

# Real-Time Adaptive EEG Source Separation Using Online Recursive Independent Component Analysis

Sheng-Hsiou Hsu, *Student Member, IEEE*, Tim R. Mullen, *Member, IEEE*, Tzzy-Ping Jung, *Fellow, IEEE*, and Gert Cauwenberghs, *Fellow, IEEE*

**Abstract**—Independent component analysis (ICA) has been widely applied to electroencephalographic (EEG) biosignal processing and brain–computer interfaces. The practical use of ICA, however, is limited by its computational complexity, data requirements for convergence, and assumption of data stationarity, especially for high-density data. Here we study and validate an optimized online recursive ICA algorithm (ORICA) with online recursive least squares (RLS) whitening for blind source separation of high-density EEG data, which offers instantaneous incremental convergence upon presentation of new data. Empirical results of this study demonstrate the algorithm's: 1) suitability for accurate and efficient source identification in high-density (64-channel) realistically-simulated EEG data; 2) capability to detect and adapt to nonstationarity in 64-ch simulated EEG data; and 3) utility for rapidly extracting principal brain and artifact sources in real 61-channel EEG data recorded by a dry and wearable EEG system in a cognitive experiment. ORICA was implemented as functions in BCILAB and EEGLAB and was integrated in an open-source Real-time EEG Source-mapping Toolbox (REST), supporting applications in ICA-based online artifact rejection, feature extraction for real-time biosignal monitoring in clinical environments, and adaptable classifications in brain–computer interfaces.

**Index Terms**—Biomedical signal processing, blind source separation, electroencephalography (EEG), independent component analysis (ICA), nonstationarity.

## I. INTRODUCTION

**I**NDPENDENT component analysis (ICA), as a means for blind source separation (BSS), has enjoyed great success in telecommunications and biomedical signal processing [1]. In

biomedical applications, such as scalp electroencephalography (EEG), ICA methods have been widely used to separate artifacts such as eye blinks and muscle activities [2] and to study brain activities [3]. For example, ICA can extract fetal electrocardiography (ECG) from maternal abdomen electrode recordings [4], and it can also isolate pathological activities associated with disease states of epilepsy [5]. In addition, applying ICA to remove task-irrelevant activities and reduce dimensionality of data can improve the performance of Brain-Computer Interfaces (BCI) [6].

The application of ICA to EEG data is justified by a reasonable assumption that multi-channel scalp EEG signals arise as a mixture of weakly dependent latent non-Gaussian sources [7]. Although several ICA algorithms have been developed [1] to learn these sources from channel mixtures, most of the algorithms require access to large amount of training data and are only suitable for offline applications. Furthermore, the offline ICA algorithms commonly assume spatiotemporal stationarity of the data, as in the widely used Infomax ICA [8] and FastICA [9] algorithms. For a few ICA methods that allow nonstationarity such as Adaptive Mixture ICA [10], they are computationally expensive. In many real-world applications, including real-time functional neuroimaging [11], artifact rejection and adaptive BCI [6], online (sequential) source separation methods are needed. Desirable properties of an online method include fast convergence, real-time computational performance, and adaptivity to nonstationary data.

Many existing online ICA methods are listed in Table I. Two major learning rules are least-mean-squares (LMS) and recursive-least-squares (RLS) methods. LMS-type algorithms use stochastic gradient descent approaches and are computationally simple, but they require careful selection of an appropriate learning rate for stable convergence. Examples include Equivariant Adaptive Separation via Independence (EASI) [12] and Natural Gradient (NG) [13] methods. RLS-type algorithms accumulate past data in an exponentially decaying fashion and use Sherman-Morrison matrix inversion to achieve higher convergence rate and better tracking capability, yet require complex computation [14], [15]. This category includes the RLS approach of Nonlinear PCA (RLS-NPCA) [16], Iterative Inversion [17], and Natural Gradient-based RLS (NG-RLS) [18]. Alternatively, Online Recursive ICA (ORICA) [19] gives an RLS-type recursive rule by solving a fixed-point approximation and has been shown to exhibit fast convergence and low computational complexity [20]. Readers can refer to [15], [17] and [21] for theoretical relationships and comparisons between the above methods.

Manuscript received May 01, 2015; revised October 20, 2015; accepted December 03, 2015. Date of publication December 17, 2015; date of current version March 04, 2016. This work was supported in part by a gift by the Swartz Foundation (Old Field, NY), by the Army Research Laboratory under Cooperative Agreement Number W911NF-10-2-0022, by NIH grant 1R01MH084819-03 and NSF EFRI-M3C 1137279.

S.-H. Hsu and T.-P. Jung are with Department of Bioengineering, Swartz Center for Computational Neuroscience, and Institute for Neural Computation, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: shh078@ucsd.edu).

T. R. Mullen was with Department of Cognitive Science, Swartz Center for Computational Neuroscience, and Institute for Neural Computation, University of California at San Diego, La Jolla, CA 92093 USA. He is now with Qusp Labs (formerly Syntrogi Labs), San Diego, CA 92111 USA.

G. Cauwenberghs is with Department of Bioengineering and Institute for Neural Computation, University of California at San Diego, La Jolla, CA 92093 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNSRE.2015.2508759

TABLE I  
COMPARISON OF ONLINE ICA METHODS

Name	Author	Year	Learning rule and optimization method	Pre-whitening
EASI	Cardoso et al. [12]	1996	Relative gradient-based (LMS) that max. kurtosis	LMS
NG	Amari et al. [13]	1996	Natural gradient-based (LMS) that min. mutual information	no
RLS-NPCA	Karhunen et al. [16]	1997	RLS that min. LSE of NPCA criterion	PCA
Iterative Inversion	Cruces-Alvarez et al. [17]	2000	Quasi-Newton method with iterative inversion that decorrelates high order statistics	no
NG-RLS	Zhu et al. [18]	2004	RLS with natural gradient that min. LSE of NPCA criterion	RLS
ORICA	Akhtar et al. [19]	2012	Recursive rule with iterative inversion from fixed-point solution of Infomax with natural gradient	no

The aforementioned papers focused on theoretical derivations and proofs of convergence and only demonstrated applications of the methods to low-density data (fewer than 10 channels) and simulated “toy” examples such as sinusoidal and square waves. When the number of channels and sources increases, many existing algorithms exhibited slow convergence and poor real-time performance [9]. In a recent study, a real-time online ICA method for high-density EEG was proposed [20]. The method was compared with other offline ICA methods, and its stability and steady-state performance were analyzed in [22].

Additionally, an important advantage of online ICA methods is their ability to adapt to spatiotemporal nonstationary data, a common occurrence in real-world applications. For instance, spatial nonstationarity in the ICA (un)mixing matrix can arise as a consequence of location shifts in either sensors or sources, or changes in electrode impedances. However, few of the online ICA methods have been carefully studied under nonstationary conditions. Further investigation is needed to characterize algorithmic performance and optimal parameter selection using nonstationary simulated and real EEG data.

In this study, we extend ORICA as formulated in [19] and [20], and the contributions are three-fold. Firstly, we demonstrate ORICA's suitability for accurate and efficient source identification in a realistic simulation of stationary 64-channel EEG data. Specifically, we include a serial orthogonalization step of the unmixing matrix in the ORICA algorithm, and we systematically examine the impact of parameters such as the forgetting factor and block sizes for pre-whitening and ORICA on algorithmic performance. Secondly, we examine ORICA's capability to adaptively decompose spatially nonstationary 64-channel EEG data corresponding to abrupt displacements of electrodes, a common source of spatial nonstationarity in real-world mobile applications. We introduce a nonstationarity index and an adaptation approach for nonstationarity detection and online adaptation. Thirdly, we evaluate ORICA's real-world applicability for rapidly extracting principal brain and artifact sources using 61-channel real EEG data recorded from a subject performing an Eriksen flanker task [23]. We demonstrate that ORICA and offline Infomax ICA [24] obtain comparable results in terms of extracting informative independent components (ICs) and their single trial and averaged event-related potentials (ERPs), yet ORICA can learn the ICs online with less than half of the data. Finally, the proposed ORICA pipeline is made freely available as functions supported in BCILAB [25] and EEGLAB [26], and it is also integrated in an open-source Real-time EEG Source-mapping Toolbox (REST) [27].

