabel (<https://github.com/scnn/ICLabel>) classifier proposed by Pion-Tonachini et al. (2019) (freely available as a package in EEGLAB; Delorme and Makeig, 2004; Delorme et al., 2011). This classifier uses three artificial neural networks (ANNs) (specifically, two ‘‘Classifier’’ networks and one ‘‘Generator’’ network): (1) a convolutional neural network (CNN) optimized by an unweighted cross-entropy loss, (2) a

CNN optimized by a weighted cross-entropy loss and neuronal IC classification errors, and (3) a semi-supervised learning generative adversarial network (SSGAN) (Odena, 2016; Salimans et al., 2016). The ICLabel classifier inputs, architectures, and training paradigms are described in detail in Appendices B and E in Pion-Tonachini et al. (2019). By using these three ANNs and the three IC features, the ICLabel classifier classified unlabeled EEG ICs into seven categories: (1) brain, (2) muscle, (3) eye, (4) heart, (5) line noise, (6) channel noise, and (7) others. In this study, because of the property of the epoch identification method described in section 4, we removed ICs whose labels were “muscle” or “eye.” Currently, the ICLabel classifier has been trained using 6352 EEG recordings in storage drives collected over the past 15 years at Swartz Center for Computational Neuroscience (SCCN) at UC San Diego.

To accurately identify IC labels using the ICLabel classifier, three discriminable features were calculated from each IC: (1)  $32 \times 32$  pixel scalp topography using the `topoplot` function in EEGLAB, (2) median power spectral densities (PSDs) from 1 to 50 Hz using a variation of the Welch method (Welch, 1967), and (3) the autocorrelation function. The scalp topographies and PSDs were scaled such that each had a maximum absolute value of 0.99. Further, the autocorrelation vectors were normalized such that the zero-lag value was 0.99. Note that the estimated mixing matrix ( $W^{-1}$ ), channel locations in 3D space, and channel labels were required to generate the scalp topographies. We collected information about the channel locations in 3D space from the `sample_locs` folder in EEGLAB toolbox. In addition, the estimated demixing matrix has a 3D structure except for ICA; thus, the corresponding frequency band (e.g., 8–30 Hz) was extracted from whole frequency bins, and the summation was computed to transform the matrix into a 2D structure. Furthermore, the matrix was scaled by the number of extracted frequency bins.

## 2.4. Signal Reconstruction

By performing the identification process using ICLabel, labels are obtained for the estimated ICs. If the label is “eye” or “muscle,” all components of the artifactual IC are set to zeros. Based on the modified sources, artifact-reduced EEG signals in EEG observations were reconstructed using the inverse linear demixing process in ICA. While applying IVA and ILRMA, the sources were translated into frequency components. Thus, all frequency components of the artifactual ICs were set to zeros and translated into time-series data by the inverse STFT. Then, the artifact-reduced EEG signals were reconstructed, and the inverse ICA linear demixing process was performed.

## 3. MATERIALS AND BASELINE METHODS

To assess the discriminability of artifact-reduced EEGs by ICA, IVA, and ILRMA, we downloaded an open-access EEG dataset published by Lee et al. (2019) from the webpage [http://gigadb.org/dataset/view/id/100542/File\\_page](http://gigadb.org/dataset/view/id/100542/File_page). The EEG data were recorded using 62 electrodes according to the International 10–20 system using BrainAmp (Brain Products; Munich, Germany) with a sampling rate of 1,000 Hz. In the analysis procedures, we

commonly downsampled all EEG data to 100 Hz. The reference and ground channels were nasion and AFz, respectively. The impedance of the EEG electrodes was maintained below 10 k $\Omega$ . Participants were instructed to comfortably sit in a chair with armrests  $\sim 60$  cm in front of a 21-inch LCD monitor (refresh rate: 60 Hz; resolution: 1,600  $\times$  1,200). In addition, they were asked to relax their muscles and minimize their eye and muscle movements during the BCI paradigms. Before beginning the experiments, five kinds of 10-s artifact-contaminated EEG data were measured: (1) eye blinking, (2) repetitive horizontal eye movements, (3) repetitive vertical eye movements, (4) teeth clenching, and (5) flexing both arms.

The dataset has the following three properties: (1) a large number of subjects (54 healthy participants; 29 males and 25 females; age: 24–35 years), (2) multiple sessions (two sessions on different days), and (3) multiple paradigms (a binary-class MI, a 36-symbol ERP, and a four-target-frequency SSVEP). Each session consisted of training and testing phases. All BCI paradigms were developed based on the OpenBMI toolbox (Lee et al., 2016) and Psychtoolbox (Brainard, 1997). We used this dataset because (1) EEGs in the three BCI paradigms were collected from the same participants, (2) each paradigm was conducted for 2 days, and (3) baseline analysis methods based on Matlab functions in the OpenBMI toolbox (<https://github.com/PatternRecognition/OpenBMI>) are available. A single dataset having all these properties is very important for fairly comparing algorithms to reveal general performances with intra- and inter-subject/paradigm variabilities in BCI research. To verify the change in the discrimination accuracy with artifact-reduced EEGs, the baseline analysis methods, including feature extraction and classification algorithms for each paradigm described in Lee et al. (2019), were used. Each paradigm and processing stream are described in detail in the following subsections.

### 3.1. MI Paradigm and Processing

The MI paradigm was designed based on a well-established protocol (Pfurtscheller and Neuper, 2001): a training/testing phase had 100 trials with 50 right and 50 left hand motion imagery tasks resulting in binary classification. Each trial lasted  $13 \pm 1.5$  s. In the first 3 s, a black fixation cross appeared at the center of the monitor. After the preparation time, the participant imagined a right or left grasping motion for 4 s depending on whether a right arrow or left arrow was displayed, respectively, and then remained in the resting state for  $6 \pm 1.5$  s.

In the MI paradigm, 20 electrodes in the motor cortex region (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6) were selected. The EEG data of the selected channels were band-pass filtered between 8 and 30 Hz through a fifth order Butterworth filter and further segmented into 2.5-s epochs, which are data segments of 1.0 to 3.5 s after the cue onsets (Pfurtscheller and Neuper, 2001; Fazli et al., 2009). We applied the filter bank common spatial pattern (FBCSP) to the epochs, which has been widely used in MI-based BCIs to maximize the discrimination of the binary class (Ang et al., 2008). A subset of the top and bottom two rows from the projection matrix was used for calculating log-variance features. Based on the features, linear discriminant analysis (LDA) classifiers were constructed and used.

### 3.2. ERP Paradigm and Processing

The ERP paradigm was designed based on a typical row-column speller system with random-set presentation (Yeom et al., 2014) and face stimuli (Kaufmann et al., 2011). The six rows and six columns were configured with 36 symbols (alphabets A to Z, numerals 1 to 9, and underscore “\_”). Each trial sequence lasted 19.5 s. In the first 4.5 s, a target character was highlighted for attracting the participant’s attention. After the preparation time, all rows and columns were flashed one by one (12 stimulus flashes) for 13 s and then remained in the resting state for 2 s. The stimulus-time interval was set to 80 ms and the interstimulus interval (ISI) to 135 ms. The highlighted target character was estimated based on data of five sequences (i.e., 60 flashes). In the training phase, participants were asked to copy the 33 characters including spaces in “NEURAL\_NETWORKS\_AND\_DEEP\_LEARNING” by gazing at the target character on the monitor, resulting in 1980 trials and binary classification (target or non-target character). In the testing phase, participants tried to copy the 36 characters including spaces in “PATTERN\_RECOGNITION\_MACHINE\_LEARNING,” resulting in 2,160 trials.

In the ERP paradigm, 32 electrodes (Fp-1/2, F-7/3/z/4/8, FC-5/1/2/6, T-7/8, C-3/z/4, TP-9/10, CP-5/1/2/6, P-7/3/z/4/8, PO-9/10, and O-1/z/2) were selected. The EEG data of the selected channels were band-pass filtered between 0.5 and 40 Hz through a fifth order Butterworth filter and then baseline-corrected by subtracting the average amplitudes of the prestimulus within an interval of 200 ms with respect to the stimulus onset. In addition, 0.8-s epochs after the onset were extracted for analysis. From the epochs, the mean amplitudes (MAs) over eight non-overlapping samples were calculated as the 320-dimensional subject-dependent spatio-temporal features (10 dimensions, 32 channels). Based on the features, LDA classifiers were constructed and used.

### 3.3. SSVEP Paradigm and Processing

The SSVEP paradigm was designed based on general requirements for SSVEP-based BCIs that run over four specific commands (Parini et al., 2009). Four flickers at 5.45, 6.67, 8.57, and 12 Hz were displayed at four positions (down, right, left, and up) on a monitor. Each target frequency was presented 25 times for both the training and the testing phases, resulting in four target identification problems. Each trial lasted 10 s. In the first 4 s, the participant gazed in the box where the target was highlighted (not flickering) in a different color, and the target flicker was then presented for 4 s with an ISI of 6 s to induce the target SSVEP.

In the SSVEP paradigm, 10 electrodes in the occipital region (P-7/3/z/4/8, PO-9/10, and O-1/z/2) were selected. The EEG data of the selected channels were segmented into 2-s epochs with respect to the stimulus onset. We applied multichannel canonical correlation analysis (CCA) (Lin et al., 2006) for identifying the target frequency index by calculating the correlation values between the input data and the prepared sinusoidal templates of the corresponding frequencies (5.45, 6.67, 8.57, and 12 Hz). Although this identification process does not need training data

owing to the use of an unsupervised classifier, only data from the testing phase were used for evaluating the BCI performance.

## 4. ASSESSMENTS

In this study, we assumed that artifact-reduced epochs are correctly classified if the artifact reduction technique effectively reduced artifactual effects from artifact-contaminated epochs. However, we do not know how many epochs of the aforementioned paradigms were contaminated by artifacts because the open-access EEG dataset does not provide such information. Empirically, it is difficult to completely avoid the generation of biological artifacts during EEG paradigms. Thus, we expected that some epochs were contaminated by some artifacts during each BCI paradigm. To identify the type of epoch (not artifact-contaminated or artifact-contaminated), we applied the detection of events in continuous time (DETECT) epoch identification method proposed by Lawhern et al. (2013); this method requires training data with clean and artifactual label information to make a multiclass SVM model (<https://github.com/VisLab/detect>). Usually, training data has a short time length (e.g., less than 5 s). Thus, in this study, 10 1-s no artifact-contaminated EEG data detected based on a manual inspection and extracted from the training phase of each BCI paradigm as “clean” epochs and five types of 10-s EEG data contaminated by artifacts, such as eye blinking, horizontal/vertical eye movements, teeth clenching, and flexing both arms, were prepared as “artifactual” epochs because these are well-known to generate ocular/muscular artifacts during EEG measurements. For training an SVM model, each 10-s-length artifactual data was separated into 10 1-s-length data without overlapping. Based on the 60 1-s epochs (10 1-s epochs  $\times$  6 classes), a 6-class SVM model was constructed. Note that segmented epochs of the BCI paradigms have different time length (i.e., MI: 2.5, ERP: 0.8, and SSVEP: 2.0 s). To apply DETECT based on a processing strategy for 1-s-length data, we extracted first 1-s data from the “clean” epochs if the target BCI paradigm was MI or SSVEP. For the epochs of the ERP paradigm, 0.2-s data before the stimulus onsets were concatenated to the “clean” epochs. Then, autoregressive features were extracted from the epochs to construct a multiclass SVM classifier of each BCI paradigm. The classifier and hard thresholding for the estimated artifactual class (certainty value obtained using the DETECT toolbox was over 0.5 or not) finally identified each epoch as being clean or artifact contaminated. **Table 1** lists the identification results. In all paradigms, data recorded in session 2 (day 2) had less artifactual data than data recorded in session 1 (day 1). In the MI and ERP paradigms, the number of artifactual data recorded in day 2 was significantly lower than data recorded in day 1 ( $p = 0.001, 0.009$  for MI and ERP paradigms in  $t$ -test). However, in the SSVEP paradigm, the number in day 2 was not significantly lower ( $p = 0.172$  in  $t$ -test). Note that we identify the epoch is neuronal unless a certainty value of all classes exceeded the hard threshold; thus, the thresholding process found an explicit artifactual class over 6 class labels. If the certainty values distributed throughout all classes and no one did not exceed the threshold, this modest

**TABLE 1** | Number of artifact-contaminated epochs in training and testing phases of each BCI paradigm and subject.

Subject	MI				ERP				SSVEP			
	Session 1		Session 2		Session 1		Session 2		Session 1		Session 2	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
s1	6	2	8	6	149	330 (36)	80	215 (33)	3	2	0	1
s2	2	0	0	1	95	5 (5)	38	115 (25)	0	0	0	0
s3	1	0	7	10	158	62 (20)	17	14 (9)	14	10	1	2
s4	0	0	0	0	32	34 (14)	7	10 (4)	9	18	0	0
s5	4	4	7	5	0	0 (0)	395	282 (25)	19	11	14	8
s6	4	2	4	3	159	113 (24)	171	146 (29)	25	26	5	6
s7	21	30	6	8	75	200 (25)	52	21 (10)	8	7	4	4
s8	2	0	4	1	427	325 (32)	35	87 (22)	3	2	9	5
s9	0	0	9	2	43	73 (24)	145	99 (25)	1	9	0	1
s10	0	2	0	0	94	15 (6)	61	78 (20)	0	0	2	4
s11	5	4	0	0	50	56 (15)	16	47 (17)	0	0	0	0
s12	0	16	2	6	75	94 (22)	98	185 (31)	0	1	1	7
s13	1	0	1	1	131	212 (29)	139	103 (26)	6	3	3	1
s14	1	0	2	1	18	35 (11)	234	190 (32)	4	6	13	22
s15	3	20	0	1	25	19 (11)	48	7 (5)	1	1	3	16
s16	0	2	13	10	492	322 (33)	113	92 (19)	0	0	8	9
s17	27	36	4	6	232	187 (30)	263	267 (34)	1	1	15	21
s18	1	11	13	9	166	404 (31)	92	122 (29)	3	2	2	4
s19	1	8	2	7	159	300 (33)	36	169 (30)	7	15	2	3
s20	0	7	1	1	236	240 (34)	112	323 (35)	2	2	1	2
s21	31	30	0	0	69	84 (29)	268	129 (26)	15	9	3	5
s22	2	8	2	0	131	128 (32)	11	48 (18)	9	2	0	1
s23	0	0	9	0	141	194 (29)	160	227 (26)	0	36	1	1
s24	0	1	8	4	65	106 (27)	87	158 (31)	2	1	1	5
s25	3	11	6	2	184	122 (24)	14	59 (15)	2	2	8	4
s26	0	1	0	0	447	552 (36)	196	265 (32)	20	22	7	6
s27	3	12	1	1	172	312 (34)	150	346 (32)	1	5	11	14
s28	2	6	0	0	93	63 (21)	37	52 (19)	6	4	2	1
s29	8	20	2	4	32	159 (25)	101	27 (9)	3	4	3	3
s30	3	9	4	0	141	171 (34)	38	30 (14)	0	1	0	0
s31	7	3	0	1	33	82 (23)	25	60 (29)	0	0	1	0
s32	33	67	0	2	122	339 (33)	80	32 (13)	13	34	19	21
s33	0	1	4	0	119	52 (17)	107	59 (15)	0	2	1	2
s34	32	18	8	19	254	181 (33)	142	441 (35)	13	21	18	16
s35	7	2	0	3	35	41 (13)	135	82 (22)	1	0	11	18
s36	3	2	6	17	178	189 (34)	110	159 (32)	3	4	0	4
s37	5	15	0	1	353	557 (36)	2	2 (1)	6	4	0	2
s38	11	8	0	0	194	119 (29)	53	29 (10)	17	11	11	11
s39	21	32	2	0	111	55 (22)	0	37 (11)	1	3	2	6
s40	1	2	1	0	145	206 (33)	11	56 (16)	31	26	3	7
s41	0	2	4	9	107	101 (28)	137	265 (31)	5	6	2	2
s42	0	0	9	10	121	191 (31)	35	62 (19)	10	15	10	2
s43	0	0	9	7	148	290 (35)	11	13 (7)	2	13	0	0
s44	3	7	3	13	293	392 (35)	32	76 (25)	10	7	7	0
s45	27	13	0	0	160	270 (34)	19	52 (19)	1	4	1	0
s46	0	1	6	11	226	307 (35)	52	175 (34)	1	4	2	4
s47	3	5	2	1	261	438 (36)	159	319 (36)	7	6	1	5
s48	6	1	0	2	25	52 (18)	41	35 (13)	2	5	20	12