## II. METHODS

Standard ICA assumes a linear generative model  $\mathbf{x} = \mathbf{A}\mathbf{s}$ , where  $\mathbf{x}$  represents scalp EEG observations,  $\mathbf{s}$  contains unknown sources, and  $\mathbf{A}$  is an unknown square mixing matrix. The objective is to learn an unmixing matrix  $\mathbf{B} = \mathbf{A}^{-1}$  such that the sources are recovered exactly, up to an unknown permutation and scaling matrix, by  $\mathbf{y} = \mathbf{B}\mathbf{x}$ , where  $\mathbf{y}$  represent the recovered source activations. A column of  $\mathbf{B}^{-1}$  represents the spatial distribution of a source over all channels, often referred to as a “component map.”

It is desirable to optimize the ICA contrast function, a measurement of the degree of independence between sources such as kurtosis or mutual information, under the decorrelation constraint  $\mathbf{R}_y = \mathbf{E}[\mathbf{y}\mathbf{y}^T] = \mathbf{I}$ . Hence the separating process can be factored into two stages as  $\mathbf{B} = \mathbf{W}\mathbf{M}$ , where  $\mathbf{M}$  is the whitening matrix that decorrelates the data and  $\mathbf{W}$  is the weight (preferably orthogonal) matrix that optimizes the ICA contrast function [12], [18]. Serial update rules of  $\mathbf{M}$  and  $\mathbf{W}$  and detailed features are presented in the following subsections.

### A. Online Recursive-Least-Squares (RLS) Pre-Whitening

Pre-whitening (decorrelating) the data reduces the number of independent parameters an ICA update must learn, and can improve convergence [1]. Pre-whitening may be efficiently carried out in an online RLS-type learning rule [18]

$$\mathbf{M}_{n+1} = \mathbf{M}_n + \frac{\lambda_n}{1 - \lambda_n} \left[ \mathbf{I} - \frac{\mathbf{v}_n \cdot \mathbf{v}_n^T}{1 + \lambda_n (\mathbf{v}_n^T \cdot \mathbf{v}_n - 1)} \right] \mathbf{M}_n \quad (1)$$

where  $n$  is the number of iterations,  $\mathbf{M}_n$  is the whitening matrix,  $\mathbf{v}_n = \mathbf{M}_n \mathbf{x}_n$  is the decorrelated data,  $\lambda_n$  is a forgetting factor, and  $\mathbf{I}$  is the identity matrix. A nonoverlapping block of data  $\mathbf{x}_n$  with a block size  $L_{\text{white}}$  is used at each iteration to reduce the computational load and to increase the robustness of the estimated correlation matrix  $\mathbf{v}_n \mathbf{v}_n^T$ . This RLS-type whitening rule exhibits faster convergence than LMS whitening methods [18].

### B. Online Recursive ICA (ORICA)

The ORICA algorithm can be derived from an incremental update form of the natural gradient learning rule of Infomax ICA [28]

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \left[ \mathbf{I} - \mathbf{f}(\mathbf{y}_n) \cdot \mathbf{y}_n^T \right] \mathbf{W}_n \quad (2)$$

where  $\mathbf{y}_n = \mathbf{W}_n \mathbf{v}_n$ ,  $\eta$  is a learning rate, and  $f(\cdot)$  is a nonlinear activation function. In the limit of a small  $\eta$  and assuming a

fixed  $f(\cdot)$ , the convergence criterion  $\langle \mathbf{f}(\mathbf{y}) \cdot \mathbf{y}^T \rangle = \mathbf{I}$  leads to a fixed-point solution in an iterative inversion form [19]

$$\mathbf{W}_{n+1}^+ = (1 - \lambda_n)\mathbf{W}_n^+ + \lambda_n \mathbf{v}_n \cdot \mathbf{f}^T(\mathbf{y}_n) \quad (3)$$

where  $\mathbf{W}_n^+$  is the Moore-Penrose pseudoinverse of  $\mathbf{W}_n$  and  $\lambda_n$  is a forgetting factor for an exponentially weighted series of updates. It should be noted that  $\lambda_n$  differs from  $\eta$ , which is the step size for stochastic gradient optimization.

Following [19], applying the Sherman-Morrison matrix inversion formula to (3), the final online recursive learning rule becomes

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\lambda_n}{1 - \lambda_n} \left[ \mathbf{I} - \frac{\mathbf{y}_n \cdot \mathbf{f}^T(\mathbf{y}_n)}{1 + \lambda_n (\mathbf{f}^T(\mathbf{y}_n) \cdot \mathbf{y}_n - 1)} \right] \mathbf{W}_n. \quad (4)$$

The near-identical forms of (4) and (1) allow us to understand ORICA as a nonlinear (or kernel) form of the RLS whitening filter: ORICA's use of nonlinearity  $f(\cdot)$  allows for independence of sources for moments above second order.

Following the (4), the orthogonality of the weight  $\mathbf{W}$  is not guaranteed. To preserve the decorrelation property of recovered source activities  $\mathbf{y} = \mathbf{W}\mathbf{M}\mathbf{x}$  and maintain learning stability, we apply an orthogonal transformation to  $\mathbf{W}$  after each ICA update

$$\mathbf{W}_{n+1} \leftarrow (\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^{-1})\mathbf{W}_{n+1} \quad (5)$$

where  $\mathbf{D}$  and  $\mathbf{V}$  contain, respectively, the eigenvalues and eigenvectors of  $\mathbf{W}_{n+1}\mathbf{W}_{n+1}^T$ . Note that this orthogonalization step is costly compared to the ORICA update. A possible alternative would be to reformulate the ORICA update rule under the orthogonal constraint or combine the serial whitening and weight updates into a single update rule [12].

1) *Block-Update Rule*: The typical single measurement vector approach [29] requires application of the update rule (4) for each data sample, which can be computationally expensive, particularly for the commonly-used MATLAB (The Mathworks, Natick, MA, USA) runtime environment. To reduce the computational load and ensure consistent real-time performance, we may adopt an multiple measurement vector approach [29] and perform updates on short blocks of samples. To achieve this without loss of accuracy, we solve (4) for time index  $l = n$  to  $l = n + L - 1$ , assuming  $\mathbf{y}_l$  is approximated as  $\mathbf{W}_n \mathbf{v}_l$  and  $\lambda_l$  is small. This leads to a block-update rule [20]

$$\mathbf{W}_{n+L} \approx \left( \prod_{l=n}^{n+L-1} \frac{1}{1 - \lambda_l} \right) \cdot \left[ \mathbf{I} - \sum_{l=n}^{n+L-1} \frac{\mathbf{y}_l \cdot \mathbf{f}^T(\mathbf{y}_l)}{\frac{1 - \lambda_l}{\lambda_l} + \mathbf{f}^T(\mathbf{y}_l) \cdot \mathbf{y}_l} \right] \mathbf{W}_n. \quad (6)$$

In this form, the sequence of updates can be vectorized for fast computation. Note that (6) appropriately accounts for the decaying forgetting factor at each time point. This keeps the approximation error to a minimum.

2) *Forgetting Factor*: The forgetting factor  $\lambda$  determines an effective length of a time window wherein data are aggregated. A large value of  $\lambda$  corresponds to a short window length. In

this case, much heavier weights are applied to new data than past data, yielding fast adaptation and convergence yet large errors and variability. As a general rule, a large  $\lambda$  is preferred during initial learning to promote fast convergence; a small  $\lambda$  is suggested at convergence to minimize variance. To this end, we adopt the forgetting factor with time-varying annealing defined in [19]

$$\lambda_n = \frac{\lambda_0}{n^\gamma} \quad (7)$$

where  $\lambda_0$  is a fixed initial forgetting factor and  $\gamma$  determines the rate of exponential decay of  $\lambda$  towards zero as a function of time. The same forgetting factor is applied to the RLS whitening filter, although theoretically it can be different.

3) *Nonlinear Function*: The choice of nonlinearity  $f(\cdot)$  depends on the probability distribution of the sources. Lee *et al.* [28] proposed an extended Infomax ICA algorithm that adopts distinct activation functions to separate sub-Gaussian and super-Gaussian sources based on an estimate of source kurtosis. Here we follow [1], [19] and choose the component-wise nonlinearity  $f(y) = -2 \tanh(y)$  for super-Gaussian sources and  $f(y) = \tanh(y) - y$  for sub-Gaussian sources.