(Continued)



TABLE 1 | Continued

Subject	MI				ERP				SSVEP			
	Session 1		Session 2		Session 1		Session 2		Session 1		Session 2	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
s49	3	10	6	8	104	54 (15)	54	22 (7)	14	8	1	1
s50	23	48	4	9	48	229 (24)	2	0 (0)	0	0	5	6
s51	9	9	3	1	354	170 (31)	136	173 (29)	0	3	3	1
s52	0	0	2	1	20	74 (23)	57	139 (31)	3	8	0	2
s53	2	2	0	2	370	176 (30)	16	25 (8)	3	3	1	4
s54	3	14	1	0	149	107 (18)	66	51 (17)	7	6	7	4
Mean	6.11	9.33	3.43	3.81	152	178 (25.8)	87.0	116 (21.1)	5.82	7.31	4.54	5.30
Std	9.27	13.2	3.54	4.63	115	136 (8.99)	79.6	104 (10.1)	6.97	8.55	5.39	5.86

The total number of epochs in training and testing phases of the MI and SSVEP BCI paradigms is 100, respectively. The total number of epochs in training or testing phases of the ERP paradigm is 1,980 or 2,160. Numbers in parentheses indicate that the number of target 36 characters including spaces in "PATTERN\_RECOGNITION\_MACHINE\_LEARNING" have been affected by artifact-contaminated epochs. When the number is 36, all copying processes contained the effect of artifacts.

identification method can not find artifact-contaminated epochs. In **Table 1**, there was an outlier: subject 5 in the session 1 of ERP paradigm had no artifact-contaminated epoch. This phenomenon might be caused by the above-mentioned reason.

Through the epoch identification process, artifact-contaminated epochs in both the training and the testing phases were detected. In the training phase, we also assumed that artifact-reduced epochs contribute to the construction of an effective classifier if the artifact reduction technique effectively reduced artifactual effects from the artifact-contaminated epochs. Therefore, the demixing matrix  $W$  was first trained by using all epochs in the training phase, and artifactual ICs were then removed from the artifact-contaminated epochs using the artifact reduction process described in sections 2.2, 2.3, and 2.4. After artifact reduction, the clean and artifact-reduced epochs in the training phase were applied to the baseline analysis methods described in section 3. For performance evaluation, the artifact-contaminated epochs in the testing phase were used to compute the classification accuracy of the artifact-reduced epochs:

$$Acc = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100\%, \quad (31)$$

where  $N_{\text{correct}}$  is the number of correct predictions, and  $N_{\text{total}}$  is the total number of artifact-contaminated epochs in the testing phase, as listed in **Table 1**, when the BCI paradigm was MI or SSVEP. Note that ERP data requires an averaging process for finding obvious feature waveforms (e.g., N200 and P300), and the averaged waveform relates to the classification performance. In other words, we cannot calculate the classification accuracy for each artifact-contaminated epoch in the paradigm. The numbers in parentheses in **Table 1** indicate the number of characters affected by artifacts ( $N_{\text{total}}$ ), which is directly related to the assessment results. **Figure 2** shows the block diagram of the assessment procedure used in this study.

A three-way repeated measures analysis of variances (ANOVAs) was applied to the classification accuracy to explore the effect of the two sessions, three BCI paradigms, and

three artifact reduction methods. In addition, artifact-reduced signals obtained using ICA, IVA, and ILRMA and that were reconstructed from muscular or ocular artifact-contaminated signals were visualized to qualitatively investigate the artifact reduction performances.

## 5. RESULTS

### 5.1. BCI Performance Before/After Applying Artifact Reduction Technique

**Table 2** lists the classification accuracies for all subjects, sessions, paradigms, and artifact reduction methods. In addition, **Figure 3** shows the averaged classification accuracies over all subjects. A three-way repeated measures ANOVA using the classification accuracies of all subjects showed the significant main effects of the BCI paradigms [ $F_{(2,829)} = 113.09, p < 0.001$ ] and artifact reduction methods [ $F_{(2,829)} = 3.05, p = 0.048$ ]; however, it did not show any significant main effect of the sessions [ $F_{(1,829)} = 1.29, p = 0.256$ ]. There were no interaction effects among them. *post-hoc* analysis using Tukey test revealed that the ICA and ILRMA results had a significant difference ( $p = 0.039$ ).

The classification accuracies obtained using ILRMA in all cases were always equal to or higher than the higher accuracy of using ICA or IVA (see **Table 2**). In particular, ILRMA improved the discriminability of artifact-reduced data for 31 subjects in session 1 and 24 subjects in session 2. When there was a difference in the artifact reduction performance, we highlighted the superior results in bold in the table. The averaged accuracy of using ILRMA in all BCI paradigms was also equal to or higher than that of using ICA and IVA (**Figure 3**). Interestingly, in some cases, artifact-contaminated data showed higher BCI performance than ICA and IVA. However, ILRMA always showed equal or higher performance compared to artifact-contaminated situations. For these results, ILRMA salvaged effective components for solving the classification problem from artifact-contaminated signals in the MI and SSVEP paradigms.

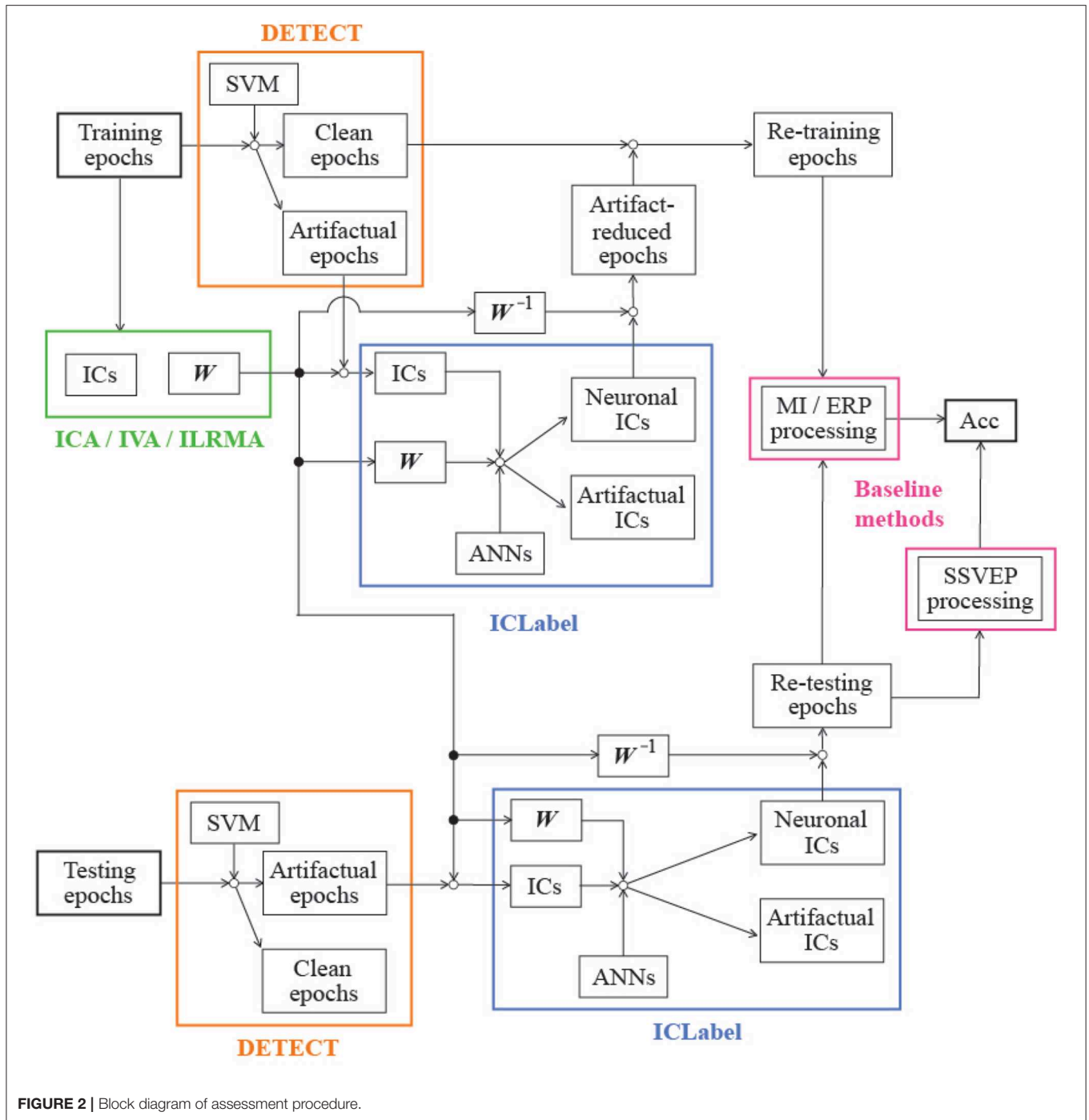


FIGURE 2 | Block diagram of assessment procedure.

Conversely, in the ERP paradigm, ICA was sufficient to remove the artifactual components and achieved almost 100% accuracy.

### 5.2. Representation of Original and Artifact-Reduced Signals

Figures 4, 5 show artifact-contaminated EEG epochs and artifact-reduced EEG epochs obtained using ICA, IVA, and ILRMA in the MI, ERP, and SSVEP paradigms. They were

qualitatively indicated that ILRMA could better remove artifact effects compared to ICA and IVA. In addition, the task-independent components were removed by ILRMA to leave characteristic features in each paradigm (e.g., event-related desynchronization caused by motor imagery and evoked potential by steady-state visual stimulus) instead of the attenuating power of all frequency components. This resulted in improvements in these BCI performances.

**TABLE 2** | Classification accuracies for all subjects, sessions, paradigms, and artifact reduction methods.

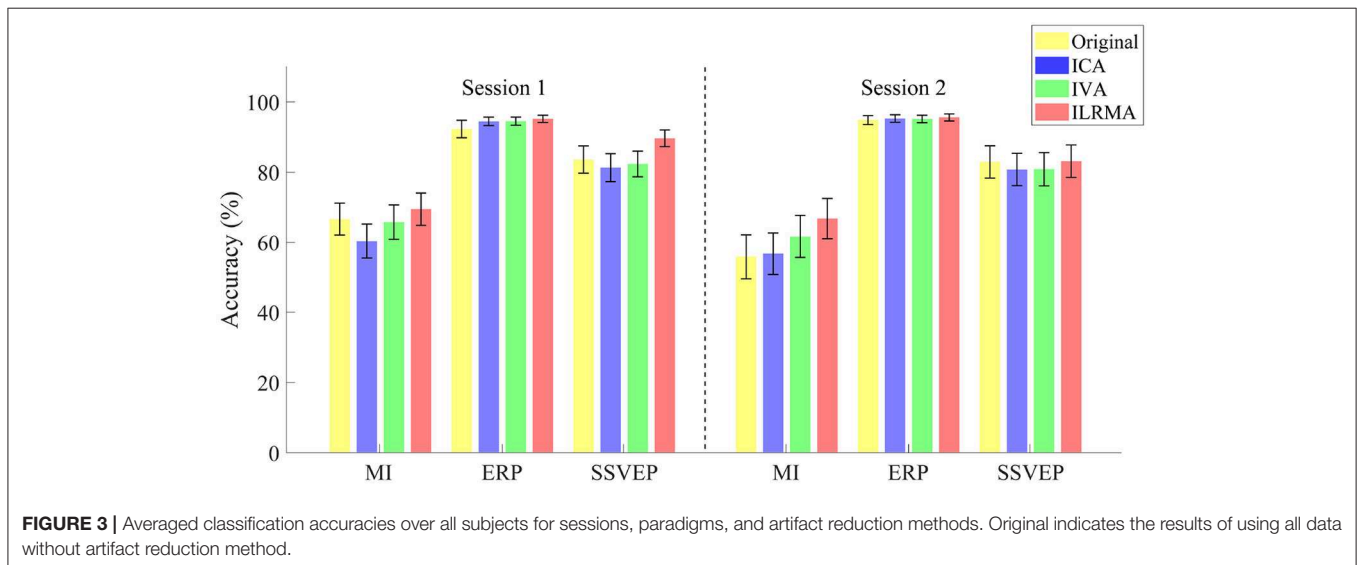
Subject	ICA						IVA						ILRMA					
	MI		ERP		SSVEP		MI		ERP		SSVEP		MI		ERP		SSVEP	
	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2
s1	0	33.3	55.6	<b>90.9</b>	<b>100</b>	100	0	16.7	<b>61.1</b>	81.8	50.0	100	0	<b>66.7</b>	<b>61.1</b>	<b>90.9</b>	<b>100</b>	100
s2	–	0	100	100	–	–	–	<b>100</b>	100	100	–	–	–	<b>100</b>	100	100	–	–
s3	–	100	100	100	80.0	50.0	–	100	100	100	<b>90.0</b>	<b>100</b>	–	100	100	100	<b>90.0</b>	<b>100</b>
s4	–	–	85.7	100	100	–	–	–	85.7	100	100	–	–	–	85.7	100	100	–
s5	100	80.0	–	56.0	<b>90.9</b>	100	100	80.0	–	56.0	72.7	100	100	80.0	–	56.0	<b>90.9</b>	100
s6	100	66.7	100	100	92.3	100	100	66.7	100	100	92.3	100	100	66.7	100	100	<b>96.2</b>	100
s7	<b>50.0</b>	62.5	100	100	85.7	100	46.7	62.5	100	100	85.7	100	<b>50.0</b>	62.5	100	100	85.7	100
s8	–	100	81.3	<b>100</b>	50.0	<b>40.0</b>	–	100	<b>87.5</b>	95.5	50.0	20.0	–	100	<b>87.5</b>	<b>100</b>	50.0	<b>40.0</b>
s9	–	100	100	100	55.6	0	–	100	100	100	55.6	0	–	100	100	100	55.6	0
s10	0	–	100	100	–	100	<b>50.0</b>	–	100	100	–	100	<b>50.0</b>	–	100	100	–	100
s11	75.0	–	100	100	–	–	<b>100</b>	–	100	100	–	–	<b>100</b>	–	100	100	–	–
s12	37.5	50.0	100	96.8	0	100	43.8	50.0	100	96.8	0	100	<b>50.0</b>	50.0	100	96.8	<b>100</b>	100
s13	–	0	100	100	100	0	–	0	100	100	100	0	–	0	100	100	100	0
s14	–	0	100	96.9	83.3	81.8	–	0	100	96.9	83.3	81.8	–	0	100	96.9	83.3	<b>90.9</b>
s15	70.0	0	100	100	100	87.5	70.0	0	100	100	100	87.5	70.0	0	100	100	100	<b>93.8</b>
s16	0	70.0	100	100	–	100	<b>100</b>	<b>90.0</b>	100	100	–	100	<b>100</b>	<b>90.0</b>	100	100	–	100
s17	86.1	<b>50.0</b>	<b>93.3</b>	91.2	0	81.0	86.1	33.3	90.0	91.2	<b>100</b>	81.0	86.1	<b>50.0</b>	<b>93.3</b>	<b>94.1</b>	<b>100</b>	<b>85.7</b>
s18	<b>81.8</b>	88.9	100	100	100	100	72.7	88.9	100	100	100	100	<b>81.8</b>	88.9	100	100	100	100
s19	100	71.4	100	100	73.3	100	100	<b>85.7</b>	100	100	73.3	100	100	<b>85.7</b>	100	100	<b>80.0</b>	100
s20	<b>71.4</b>	0	100	100	100	100	57.1	0	100	100	100	100	<b>71.4</b>	0	100	100	100	100
s21	86.7	–	100	100	<b>100</b>	100	93.3	–	100	100	88.9	100	<b>96.7</b>	–	100	100	<b>100</b>	100
s22	75.0	–	93.8	88.9	100	0	<b>87.5</b>	–	93.8	88.9	100	0	<b>87.5</b>	–	93.8	88.9	100	0
s23	–	–	<b>86.2</b>	69.2	47.2	0	–	–	72.4	<b>73.1</b>	52.8	0	–	–	<b>86.2</b>	<b>73.1</b>	<b>55.6</b>	0
s24	0	25.0	100	90.3	0	80.0	0	<b>75.0</b>	100	<b>93.6</b>	0	80.0	0	<b>75.0</b>	100	<b>93.6</b>	<b>100</b>	80.0
s25	27.3	<b>50.0</b>	100	93.3	100	75.0	<b>45.5</b>	0	100	93.3	100	<b>100</b>	<b>45.5</b>	<b>50.0</b>	100	93.3	100	<b>100</b>
s26	100	–	<b>83.3</b>	87.5	90.9	100	100	–	80.6	87.5	90.9	100	100	–	<b>83.3</b>	<b>90.6</b>	90.9	100
s27	41.7	0	100	100	100	85.7	50.0	0	100	100	100	92.9	<b>58.3</b>	<b>100</b>	100	100	100	<b>100</b>
s28	83.3	–	100	94.7	100	100	<b>100</b>	–	100	94.7	100	100	<b>100</b>	–	100	94.7	100	100
s29	65.0	50.0	88.0	100	75.0	100	85.0	75.0	<b>96.0</b>	100	75.0	100	<b>100</b>	<b>100</b>	<b>96.0</b>	100	75.0	100
s30	55.6	–	82.4	92.9	100	–	55.6	–	82.4	92.9	100	–	55.6	–	<b>85.3</b>	92.9	100	–
s31	100	0	100	100	–	–	100	0	100	100	–	–	100	0	100	100	–	–
s32	43.3	100	100	100	88.2	95.2	44.8	100	100	100	<b>91.2</b>	95.2	<b>52.2</b>	100	100	100	<b>91.2</b>	95.2
s33	100	–	100	93.3	100	0	100	–	100	93.3	100	0	100	–	100	93.3	100	0
s34	55.6	42.1	87.8	91.4	76.2	81.3	55.6	<b>47.4</b>	<b>90.9</b>	<b>97.1</b>	76.2	81.3	55.6	<b>47.4</b>	<b>90.9</b>	<b>97.1</b>	<b>90.5</b>	81.3
s35	<b>50.0</b>	66.7	100	95.5	–	83.3	0	66.7	100	95.5	–	83.3	<b>50.0</b>	66.7	100	95.5	–	83.3
s36	50.0	94.1	100	87.5	100	100	50.0	94.1	100	87.5	100	100	50.0	94.1	100	87.5	100	100
s37	86.7	100	69.4	100	100	100	<b>100</b>	100	72.2	100	100	100	<b>100</b>	100	<b>75.0</b>	100	100	100
s38	<b>75.0</b>	–	96.6	100	100	72.7	62.5	–	96.6	100	100	72.7	<b>75.0</b>	–	96.6	100	100	72.7
s39	62.5	–	86.4	100	100	<b>83.3</b>	62.5	–	86.4	100	100	50.0	62.5	–	86.4	100	100	<b>83.3</b>
s40	0	–	97.0	100	<b>57.7</b>	85.7	0	–	90.9	100	46.2	85.7	0	–	97.0	100	<b>57.7</b>	85.7
s41	50.0	66.7	100	100	<b>83.3</b>	100	50.0	66.7	100	100	66.7	100	50.0	66.7	100	100	<b>83.3</b>	100
s42	–	80.0	100	94.7	66.7	100	–	80.0	100	94.7	<b>73.3</b>	100	–	80.0	100	94.7	<b>73.3</b>	100
s43	–	85.7	100	100	100	–	–	85.7	100	100	100	–	–	85.7	100	100	100	–
s44	100	100	100	100	0	–	100	100	100	100	<b>100</b>	–	100	100	100	100	<b>100</b>	–
s45	76.9	–	91.2	94.7	100	–	<b>92.3</b>	–	91.2	94.7	100	–	<b>92.3</b>	–	91.2	94.7	100	–
s46	100	63.6	100	94.1	100	100	100	63.6	100	94.1	100	100	100	63.6	100	94.1	100	100
s47	80.0	100	100	<b>100</b>	33.3	80.0	80.0	100	100	97.2	33.3	80.0	80.0	100	100	<b>100</b>	33.3	80.0
s48	100	0	100	100	100	100	100	0	100	100	100	100	100	0	100	100	100	100