4) *Number of Sub- and Super-Gaussian Sources*: While approaches for adaptively selecting  $f(y)$  within ORICA have been proposed [19], these are heuristic and presently lack convergence proofs. In practice, we found that both convergence speed and run-time performance were improved by preassuming a fixed number of sub- and super-Gaussian sources. The detail of selecting an appropriate number was discussed in Section V and the effect of inaccurate assumptions was explored in Section IV.

### C. Nonstationarity Detection

In this study, the nonstationarity refers to any changes in the ICA model  $\mathbf{x} = \mathbf{A}\mathbf{s}$ , including spatial nonstationarity of the mixing process between sources and the measurements  $\mathbf{A}$  and temporal nonstationarity of probability distributions of source activities  $\mathbf{s}$ . Hence the nonstationarity might arise from switching of active brain sources, transient muscle activities, sensor displacement, or impedance changes. Our goal is to propose a generic approach to detect and adapt to the nonstationarity in EEG data.

1) *Nonstationarity Index*: As previously described, ORICA is derived from a fixed point solution to the convergence criterion  $\langle \mathbf{y} \cdot \mathbf{f}(\mathbf{y})^T \rangle = \mathbf{I}$ , reflecting independence of sources for moments above second order. Violation of this criterion once an algorithm reaches steady-state can be interpreted as a change in the latent source or mixture distribution, which leads us to define the following heuristic nonstationarity index:

$$\delta_{ns} = \left\| \mathbf{y}_n \mathbf{f}^T(\mathbf{y}_n) - \mathbf{I} \right\|_F \quad (8)$$

where  $\|\cdot\|_F$  represents the Frobenius norm and  $n$  is the current sample point. After ICA decomposition converges,  $\delta_{ns}$  would remain small when data are stationary, while  $\delta_{ns}$  would increase and fluctuate when data are nonstationary.

2) *Adaptation of the Forgetting Factor*: If the nonstationarity index  $\delta_{ns}$  increases above a threshold, we may interpret this as evidence of a change in the latent mixing matrix and increase the RLS forgetting factor allowing ORICA to more rapidly adapt

to the new mixing matrix. In this study, the threshold value was heuristically chosen to be a percentage (e.g., 1%–10%) of initial  $\delta_{n_s}$ , which was several standard deviations above the mean of  $\delta_{n_s}$  at convergence in the simulated stationary data. Once  $\delta_{n_s}$  reached the threshold, we increased the forgetting factor to its initial value.

### III. MATERIALS

#### A. Data Collection

The performance of ORICA was evaluated under simulated and real-world conditions. Previous works of online ICA mostly used “toy” simulations with artificially constructed sources (sinusoids, i.i.d. random data, etc.), stationary mixing matrices, and relatively small numbers of channels and sources. Here we generated high-density EEG (64 channels and 64 sources) under more realistic conditions, including use of auto-correlated stochastic sources, realistic source locations and mixing matrices derived from Boundary Element Method (BEM) modeling, and spatial nonstationarity. Simulated data were generated using the EEG simulation module in Source Information Flow Toolbox (SIFT) [30], using an approach similar to [31].

1) *Simulated Spatially Stationary EEG Data*: To test ORICA’s performance in separating stationary EEG sources, we generated 64 super-Gaussian independent source time-series from stationary and random-coefficient order-3 autoregressive (AR-3) models (300 Hz sampling rate, 10-min), assigned each source a random cortical dipole location, and projected these through a zero-noise 3-layer BEM forward model (MNI “Colin27”) with standard 10–20 electrode locations matching the 64-channel Cognionics montage used subsequently for real-world ORICA evaluation. This yielded 64-channel EEG data.

2) *Simulated Spatially Nonstationary EEG Data*: To evaluate ORICA’s capability to adapt to spatial nonstationarity, we simulated abrupt shifts of the electrode montage during continuous recording. We first generated 30 min of temporally stationary AR-3 source data, as described above. The data was partitioned into thirds. For each 10 min segment, 64 channel EEG data was generated using a unique BEM forward (mixing) matrix, corresponding respectively to 1) the standard electrode montage, 2) a 5° anterior cap rotation, and 3) a subsequent 10° posterior cap rotation (5° posterior rotation from standard position). The procedure is illustrated in Fig. 5(a).

3) *Real EEG Data*: One session of high-density EEG data was collected from a 24-year-old right-handed male subject using a 64-channel wearable and wireless EEG headset with dry electrodes (Cognionics, Inc.) [32]. In the 20 min session, the subject performed a modified Eriksen flanker task [23] with a 133 ms delay between flanker and target presentation. The subject was asked to press buttons according to the target stimuli as quickly as possible. Flanker tasks are known to produce robust error-related negativity (ERN, Ne) at frontal-central electrode sites. The goal here is to extract these event-related potential (ERP) components from high-density EEG data in a real-world setting using the proposed ORICA pipeline.

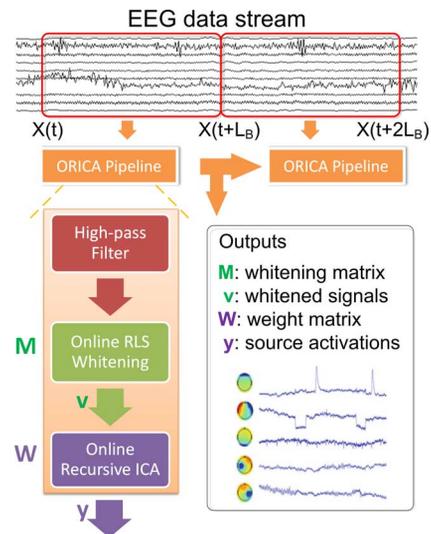


Fig. 1. ORICA pipeline for online EEG data processing.  $X(t)$  is the input data vector at time  $t$  and  $L_B$  is the size of data in the online buffer.

TABLE II  
LIST OF PARAMETERS FOR THE ORICA PIPELINE: IIR HIGH-PASS FILTER (IIR), ONLINE RLS WHITENING FILTER (RLS), AND ONLINE RECURSIVE ICA FILTER (ICA)

Filters	Parameters	Values	Description
IIR	BW	0.2–2 Hz	Transition bandwidth
RLS	$L_{white}$	1~16	Block-average size
RLS	$\lambda_0$	0.995	Initial forgetting factor
ICA	$\gamma$	0.60	Decay rate of forgetting factor
ICA	$L_{ICA}$	1~16	Block-update size
ICA	$n_{sub}$	0	Number of subgaussian sources

#### B. The ORICA Pipeline

As shown in Fig. 1, the ORICA pipeline continuously fetched the streamed data with variable size  $L_B$  in the online buffer and processed the data with the three filters in sequence: a Butterworth IIR high-pass filter, an online RLS whitening filter, and an ORICA filter. The high-pass filter removed the trend and low-frequency drift, ensuring the zero-mean criterion for ICA was satisfied. For each update, the pipeline computed and output the whitening matrix  $\mathbf{M}$  and the weight matrix  $\mathbf{W}$  according to (1) and (6). The next nonoverlapping data chunk was then used for the subsequent update.

The pipeline was implemented and analyzed in a simulated online environment using BCILAB, an open source MATLAB toolbox designed for BCI research [11], [25]. It was initialized with the first second of data segment from the dataset. For the simulated 64-ch stationary data, we investigated the effect of the parameters on empirical convergence, and we used block sizes  $L_{white} = L_{ICA} = 16$  as an example to show the decomposed components since the block sizes returned satisfactory results in the shortest computational time. For the simulated 64-ch nonstationary data and real 61-ch data, we used  $L_{white} = 8$  and  $L_{ICA} = 1$ , which were found to be optimal for the simulated stationary data. Table II summarizes the parameters of the three filters.

### C. Data Processing and Analysis

We applied additional procedures to process and analyze the recorded EEG data from the subject. Firstly, an automatic removal of bad (e.g., flatlined or abnormally correlated) channels was applied prior to the ORICA pipeline using BCILAB routines, which removed 3 out of 64 channels. Secondly, following application of the ORICA pipeline, the source activities were epoched in a  $-400$  to  $600$  ms window time locked to subject's responses (button press), yielding 693 epochs (104 error and 589 correct trials). The epochs were averaged to produce ERPs and were analyzed offline in EEGLAB [26].