(Continued)

TABLE 2 | Continued

Subject	ICA						IVA						ILRMA					
	MI		ERP		SSVEP		MI		ERP		SSVEP		MI		ERP		SSVEP	
	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2	Se1	Se2
s49	50.0	50.0	100	100	100	100	60.0	50.0	100	100	87.5	100	<b>70.0</b>	50.0	100	100	100	100
s50	60.4	55.6	87.5	–	–	83.3	62.5	44.4	87.5	–	–	83.3	<b>68.8</b>	<b>66.7</b>	<b>91.7</b>	–	–	83.3
s51	44.4	0	87.1	89.7	66.7	100	44.4	<b>100</b>	<b>90.3</b>	89.7	66.7	100	44.4	<b>100</b>	<b>90.3</b>	89.7	66.7	100
s52	–	100	100	100	75.0	100	–	100	100	100	<b>100</b>	100	–	100	100	100	<b>100</b>	100
s53	0	100	96.7	100	100	100	0	100	96.7	100	100	75.0	0	100	96.7	100	100	100
s54	57.1	–	100	100	83.3	100	42.9	–	100	100	<b>100</b>	100	57.1	–	100	100	<b>100</b>	100
Mean	61.6	59.4	95.3	96.0	82.0	81.4	66.3	62.1	95.3	96.0	83.0	81.5	<b>70.0</b>	<b>67.3</b>	<b>96.0</b>	<b>96.4</b>	<b>90.4</b>	<b>83.8</b>
SE	4.91	5.98	1.22	1.08	4.01	4.66	4.93	6.09	1.17	1.06	3.70	4.78	4.65	5.80	1.04	1.03	2.40	4.65

Hyphens indicate that no artifact-contaminated epochs existed in the testing phase on the session. Bold values indicate that the accuracies over the three artifact reduction methods have different values and the results were higher than the other ones.