### D. Performance Evaluation

1) *Performance Index*: If the ground truth ( $N$ -by- $N$ ) mixing matrix  $\mathbf{A}$  is known, a performance index  $PI$  for assessing quality of source separation can be defined as [33]

$$PI = \frac{1}{N-1} \left[ N - \frac{1}{2} \sum_{i=1}^N \left( \frac{\max_{1 \leq j \leq m} |\mathbf{C}_{ij}(n)|^2}{\sum_{j=1}^N |\mathbf{C}_{ij}(n)|^2} + \frac{\max_{1 \leq j \leq m} |\mathbf{C}_{ji}(n)|^2}{\sum_{j=1}^N |\mathbf{C}_{ji}(n)|^2} \right) \right] \quad (9)$$

where  $\mathbf{C}(n) = \mathbf{W}_n \mathbf{M}_n \mathbf{A}$ . This measures a normalized total cross-talk error of the estimated whitening matrix  $\mathbf{M}$  and weight matrix  $\mathbf{W}$ , accounting for scale and permutation ambiguities. For perfect separation at convergence,  $PI$  approaches zero.

2) *Best-Matched Correlation Coefficients and Hungarian Algorithm*:  $PI$  reflects ORICA's global performance across all components. However, it is also useful to evaluate convergence of individual independent components (ICs), i.e., rows of  $\mathbf{W}$ . One metric is the Pearson correlation between an estimated IC and its counterpart in a "ground truth" weight matrix,  $\mathbf{W}^*$ . Due to permutation ambiguities, a matching algorithm is required for optimal pairing of rows of  $\mathbf{W}$  and  $\mathbf{W}^*$ . This study used the Hungarian method [34] to maximize the sum of absolute pairwise correlations. We used Niclas Borlin's implementation in EEGLAB's *matcorr*( $\cdot$ ) function.

## IV. RESULTS

### A. Simulated 64-ch Stationary EEG Data

1) *Evaluation of the Decomposed Components*: Fig. 2(a) plots the correlation magnitudes between ORICA components and their ground-truth counterparts. The components were sorted such that smaller component ID represented faster convergence. We observed that all components converged, i.e., correlation magnitudes approached 1, by the end of the 10-min session. A common empirical heuristic for separating  $N$  stable ICA sources using Infomax ICA was  $kN^2$ , where  $k > 25$  [24]. For 64 channels, the heuristic time required for convergence amounted to  $64^2 \times 25 = 102400$  samples = 5.7 min with a 300 Hz sample rate. By 5.7 min, 77% (91%) of ICs reached a correlation magnitude of 0.95 (0.8); by 3.5 min, more than half of the ICs reached a correlation magnitude of 0.95.

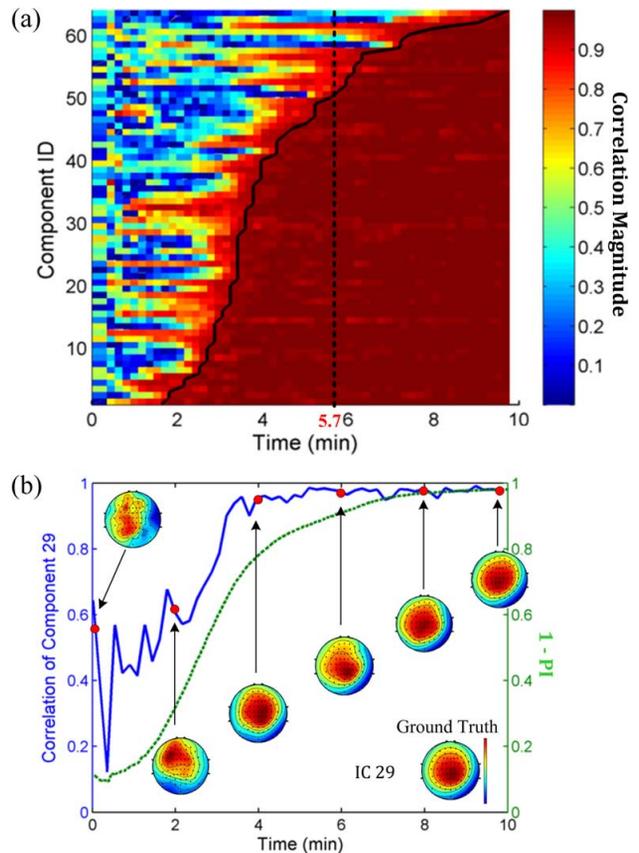


Fig. 2. (a) Evolution of component-wise correlation magnitudes between ORICA-decomposed ICs and ground truth on simulated stationary 64-ch EEG data [20]. ICs sorted with respect to time required to reach a correlation magnitude of 0.95 (solid curve). Dotted line is the heuristic time for separating 64 stable ICs. (b) Evolution of correlation magnitudes (blue) and component maps of a randomly selected IC #29. One minus the performance index (green) is superimposed.

Fig. 2(b) shows the evolution of the component maps of a randomly selected IC #29 and its correlation magnitudes with ground truth. This IC converged to a steady-state correlation magnitude of 0.95 under 4 min. The superimposed global performance index ( $1 - PI$ , in green) exhibited a similar convergence trajectory. Fig. 3 shows 300-ms time-series of four representative ICs reconstructed by ORICA at 3, 6, and 9 min. At 9 min, ORICA correctly reconstructed all the source dynamics with the errors approached 0. At the heuristic time 5.7 min, only ID #58 had not converged. Interestingly, components such as IC #3 converged within 3 min. Both decomposed component maps and recovered source dynamics demonstrated ORICA's suitability for accurate and efficient decomposition of high-density (64-channel) data, albeit with systematic variation in convergence speed.

2) *Effect of ORICA Parameters*: As shown in Fig. 4, we systematically evaluated the effects of four ORICA parameters on convergence. The decay rate of forgetting factor  $\gamma$  had a significant impact on convergence speed, with the fastest convergence for  $\gamma = 0.6$  [Fig. 4(a)]. A sigmoidal profile for the convergence trajectory was observed for  $\gamma \leq 0.6$ , while an exponential decay profile was observed for  $\gamma > 0.6$ . The ORICA block size  $L_{ICA}$  had negligible effect on the convergence for  $L_{ICA} \leq 64$

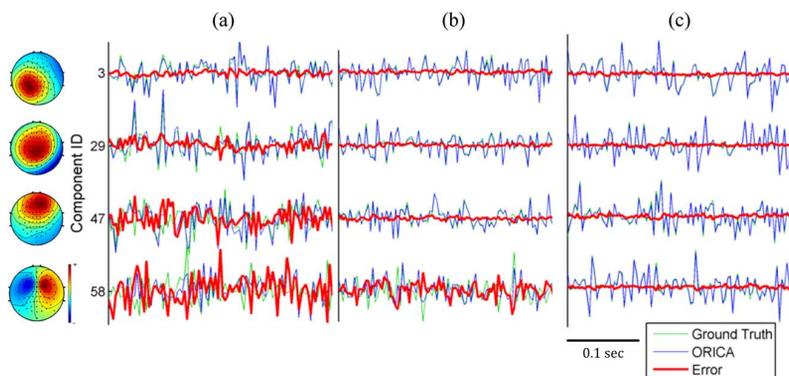


Fig. 3. Source dynamics and the corresponding component maps of four representative components reconstructed by ORICA at (a) 3 min, (b) 6 min, or (c) 9 min of simulated mixed 64-channel stationary EEG data. Reconstructed source dynamic (blue) is superimposed on ground truth (green) with error, i.e., difference, (red). Oscillatory (autocorrelated) and burst-like source dynamics, as well as homogeneous component maps of the four depicted ICs, are characteristic of real EEG source dynamics.