## 6. DISCUSSION

### 6.1. Automatic Processing Architecture

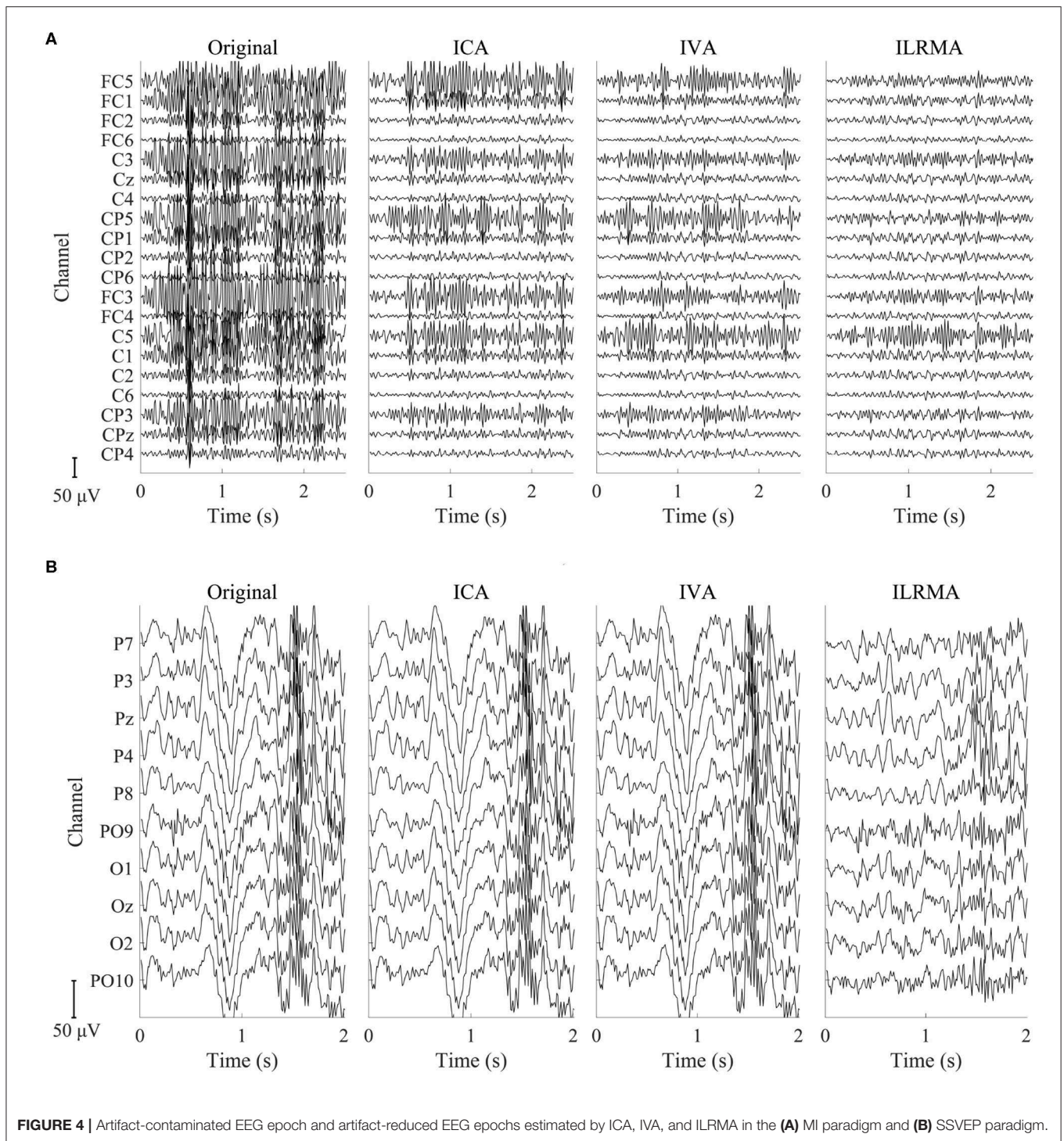
ICA-based artifact reduction techniques have been widely used in the field of EEG signal processing because of their powerful signal separation accuracy, simplicity (low computational cost), and ease of use (Delorme et al., 2007; Dimigen, 2019; Jiang et al., 2019). The techniques for limiting ocular and muscular artifacts (Chen et al., 2019; Tian et al., 2020) other than the ICA family are useful if they are integrated in a cascade-type processing module, which can automatically identify the type of artifact contained in the EEG observation. A simple filtering (linear combination) approach such as ICA, which multiplies the demixing matrix  $W$  as a filter, is faster and user-friendly. IVA and ILRMA use this property and can sufficiently cope with online processing as long as they can learn the demixing matrix. In addition, these algorithms can benefit from the ICLabel classifier (Pion-Tonachini et al., 2019) for IC identification to realize an automatic artifact reduction method.

Thus, the ILRMA-based artifact reduction technique (1) has higher accuracy than ICA, (2) has low computational cost (equivalent to ICA) in an online process, (3) is a changeable module for the ICA decomposition function, and (4) can simultaneously remove multiple types of artifacts. Note that the ICLabel classifier is expected to be updated frequently in the future. Although the EEGLAB toolbox keeps track of updates in the `run_ICL` function, the label assignment results may change depending on the updates. In this study, we selected the “default” version for IC identification.

### 6.2. Efficacy of Artifact Reduction for BCIs

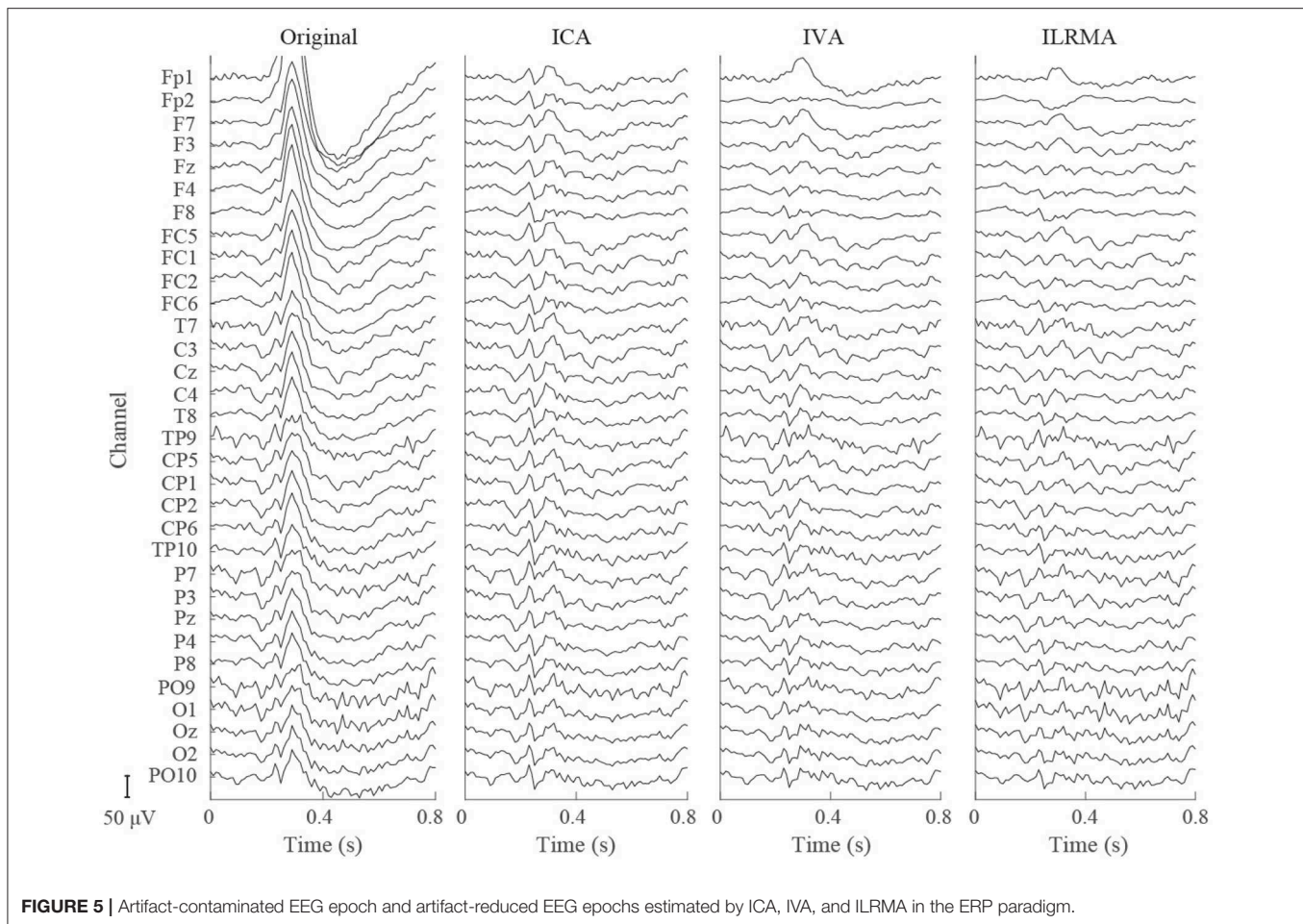
Researchers who propose original artifact reduction techniques for BCIs should describe not only the signal quality but also the discriminative performance of the extracted components to demonstrate their efficacy in BCIs. The performances of proposed artifact reduction techniques in most previous studies were evaluated and ranked based on a metric (e.g., signal-to-noise





ratio and correlation) that indicates how the signal quality of the estimated neuronal sources was preserved (Islam et al., 2016). We do not know the original (true) neuronal sources of EEG observations; thus, synthetic data whose pseudo-neuronal/pseudo-artifactual sources and mixing process are known were usually used to calculate the metric (Chen et al., 2019; Mucarquer et al., 2019). After the quantified evaluation

of signal quality in the estimated sources through the proposed artifact reduction technique, the separation ability for real data is qualitatively shown (Blum et al., 2019; Kanoga et al., 2019a). However, the evaluation of the discriminative performance of the remaining sources (extracted components) is not a major/standard quantitative one in this field. To implement an artifact reduction technique into the BCI framework, it is more



**FIGURE 5** | Artifact-contaminated EEG epoch and artifact-reduced EEG epochs estimated by ICA, IVA, and ILRMA in the ERP paradigm.

important to know which aspect of the technique is the most crucial to the classification/identification performance. In this study, we demonstrated improved MI- and SSVEP-based BCI performances using our proposed technique, which represents common and recurrent properties of artifactual waveforms into trials over all classes in low-rank bases and automatically removes them. Except for session 2 in the MI-based BCI performance, our proposed method showed over 70% averaged accuracies (Table 2), which is required for satisfactory BCI control (Sellers et al., 2006). When we consider the time latency during an MI period and change the starting time point of MI features from fixed to flexible by using time window selection algorithm such as correlation-based time window selection (Feng et al., 2018), the average performance of the MI-based BCI might be improved. Furthermore, using other kind of feature extraction method such as sparse FBCSP (Zhang et al., 2015) is one of good solution to improve the MI-based BCI performance because the method automatically chose the filter bands with superior accuracies compared with FBCSP. In the case of the ERP-based BCI, ICA was already effective enough. Therefore, the superiority of ILRMA could not be confirmed; however, its performance is equivalent to that of ICA. Although this paper did not present the quantitative signal quality of the estimated neuronal

sources because we did not prepare synthetic data to avoid the artificial bias of neuronal characteristics (all EEG observations have unclear individual differences such as amplitudes and latencies, so we could not easily predict the features and generate pseudo/synthetic data), the classification/identification results obtained with three well-known BCI paradigms should be helpful information for practitioners and implementers.