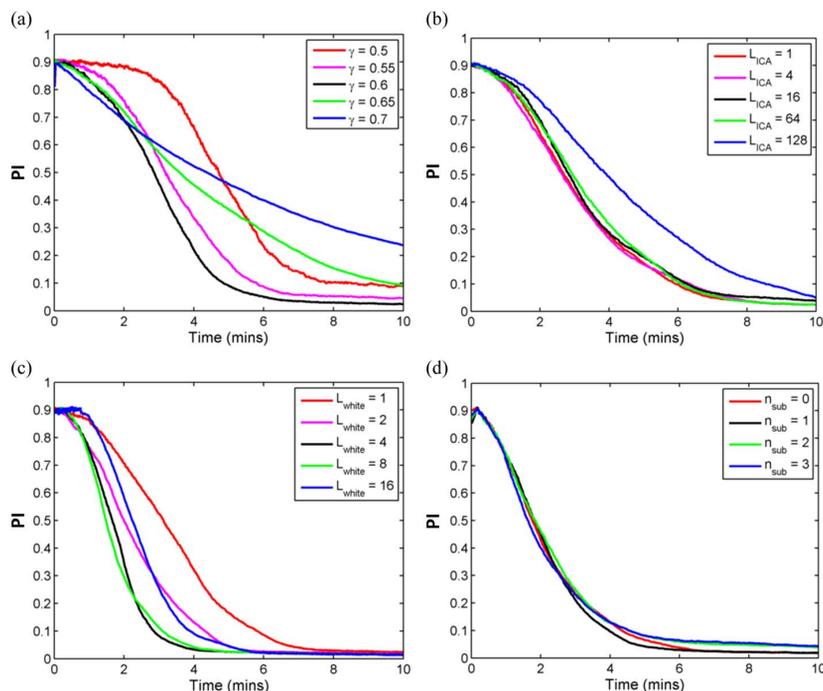


Fig. 4. Effect of ORICA pipeline parameters on convergence trajectory, i.e., performance index over time, applied to 64-ch simulated stationary EEG data. (a) Decay rate  $\gamma$  of forgetting factor, (b) block size  $L_{ICA}$  of ORICA, (c) block size  $L_{white}$  of the online whitening, and (d) pre-assumed number  $n_{sub}$  of sub-Gaussian sources.

[Fig. 4(b)]. This demonstrated the approximation error of the block update rule (6) was negligible for small to moderate block sizes. The block size of online whitening  $L_{white}$  significantly affected the ICA convergence [Fig. 4(c)].  $L_{white}$  between four and eight samples achieved the best performance. Interestingly,  $L_{white} = 1$  was not the optimal value, mainly due to the effect of variable time scale of adjustments in the whitening matrix on the convergence of the subsequent ORICA. The pre-assumed number of subgaussian sources  $n_{sub}$  had little effect on the convergence, within the range of  $n_{sub} = 0 \sim 3$ , with the true number  $n_{sub} = 0$ . The performance of ORICA was rather insensitive to assumptions on the kurtosis of the sources. For the above results, we set  $\gamma = 0.6$ ,  $L_{ICA} = 1$ ,  $L_{white} = 1$ , and  $n_{sub} = 0$  unless otherwise noted.

3) *Quantification of Computational Load*: Table III shows the average execution time required to apply the ORICA pipeline to 1 s of data, computed by averaging the processing rates (data size divided by time) of the incoming data chunks for 1 min. Runtime was uniformly less than 1 s, illustrating the 64-ch data were processed faster than accumulated in the input buffer, and thus the pipeline was capable of real-time operation. The online whitening filter ran 3–9 times faster than ORICA did and the runtime monotonically decreased as the block size increased. The execution time of the ORICA filter was nearly halved as block size doubled when  $L \leq 8$ , with diminishing returns for  $L \geq 16$ . This allows us to balance the tradeoff between runtime of the pipeline and accuracy of the block-update rule.

TABLE III  
AVERAGED EXECUTION TIME (ms) FOR 1 s (300 SAMPLES) 64-CHANNEL DATA USING ONLINE RLS WHITENING AND ORICA WITH DIFFERENT BLOCK SIZES  $L_{\text{white}}$  AND  $L_{\text{ICA}}$

Time (ms)	Block Size							
	1	2	4	8	16	32	64	128
Algorithm								
RLS	35.2±5.2	23.8±7.9	13.5±6.0	8.6±4.3	5.3±1.5	4.1±4.5	3.0±4.2	2.5±4.9
ORICA	332±29	174±37	79.5±14.5	36.3±11.8	21.0±5.2	15.1±7.2	10.6±9.4	6.6±2.6

Run in MATLAB 2012a on a dual-core 2.50 GHz Intel Core i5-3210M CPU with 8 GB RAM.

### B. Simulated 64-ch Nonstationary EEG Data

Fig. 5 plots ORICA's performance in tracking spatial nonstationarity in simulated 64-channel EEG data, with the simulated abrupt shifts of the EEG cap [Fig. 5(a)]. Fig. 5(b) plots the nonstationarity index  $\delta_{ns}$  as a function of time. The index robustly identified changes in the mixing matrix due to cap displacements. Fig. 5(c) plots the performance index of ORICA's decomposition as a function of time. Following the detection of nonstationarity, ORICA's forgetting factor was reset to its initial value, and ORICA smoothly adapted to the new mixing matrix. Fig. 5(d) plots the ground truth and the estimated component maps for a representative IC at different time-points, superimposed on a plot of log-transformed  $PI$ . For this IC, suitable convergence was obtained within 15 min, and improved further over time. The effect of cap rotation was captured by the concomitant shift of the component maps, indicating ORICA's capability to detect and adapt to the spatial abrupt nonstationarity.

### C. Real 61-ch EEG Data From the Flanker Task

Since the ground truth was unknown for real EEG data, we adopted the offline Extended Infomax ICA algorithm [28], as implemented in the EEGLAB [26] function RUNICA, as a "gold standard". The robustness and stability of the algorithm on high-density EEG data had been shown to outperform most blind source separation algorithms [7].

To confirm whether ORICA yielded a comparable solution at convergence (average over the last minute) as RUNICA, we investigated event-related activities of three sets of components with stereotypical fronto-central, occipital, and prefrontal spatial topographies. Fig. 6(a) revealed that the ORICA- and RUNICA-decomposed fronto-central ICs and their characteristic ERN were comparable and consistent with the results from previous studies [11], [35]. Fig. 6(b) shows that averaged occipital visual-evoked potentials (VEP) elicited by flanker (0–50 ms window after the onset of stimulus) and target (100–150 ms window after) presentation were clearly observed using both methods. Fig. 6(c) shows eye blinks time-locked to response (click of button), as previously described in [35]. The ERN and VEP shown in Fig. 6 were representative, with comparable results obtained from all subjects, indicating the reproducibility of the ORICA pipeline. In summary, the empirical results demonstrated comparable performance of ORICA to RUNICA in separating informative ICs and resolving single trial and averaged ERPs. Furthermore, ORICA required significantly less computation time than RUNICA.

Adopting a procedure similar to Fig. 2, Fig. 7(a) plots the correlation magnitudes between all ICs learned by ORICA and their best-matched RUNICA counterparts. Firstly, only 10% (26%) of the ORICA components reached a correlation mag-

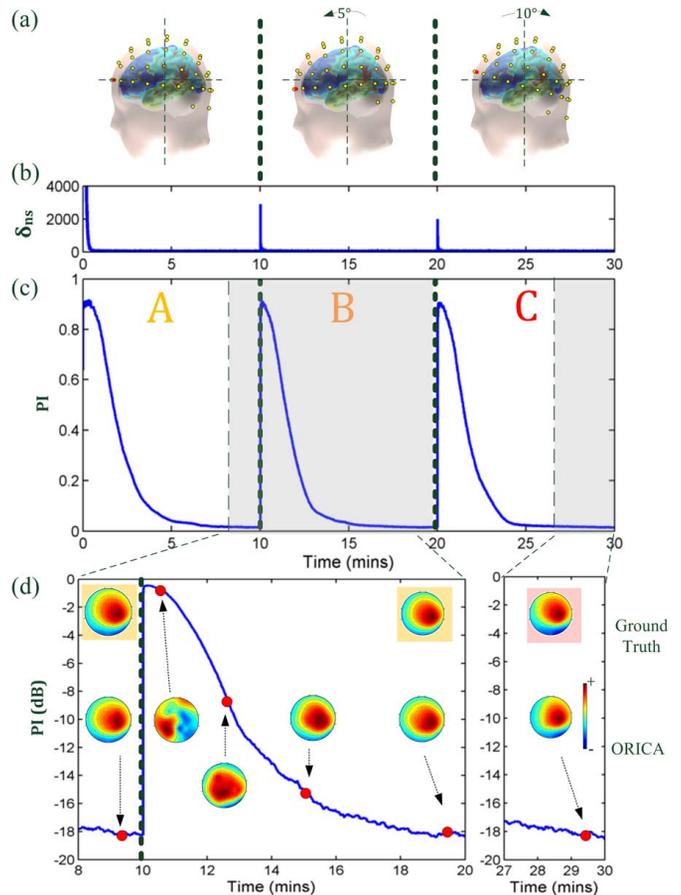


Fig. 5. Application of ORICA to 64-ch simulated nonstationary EEG data consisting of three concatenated 10-min sessions which simulate a 5° forward EEG cap rotation followed by a 10° backward cap rotation. (a) Electrode locations for each session. (b) Nonstationarity index  $\delta_{ns} = \|\mathbf{y}\mathbf{f}^T - \mathbf{I}\|_F$  detects the abrupt change between sessions. (c) The  $PI$  convergence curve shows the adaptation behavior. (d) Zoomed-in plots of log-scaled convergence curve with ground truth component maps (top row) and reconstructed IC at different time points.

nitude of 0.9 (0.8) at the end of the session (average over the last minute). Secondly, 8% (21%) of the components reached a correlation magnitude of 0.9 (0.8) within 3–4 min, only half of the empirical heuristic time suggested by [24].

Among those ICs with the highest correlation magnitudes (larger than 0.95), we found a number of ICs with stereotypical and plausible component maps. Fig. 7(b) plots the convergence profile (evolution of correlation magnitudes) for three such informative ICs: prefrontal (IC 1, accounting for eye-blink), occipital (IC 12, accounting for VEP), and fronto-central (IC 7, accounting for ERN) components. These ICs converged to their RUNICA counterparts within 3–4 min. The correlation magnitude curves of both the occipital and fronto-central components

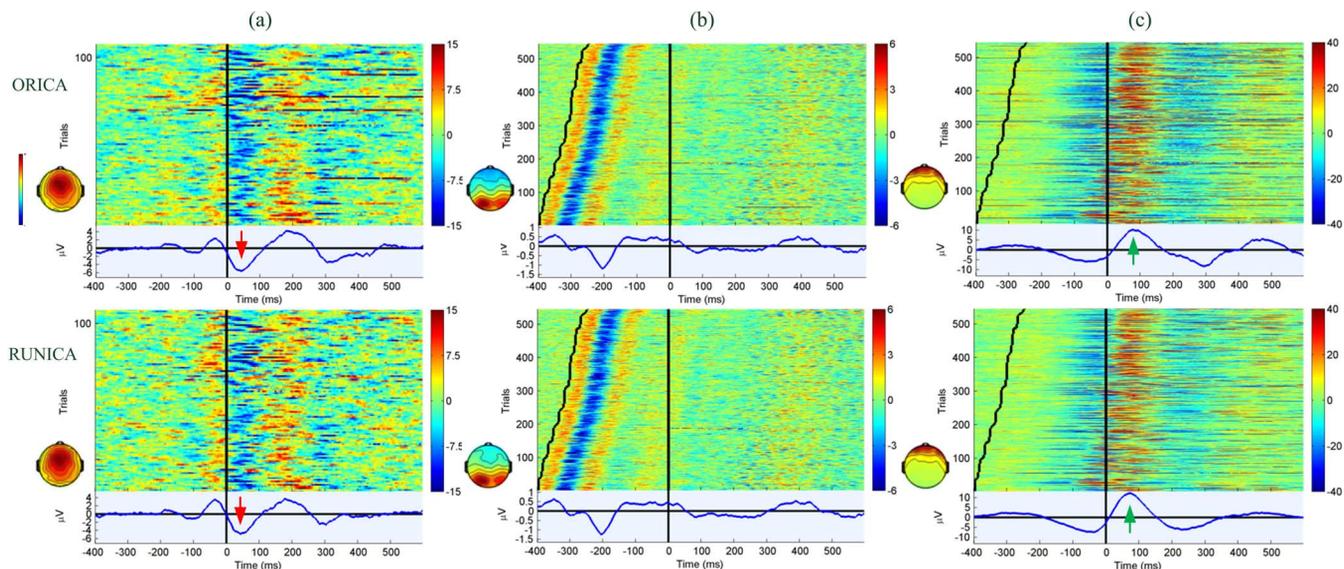


Fig. 6. Color-coded event-related potential (ERP) images (trials by time) of (a) fronto-central, (b) occipital, and (c) prefrontal components reconstructed by ORICA (top row) and RUNICA (bottom row) on real 61-ch EEG data from the flanker task, time-locked to the response at time 0 (vertical straight line). Averaged ERP traces are shown in bottom panel. Only error trials are included in (a) such that error-related negativity (ERN) can be observed as red arrows indicate. In (b) and (c), all trials are sorted based on reaction time, i.e., onset of flanker stimulus (sawtooth line) to response. A visually evoked potential (VEP) is clearly observed in (b). Green arrows in (c) indicate to eye blinks.

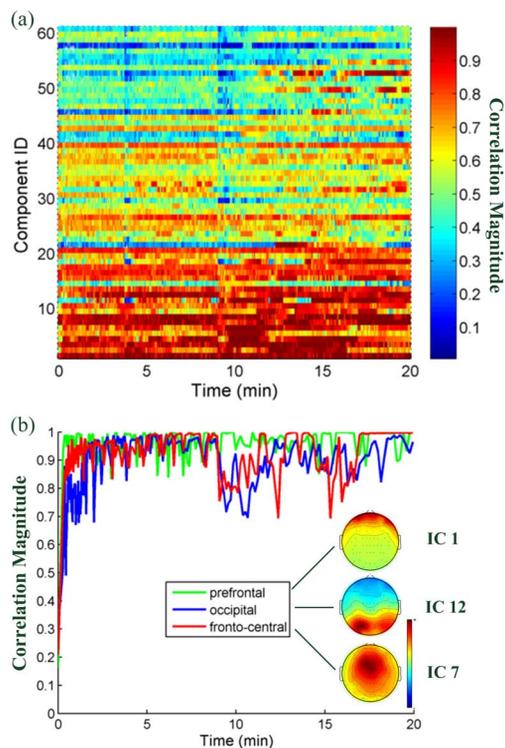


Fig. 7. (a) Evolution of component-wise correlation magnitudes between ORICA- and RUNICA-decomposed ICs on real 61-ch EEG data from the flanker Task. (b) Evolution of correlation magnitudes and spatial filters of three rapidly converged ICs: prefrontal (eye-blinks), occipital (VEP), and fronto-central (ERN) components.

fluctuated across time and eventually reached a steady-state. In contrast, the correlation magnitude time series of the prefrontal component was relatively stable across the whole session.

Fig. 2(a) and Fig. 7(a) exhibit significant performance differences in decomposing simulated versus real EEG data, which can be attributed to the differences in the quality of the gold standard. Those ICs producing poor correlation corresponded to nondipolar sources in the gold standard, i.e., sources with high residual variance in dipole fitting, such as mixtures of sources or noise [7]. This phenomenon is commonly observed and reported when applying ICA methods to real EEG data that are inevitably noisy and likely nonstationary [1], [6].

## V. DISCUSSION

### A. Fast Convergence Speed

The ORICA pipeline was capable of accurately decomposing 64-channel simulated EEG data within the required heuristic convergence time ( $kN^2$ ) [24]. The fast convergence could be attributed to three important factors: 1) combining online RLS pre-whitening and ORICA, 2) choosing an optimal forgetting factor profile and parameters, and 3) fixing the numbers of modeled sub- and super-Gaussian sources.

Simulation results showed that faster convergence of the whitening matrix facilitated ICA convergence. This was consistent with the findings in previous studies [1], [18], which suggested pre-whitening could significantly improve the ICA convergence by reducing the dimensionality of the parameter space. For online whitening, a local block-average approach could provide a more robust estimate than a stochastic (single sample) update approach. Besides RLS whitening, online LMS whitening method [12] can also be considered, which has lower computational complexity, but slower convergence [18].