### 6.3. Limitations

The results obtained using the DETECT toolbox were treated as the grand truth. However, the muscular label reflected the characteristics of “clenching” and “flexing both arms.” Other types of muscular artifacts, such as “changing head direction,” may not be extracted as artifactual epochs. In addition, the “100% accuracy” was sometimes calculated using only one testing epoch although other accuracies were calculated using more epochs (e.g., 20). The comparison of artifact reduction techniques was fair because the number of artifact-contaminated epochs was the same over the factor. However, the comparisons of BCI paradigms and multiday effects were not fair: each factor has different numbers of artifact-contaminated epochs. Evaluating the techniques as fairly as possible by using the DETECT

toolbox is very difficult. For solving this problem, an artifact-contaminated EEG dataset with multiple types of intensity-manipulated artifacts is required in this research field to enable rapid developments in artifact reduction techniques for BCIs.

## 6.4. Future Works

Further improvement of ILRMA-based artifact reduction techniques is expected through the introduction of an identification algorithm for decomposed frequency components and a soft-threshold-like wavelet-enhanced ICA (Castellanos and Makarov, 2006). Despite the fact that ILRMA decomposes the STFT of the original signals up to each frequency bin, our automatic processing architecture reconstructs artifact-reduced signals by replacing artifactual source(s) with zeros (replacing entire frequency bins with zeros) to adapt the ICLabel classifier, which needs time-series ICs. In other words, a lot of neural information is lost in the reduction step. Signal reconstruction should be made more sophisticated by considering the effective frequency band adjusted to the BCI paradigm.

Moreover, we need further investigations of artifact reduction methods in practical situations such as using wearable devices that have small number of channels (in an extreme case, the number of channels is only one) for EEG measurements. In such situation, the performance of artifact reduction techniques will change and might be decreased. Recent studies attempt to propose a generic artifact removal algorithm (Chen et al., 2019). Unlike the time-domain algorithm, frequency-domain methods (i.e., IVA and ILRMA) can separate single-channel data if the differences in data-driven spectral basis functions can be learned well (Kanoga and Mitsukura, 2014; Kanoga et al., 2019a). Thus, we will investigate our proposed algorithm in practical

situations and extend it as a generic and user-friendly algorithm for reducing artifacts from EEG data.

The ICA family, including IVA and ILRMA, represents EEG observations through linear combinations of sources based on a time-invariant demixing matrix; the trained demixing matrix may sometimes cause instability through inter-/intrasubject variabilities. By integrating with a transfer learning algorithm (Pan and Yang, 2009; Tan et al., 2018), relearning from the general filter (demixing matrix) to the user-specific filter according to the data while performing online processing could potentially reduce the variability and provide more convenient and practical BCIs.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: [http://gigadb.org/dataset/view/id/100542/File\\_page/1](http://gigadb.org/dataset/view/id/100542/File_page/1).

## AUTHOR CONTRIBUTIONS

SK obtained the initial idea for this study. All authors contributed to the planning and design of the study. SK and TH analyzed the data and interpreted the results. SK compiled the first draft of the article. All authors participated in revisions to finalize the draft of the manuscript.

## FUNDING

This research was supported in part by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

## REFERENCES

- Amari, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Comput.* 10, 251–276. doi: 10.1162/089976698300017746
- Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). “Filter bank common spatial pattern (FBCSP) in brain-computer interface,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (Hong Kong: IEEE), 2390–2397.
- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871. doi: 10.1126/science.273.5283.1868
- Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45, 434–444. doi: 10.1109/78.554307
- Blum, S., Jacobsen, N., Bleichner, M. G., and Debener, S. (2019). A Riemannian modification of artifact subspace reconstruction for EEG artifact handling. *Front. Hum. Neurosci.* 13:141. doi: 10.3389/fnhum.2019.00141
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Budd, J. M., and Kisvárdy, Z. F. (2001). Local lateral connectivity of inhibitory clutch cells in layer 4 of cat visual cortex (area 17). *Exp. Brain Res.* 140, 245–250. doi: 10.1007/s002210100817
- Castellanos, N. P., and Makarov, V. A. (2006). Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods* 158, 300–312. doi: 10.1016/j.jneumeth.2006.05.033
- Chen, X., Liu, Q., Tao, W., Li, L., Lee, S., Liua, A., et al. (2019). ReMAE: a user-friendly toolbox for removing muscle artifacts from EEG. *IEEE Trans. Instrum. Meas.* 69, 2105–2119. doi: 10.1109/TIM.2019.2920186
- Choi, S., Cichocki, A., Park, H.-M., and Lee, S.-Y. (2005). Blind source separation and independent component analysis: a review. *Neural Inform. Process. Lett. Rev.* 6, 1–57.
- Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Delorme, A., Mullen, T., Kothe, C., Acar, Z. A., Bigdely-Shamlo, N., Vankov, A., et al. (2011). EEGLAB, SIFT, NIFT, BCILAB, and ERICA: New tools for advanced EEG processing. *Comput. Intell. Neurosci.* 2011:10. doi: 10.1155/2011/130714
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004
- Dimigen, O. (2019). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *Neuroimage* 207:116117. doi: 10.1016/j.neuroimage.2019.116117
- Duong, N. Q., Vincent, E., and Gribonval, R. (2010). Under-determined reverberant audio source separation using a full-rank spatial covariance model. *IEEE Trans. Audio Speech Lang. Process.* 18, 1830–1840. doi: 10.1109/TASL.2010.2050716