The forgetting factor, especially its decay rate  $\gamma$ , had a significant effect on ORICA's convergence. For highly nonstationary data, a large  $\gamma$  and thus a shorter effective window size were

preferred. Factors including data dimension (number of channels) and underlying stationarity of the data affected the choice of optimal parameters. Alternative approaches for adapting the forgetting factor were suggested in [17] and [36].

Fixing the numbers of sub- and super-Gaussian sources increased both the stability and the speed of convergence, especially for high-density data. The experiments with synthetic data showed that the choice was not critical, and a model mismatch in these numbers could be tolerated. This study further examined the kurtosis of the real 61-ch EEG data and found that all sources were super-Gaussian distributed, which was consistent with previous studies [2], [26] that EEG signals from most brain activities and nonbrain artifacts were primarily super-Gaussian. For greater accuracy under general conditions, online kurtosis estimation as described in [19] and [37] can be incorporated into the ORICA pipeline. This may lead to better steady-state performance but potentially decreased stability of convergence.

Several approaches not yet implemented in this study can further improve the convergence speed of ORICA. For example, dimensionality reduction methods such as PCA, or selecting a subset of channels prior to ORICA decomposition, can reduce the empirical convergence time. Another promising approach is to pre-process the data using artifact reduction methods such as Artifact Subspace Reconstruction [11] to mitigate sensitivity to transient artifacts in noisy high dimensional EEG data.

### B. Real-Time Processing

ORICA was implemented as a BCILAB function with block-update vectorization, and could easily perform real-time processing with a user-defined block-size for 64-channel EEG on a standard laptop. The block update rule, while approximate, incurred negligible loss in accuracy up to  $L_{ICA} = 64$  for 64-channel simulated data. Even without the block-update ( $L_{white} = L_{ICA} = 1$ ), the ORICA pipeline still ran in real-time. The block update may be most valuable when computational resources are constrained; for instance, when applying multiple data processing operations in serial or operating on a low-power mobile device.

### C. Application to Real EEG Data

Empirical results on the 61-channel EEG data collected in the flanker task experiment demonstrated that ORICA could decompose brain sources and artifact ICs that resembled results from standard RUNICA. We observed that the most informative ICs, such as VEP and ERN brain sources and eye-blink artifacts, had the highest correlation magnitudes among all ICs and converged much quicker than the heuristic convergence time—a fortunate circumstance for real-time applications in mobile EEG BCI. We speculate that those components exhibit robust and frequently occurring statistical patterns which facilitate ICA separation. These observed phenomena support applications of ORICA for rapid decomposition of high-density data as much less time is required to decompose brain and artifact components.

The ORICA pipeline also revealed nonstationarity in the experimental data, captured by the dynamics of component maps and the nonstationarity index. One challenge for ORICA and RUNICA is the order switching of ICs, especially for

nonstationary real EEG data. This hinders the identification of informative ICs over time, i.e., tracking the same set of components regardless of the weight matrix permutations. One solution to the problem is to sort the current weight matrix  $\mathbf{W}(t)$  based on the correlation matching with the previous weight matrix  $\mathbf{W}(t-1)$  using the Hungarian method described in Section III-D to keep track of the same components. This is useful for online identifying and separating artifact components from noisy EEG recordings.

### D. Nonstationarity Detection and Adaptation

This study proposed using the Frobenius norm of ICA error matrix,  $\|\mathbf{y}\mathbf{f}(\mathbf{y})^T - \mathbf{I}\|_F$ , for the ORICA pipeline as an index to detect nonstationarity events in the data, identified as abrupt changes in the mixing matrix. This nonstationarity index measurement captured the fitness of the ICA model to the current data, i.e., the degree to which nonlinear decorrelation was achieved. Alternative forms for the nonstationarity index can be used depending on applications, for instance mutual information reduction (MIR) in windowed data provides a measure of statistical independence between sources [7] and thus can capture the changes in data statistics as MIR varies.

This study also presented a method of nonstationary adaptation by increasing the forgetting factor when the nonstationarity index exceeds a hard threshold, e.g., 1%–10% of the initial value when ICA had not converged. However, this method required prior knowledge of the hard threshold and did not address continuous variation in degree of nonstationarity. For selection of the threshold, it is possible to design an adaptive threshold that depends on the online estimated mean and standard deviation of  $\delta_{n,s}$ . For adaption of the forgetting factor, one possible solution is to adopt the strategy similar to the adaptive learning rate proposed by Murata *et al.* [38] for a gradient-based algorithm in an online environment. An adaptive forgetting factor for ORICA, as an RLS-like recursive online algorithm, is crucial for its ability to track nonstationarity, calling for further investigation.

### E. Applications and Future Directions

The proposed online ICA method for real-time processing of high-density EEG opens up new opportunities for the following potential applications: 1) ICA-based real-time artifact removal (especially for sporadic muscle activities), 2) ICA-based brain activity monitoring (e.g., epilepsy, etc.) for clinical practice, and 3) adaptable ICA-based features for brain state (e.g., cognitive functions, fatigue level, etc.) classification in real-time brain–computer interfaces.

A significant next step is to leverage ORICA for real-time source localization, for instance using anatomically constrained low resolution electrical tomographic analysis (LORETA) [39]. A recent study [27] attempted to combine online ICA and source localization, yet further validation of sources' reliability were needed. Knowledge of the source locations in the brain can be used to assess the reliability of the sources (e.g., validate consistency of the source locations over time and with anatomical expectations), to provide biological interpretation of the decomposed sources, and to integrate with other real-time source-level methods such as connectivity analysis in SIFT [11], [30].

## VI. CONCLUSION

This study proposed and demonstrated an efficient computational pipeline for real-time, adaptive blind source separation of EEG data using Online Recursive ICA. The efficacy of the proposed pipeline was demonstrated on three datasets: a simulated 64-channel stationary dataset, a simulated 64-channel nonstationary dataset, and a real 61-channel EEG dataset collected under an Eriksen flanker task. Through application of ORICA to simulated stationary data we 1) systematically evaluated the effects of key parameters on convergence; 2) characterized the convergence speed, steady state performance, and computational load of the algorithm; and 3) quantitatively compared the proposed ORICA method with a standard offline Infomax ICA algorithm. Our analysis of a simulated nonstationary 64-channel EEG dataset demonstrated ORICA's ability to adaptively track changes in the mixing matrix due to electrode displacements.

Applied to 61-channel experimental EEG data, we demonstrated ORICA's ability to decompose brain and artifact subspaces online, with comparable performance to offline Infomax ICA. Furthermore, we found that subspaces of biologically plausible ICs (e.g., eye, occipital, and frontal midline sources) could be reliably learned in much less time than required by the  $kN^2$  empirical heuristic for ICA convergence. To serve the EEG and BCI communities, the proposed pipeline has been implemented as BCILAB [25] and EEGLAB compatible functions, it has also been integrated into an open-source Real-time EEG Source-mapping Toolbox (REST) [27]. Future work will focus on further validation of this promising method as well as application to artifact rejection, clinical monitoring, and brain-computer interfaces [11].

## ACKNOWLEDGMENT

The authors would like to thank C. Kothe for help with BCILAB integration.

## REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2004, vol. 46.
- [2] T.-P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.
- [3] S. Makeig *et al.*, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.
- [4] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, May 2000.
- [5] T. Mullen, G. Worrell, and S. Makeig, "Multivariate principal oscillation pattern analysis of ICA sources during seizure," in *Proc. IEEE EMBC*, 2012, pp. 2921–2924.
- [6] Y. Wang and T.-P. Jung, "Improving brain-computer interfaces using independent component analysis," in *Towards Practical Brain-Computer Interfaces*. New York: Springer, 2013, pp. 67–83.
- [7] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PLoS One*, vol. 7, no. 2, p. e30135, 2012.
- [8] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [9] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4, pp. 411–430, 2000.
- [10] J. A. Palmer, K. Kreutz-Delgado, B. D. Rao, and S. Makeig, "Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities," in *Independent Component Analysis and Signal Separation*. New York: Springer, 2007, pp. 97–104.
- [11] T. Mullen *et al.*, "Real-time modeling and 3d visualization of source dynamics and connectivity using wearable EEG," in *Proc. IEEE EMBC*, 2013, pp. 2184–2187.
- [12] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [13] S.-I. Amari *et al.*, "A new learning algorithm for blind signal separation," *Adv. Neural Inf. Process. Syst.*, pp. 757–763, 1996.
- [14] X. Giannakopoulos, J. Karhunen, and E. Oja, "An experimental comparison of neural algorithms for independent component analysis and blind separation," *Int. J. Neural Syst.*, vol. 9, no. 2, pp. 99–114, 1999.
- [15] X.-L. Zhu and X.-D. Zhang, "Adaptive RLS algorithm for blind source separation using a natural gradient," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 432–435, Dec. 2002.
- [16] J. Karhunen and P. Pajunen, "Blind source separation and tracking using nonlinear PCA criterion: A least-squares approach," in *Proc. IEEE Int. Conf. Neural Netw.*, 1997, vol. 4, pp. 2147–2152.
- [17] S. Cruces-Alvarez, A. Cichocki, and L. Castedo-Ribas, "An iterative inversion approach to blind source separation," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1423–1437, 2000.
- [18] X. Zhu, X. Zhang, and J. Ye, "Natural gradient-based recursive least-squares algorithm for adaptive blind source separation," *Sci. China Ser. F, Inf. Sci.*, vol. 47, no. 1, pp. 55–65, 2004.
- [19] M. T. Akhtar, T.-P. Jung, S. Makeig, and G. Cauwenberghs, "Recursive independent component analysis for online blind source separation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2012, pp. 2813–2816.
- [20] S.-H. Hsu, T. Mullen, T.-P. Jung, and G. Cauwenberghs, "Online recursive independent component analysis for real-time source separation of high-density EEG," in *Proc. IEEE EMBC*, 2014, pp. 3845–3848.
- [21] H. H. Yang and S.-I. Amari, "Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information," *Neural Comput.*, vol. 9, no. 7, pp. 1457–1482, 1997.
- [22] S.-H. Hsu, T. Mullen, T.-P. Jung, and G. Cauwenberghs, "Validating online recursive independent component analysis on EEG data," in *Proc. IEEE EMBS Conf. Neural Eng.*, 2015, pp. 918–921.
- [23] G. McLoughlin *et al.*, "Performance monitoring is altered in adult ADHD: A familial event-related potential investigation," *Neuropsychologia*, vol. 47, no. 14, pp. 3134–3142, 2009.
- [24] J. Onton and S. Makeig, "Information-based modeling of event-related brain dynamics," *Progr. Brain Res.*, vol. 159, pp. 99–120, 2006.
- [25] C. A. Kothe and S. Makeig, "BCILAB: A platform for brain-computer interface development," *J. Neural Eng.*, vol. 10, no. 5, p. 056014, 2013.
- [26] A. Delorme *et al.*, "EEGLAB, SIFT, NIFT, BCILAB, and ERICA: New tools for advanced EEG processing," *Comput. Intell. Neurosci.*, vol. 2011, p. 10, 2011.
- [27] L. Pion-Tonachini, S.-H. Hsu, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time EEG source-mapping toolbox (REST): Online ICA and source localization," in *Proc. IEEE EMBC*, 2015, pp. 4114–4117.
- [28] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subGaussian and superGaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, 1999.
- [29] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
- [30] T. R. Mullen, *The Dynamic Brain: Modeling Neural Dynamics and Interactions From Human Electrophysiological Recordings*. San Diego: Univ. California, 2014.
- [31] S. Haufe, R. Tomioka, G. Nolte, K.-R. Müller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for EEG/MEG," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1954–1963, Aug. 2010.
- [32] S. Ha *et al.*, "Integrated circuits and electrode interfaces for noninvasive physiological monitoring," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1522–1537, May 2014.
- [33] S. C. Douglas, "Blind signal separation and blind deconvolution," in *Handbook of Neural Network Signal Processing*. Boca Raton: CRC, 2001, vol. 6, pp. 1–28.
- [34] G. Carpaneto and P. Toth, "Algorithm 548: Solution of the assignment problem [h]," *ACM Trans. Math. Software*, vol. 6, no. 1, pp. 104–111, 1980.

- [35] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biol. Psychol.*, vol. 51, no. 2, pp. 87–107, 2000.
- [36] C. Paleologu, J. Benesty, and S. Ciochina, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.
- [37] A. Cichocki and R. Thawonmas, "On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics," *Neural Process. Lett.*, vol. 12, no. 1, pp. 91–98, 2000.
- [38] N. Murata, M. Kawanabe, A. Ziehe, K.-R. Müller, and S.-I. Amari, "On-line learning in changing environments with applications in supervised and unsupervised learning," *Neural Netw.*, vol. 15, no. 4, pp. 743–760, 2002.
- [39] N. J. Trujillo-Barreto, E. Aubert-Vázquez, and P. A. Valdés-Sosa, "Bayesian model averaging in eeg/meg imaging," *NeuroImage*, vol. 21, no. 4, pp. 1300–1319, 2004.



**Sheng-Hsiou Hsu** (S'14) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 2011. He was a sole recipient of the scholarship of government sponsorship for overseas study in Biomedical Engineering, in 2012. He is currently a Ph.D. degree student in bioengineering at University of California at San Diego, La Jolla, CA, USA, working with the Swartz Center for Computational Neuroscience and Institute for Neural Computation.

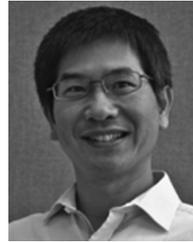
His major research interests include brain-computer interfaces, machine learning, and signal processing with respect to functional imaging of brain dynamics.



**Tim R. Mullen** (S'13–M'14) received dual B.A. degrees in computer science (high honors) and cognitive neuroscience (highest honors) from the University of California at Berkeley, Berkeley, CA, USA, in 2008. He received the M.S. and Ph.D. degrees in cognitive sciences from University of California at San Diego, La Jolla, CA, USA, while at the Swartz Center for Computational Neuroscience, in 2011 and 2014, respectively.

His research has focused on modeling neural dynamics and interactions from human electrophysiological recordings, with applications to clinical and cognitive neuroscience and neural interfaces. He is co-founder and CEO at San Diego neurotechnology company Syntrogi Inc. (aka Qusp) and Research Director of its R&D division Syntrogi Labs.

Dr. Mullen's awards included Glushko, Swartz, and San Diego fellowships; best paper awards; and the 2014–15 University of California at San Diego Chancellor's Dissertation Medal.



**Tzyy-Ping Jung** (F'15) received the B.S. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, OH, USA, in 1989 and 1993, respectively.

He is currently a Research Scientist and the co-Director of the Center for Advanced Neurological Engineering, Institute of Engineering in Medicine, University of California at San Diego (UCSD), La Jolla, CA, USA. He is also an Associate Director of the Swartz Center for Computational Neuroscience, Institute for Neural Computation, and an Adjunct Professor of Bioengineering at UCSD. In addition, he is an Adjunct Professor of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. His research interests are in the areas of biomedical signal processing, cognitive neuroscience, machine learning, time-frequency analysis of human EEG, functional neuroimaging, and brain-computer interfaces and interactions.

Dr. Jung is currently an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.



**Gert Cauwenberghs** (S'89–M'94–SM'04–F'11) received the Ph.D. degree in electrical engineering from California Institute of Technology, Pasadena, CA, USA, in 1994.

He is Professor of Bioengineering and Co-Director of the Institute for Neural Computation at University of California at San Diego, La Jolla, CA, USA. He was previously Professor of Electrical and Computer Engineering at Johns Hopkins University, Baltimore, MD, USA, and Visiting Professor of Brain and Cognitive Science at Massachusetts Institute of Technology, Cambridge, MA, USA. He co-founded Cognionics Inc., and chairs its Scientific Advisory Board. His research focuses on micropower biomedical instrumentation, neuron-silicon and brain-machine interfaces, neuromorphic engineering, and adaptive intelligent systems.

Dr. Cauwenberghs received the NSF Career Award in 1997, ONR Young Investigator Award in 1999, and Presidential Early Career Award for Scientists and Engineers in 2000. He serves IEEE in a variety of roles including as General Chair of the IEEE Biomedical Circuits and Systems Conference (BioCAS 2011, San Diego), as Program Chair of the IEEE Engineering in Medicine and Biology Conference (EMBC 2012, San Diego), and as Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.