- Fazli, S., Popescu, F., Danóczy, M., Blankertz, B., Müller, K.-R., and Grozea, C. (2009). Subject-independent mental state classification in single trials. *Neural Netw.* 22, 1305–1312. doi: 10.1016/j.neunet.2009.06.003
- Feng, J., Yin, E., Jin, J., Saab, R., Daly, I., Wang, X., et al. (2018). Towards correlation-based time window selection method for motor imagery BCIs. *Neural Netw.* 102, 87–95. doi: 10.1016/j.neunet.2018.02.011
- Hagemann, D., and Naumann, E. (2001). The effects of ocular artifacts on (lateralized) broadband power in the EEG. *Clin. Neurophysiol.* 112, 215–231. doi: 10.1016/S1388-2457(00)00541-1
- Halliday, D. M., Conway, B. A., Farmer, S. F., and Rosenberg, J. R. (1998). Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans. *Neurosci. Lett.* 241, 5–8. doi: 10.1016/S0304-3940(97)00964-6
- Hiroe, A. (2006). “Solution of permutation problem in frequency domain ICA, using multivariate probability density functions,” in *International Conference on Independent Component Analysis and Signal Separation* (Charleston: Springer), 601–608. doi: 10.1007/11679363\_75
- Hunter, D. R., and Lange, K. (2000). Quantile regression via an mm algorithm. *J. Comput. Graph. Stat.* 9, 60–77. doi: 10.1080/10618600.2000.10474866
- Hyvärinen, A., and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9, 1483–1492. doi: 10.1162/neco.1997.9.7.1483
- Islam, M. K., Rastegarnia, A., and Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: a review. *Neurophysiol. Clin.* 46, 287–305. doi: 10.1016/j.neucli.2016.07.002
- James, C. J., and Hesse, C. W. (2004). Independent component analysis for biomedical signals. *Physiol. Meas.* 26:R15. doi: 10.1088/0967-3334/26/1/R02
- Jiang, X., Bian, G.-B., and Tian, Z. (2019). Removal of artifacts from EEG signals: a review. *Sensors* 19:987. doi: 10.3390/s19050987
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178. doi: 10.1111/1469-8986.3720163
- Kanoga, S., Kanemura, A., and Asoh, H. (2019a). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing* 347, 240–250. doi: 10.1016/j.neucom.2019.02.060
- Kanoga, S., and Mitsukura, Y. (2014). Eye-blink artifact reduction using 2-step nonnegative matrix factorization for single-channel electroencephalographic signals. *J. Signal Process.* 18, 251–257. doi: 10.2299/jsp.18.251
- Kanoga, S., and Mitsukura, Y. (2017). “Review of artifact rejection methods for electroencephalographic systems,” in *Electroencephalography* 69, 69–89. doi: 10.5772/68023
- Kanoga, S., Nakanishi, M., and Mitsukura, Y. (2016). Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram. *Neurocomputing* 193, 20–32. doi: 10.1016/j.neucom.2016.01.057
- Kanoga, S., Nakanishi, M., Murai, A., Tada, M., and Kanemura, A. (2019b). Robustness analysis of decoding SSVEPs in humans with head movements using a moving visual flicker. *J. Neural Eng.* 17:016009. doi: 10.1088/1741-2552/ab5760
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain* 106, 643–653. doi: 10.1093/brain/106.3.643
- Kaufmann, T., Schulz, S., Grünzinger, C., and Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *J. Neural Eng.* 8:056016. doi: 10.1088/1741-2560/8/5/056016
- Kim, T., Attias, H. T., Lee, S.-Y., and Lee, T.-W. (2006a). Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. Audio Speech Lang. Process.* 15, 70–79. doi: 10.1109/TASL.2006.872618
- Kim, T., Lee, I., and Lee, T.-W. (2006b). “Independent vector analysis: definition and algorithms,” in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA: IEEE), 1393–1396. doi: 10.1109/ACSSC.2006.354986
- Kitamura, D., Ono, N., and Saruwatari, H. (2017). “Experimental analysis of optimal window length for independent low-rank matrix analysis,” in *2017 25th European Signal Processing Conference (EUSIPCO)* (Kos: IEEE), 1170–1174. doi: 10.23919/EUSIPCO.2017.8081392
- Kitamura, D., Ono, N., Sawada, H., Kameoka, H., and Saruwatari, H. (2016). Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. *IEEE/ACM Trans. Audio Speech Lang. Process.* 24, 1622–1637. doi: 10.1109/TASLP.2016.2577880
- Lawhern, V., Hairston, W. D., and Robbins, K. (2013). Detect: A MATLAB toolbox for event detection and identification in time series, with applications to artifact detection in EEG signals. *PLoS ONE* 8:e62944. doi: 10.1371/journal.pone.0062944
- Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788. doi: 10.1038/44565
- Lee, M.-H., Fazli, S., Kim, K.-T., and Lee, S.-W. (2016). “Development of an open source platform for brain-machine interface: OpenBMI” in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)* (Yongpyong: IEEE), 1–2. doi: 10.1109/IWW-BCI.2016.7457440
- Lee, M.-H., Kwon, O.-Y., Kim, Y.-J., Kim, H.-K., Lee, Y.-E., Williamson, J., et al. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience* 8:giz002. doi: 10.1093/gigascience/giz002
- Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719
- Lin, Z., Zhang, C., Wu, W., and Gao, X. (2006). Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.* 53, 2610–2614. doi: 10.1109/TBME.2006.886577
- Mohammadpour, M., and Rahmani, V. (2017). “A hidden Markov model-based approach to removing EEG artifact,” in *2017 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)* (Qazvin: IEEE), 46–49. doi: 10.1109/CFIS.2017.8003655
- Mucarquer, J. A., Prado, P., Escobar, M.-J., El-Deredy, W., and Zañartu, M. (2019). Improving EEG muscle artifact removal with an EMG array. *IEEE Trans. Instrum. Meas.* 69, 815–824. doi: 10.1109/TIM.2019.2906967
- Odena, A. (2016). Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*.
- Ogino, M., Kanoga, S., Muto, M., and Mitsukura, Y. (2019). Analysis of prefrontal single-channel EEG data for portable auditory ERP-based brain-computer interfaces. *Front. Hum. Neurosci.* 13:250. doi: 10.3389/fnhum.2019.00250
- Ono, N. (2011). “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (New Paltz, NY: IEEE), 189–192. doi: 10.1109/ASPAA.2011.6082320
- Onton, J., and Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Prog. Brain Res.* 159, 99–120. doi: 10.1016/S0079-6123(06)59007-7
- Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Parini, S., Maggi, L., Turconi, A. C., and Andreoni, G. (2009). A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. *Comput. Intell. Neurosci.* 2009:864564. doi: 10.1155/2009/864564
- Pfurtscheller, G., and Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8
- Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829
- Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* 198, 181–197. doi: 10.1016/j.neuroimage.2019.05.026
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr. Clin. Neurophysiol.* 20, 238–248. doi: 10.1016/0013-4694(66)90088-5
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). “Improved techniques for training GANs,” in *Advances in Neural Information Processing Systems* (Barcelona), 2234–2242.
- Sawada, H., Kameoka, H., Araki, S., and Ueda, N. (2013). Multichannel extensions of non-negative matrix factorization with complex-valued data. *IEEE Trans. Audio Speech Lang. Process.* 21, 971–982. doi: 10.1109/TASL.2013.2239990
- Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2006). A P300 event-related potential neural-computer interface (BCI): the



- effects of matrix size and inter stimulus interval on performance. *Biol. Psychol.* 73, 242–252. doi: 10.1016/j.biopsycho.2006.04.007
- Squires, K. C., Wickens, C., Squires, N. K., and Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science* 193, 1142–1146. doi: 10.1126/science.959831
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). “A survey on deep transfer learning,” in *International Conference on Artificial Neural Networks* (Rhodes: Springer), 270–279. doi: 10.1007/978-3-030-01424-7\_27
- Tian, Z., Ling, B. W.-K., Zhou, X., Lam, R. W.-K., and Teo, K.-L. (2020). Suppressing the spikes in electroencephalogram via an iterative joint singular spectrum analysis and low-rank decomposition approach. *Sensors* 20:341. doi: 10.3390/s20020341
- Urigüen, J. A., and Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *J. Neural Eng.* 12:031001. doi: 10.1088/1741-2560/12/3/031001
- Vigario, R., and Oja, E. (2008). BSS and ICA in neuroinformatics: from current practices to open challenges. *IEEE Rev. Biomed. Eng.* 1, 50–61. doi: 10.1109/RBME.2008.2008244
- Vourvopoulos, A. T., Jorge, C., Abreu, R., Figueiredo, P., Fernandes, J.-C., and Bermúdez i Badia, S. (2019). Efficacy and brain imaging correlates of an immersive motor imagery BCI-driven VR system for upper limb motor rehabilitation: a clinical case report. *Front. Hum. Neurosci.* 13:244. doi: 10.3389/fnhum.2019.00244
- Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., and Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *Int. J. Psychophysiol.* 53, 105–119. doi: 10.1016/j.ijpsycho.2004.03.007
- Wang, M., Li, R., Zhang, R., Li, G., and Zhang, D. (2018). A wearable SSVEP-based BCI system for quadcopter control using head-mounted device. *IEEE Access* 6, 26789–26798. doi: 10.1109/ACCESS.2018.2825378
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15, 70–73. doi: 10.1109/TAU.1967.1161901
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3
- Yeom, S.-K., Fazli, S., Müller, K.-R., and Lee, S.-W. (2014). An efficient ERP-based brain-computer interface using random set presentation and face familiarity. *PLoS ONE* 9:e111157. doi: 10.1371/journal.pone.0111157
- Zhang, Y., Zhou, G., Jin, J., Wang, X., and Cichocki, A. (2015). Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. *J. Neurosci. Methods* 255, 85–91. doi: 10.1016/j.jneumeth.2015.08.004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kanoga, Hoshino and Asoh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.