Journal of Neuroscience Methods 256 (2015) 41–55

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational neuroscience

Informed decomposition of electroencephalographic data

S.M. Gordon[a,*], V. Lawhern[b], A.D. Passaro[a], K. McDowell[b]

[a] DCS Corporation, Alexandria, VA 22310, USA
[b] Human Research and Engineering Directorate, US Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA

HIGHLIGHTS

- BSS techniques have become the standard for decomposing EEG data but are not optimal when additional information is known about the problem.
- We propose *Informed Multidimensional ICA* (*IMICA*) that builds on ISA and ISS techniques to better model distinct subspaces within EEG data.
- We show that IMICA outperforms other common methods such as Infomax ICA, FastICA, PCA, JADE, and SOBI.
- The results show that IMICA better isolates and removes subspaces within the EEG data.

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form 15 August 2015
Accepted 18 August 2015
Available online 22 August 2015

Keywords:
Independent components analysis
EEG
Independent subspace analysis
Informed source separation

ABSTRACT

Background: Blind source separation techniques have become the *de facto* standard for decomposing electroencephalographic (EEG) data. These methods are poorly suited for incorporating prior information into the decomposition process. While alternative techniques to this problem, such as the use of constrained optimization techniques, have been proposed, these alternative techniques tend to only minimally satisfy the prior constraints. In addition, the experimenter must preset a number of parameters describing both this minimal limit as well as the size of the target subspaces.
New method: We propose an informed decomposition approach that builds upon the constrained optimization approaches for independent components analysis to better model and separate distinct subspaces within EEG data. We use a likelihood function to adaptively determine the optimal model size for each target subspace.
Results: Using our method we are able to produce ordered independent subspaces that exhibit less residual mixing than those obtained with other methods. The results show an improvement in modeling specific features of the EEG space, while also showing a simultaneous reduction in the number of components needed for each model.
Comparison with existing method(s): We first compare our approach to common methods in the field of EEG decomposition, such as Infomax, FastICA, PCA, JADE, and SOBI for the task of modeling and removing both EOG and EMG artifacts. We then demonstrate the utility of our approach for the more complex problem of modeling neural activity.
Conclusions: By working in a one-size-fits-all fashion current EEG decomposition methods do not adapt to the specifics of each data set and are not well designed to incorporate additional information about the decomposition problem. However, by adding specific information about the problem to the decomposition task, we improve the identification and separation of distinct subspaces within the original data and show better preservation of the remaining data.

1. Introduction

Electroencephalographic (EEG) data is currently the most common tool for noninvasively recording the electrical signals of the brain. EEG data can be acquired with high temporal resolution and in a wide range of environments, but there are also disadvantages to using EEG such as poor spatial resolution and low signal-to-noise ratio (SNR). These disadvantages primarily result from the fact that the data recorded by each EEG sensor represents a mixture of multiple sources, which at most EEG frequencies of interest can be assumed to be linear due to the instantaneous nature of volume conduction (Nunez and Srinivasan, 2006). These sources typically include (1) electrical signals arising from distinct regions within the brain in response to task

* Corresponding author. Tel.: +1 571 227 6167.
E-mail address: sgordon@dcscorp.com (S.M. Gordon).

http://dx.doi.org/10.1016/j.jneumeth.2015.08.019

and/or experimental conditions (i.e. the neural activity of interest), (2) background neural activity occurring throughout the brain, (3) simultaneously active non-neural electrophysiological components, and (4) activity from nearby electrical equipment. Often the signal power from the neural activity of interest is greatly overwhelmed by the signal power from these other components.

In order to isolate specific types of information within EEG recordings researchers have turned to decompositions approaches such as principal component analysis (PCA) and independent components analysis (ICA) (Jung et al., 1998a; Jung et al., 2000; Cardoso and Souloumiac, 1993). Both PCA and ICA have been shown to be applicable for a variety of EEG decomposition problems (Niazy et al., 2005; Jung et al., 1998b; Crespo-Garcia et al., 2008; Shao et al., 2009; Safieddine et al., 2012; Vos et al., 2010; Gwin et al., 2010; Srivastava et al., 2005). In particular, ICA methods have received a lot of attention because these methods create maximally, statistically independent components versus the simple decorrelation that occurs with PCA (Comon, 1994; Haykin, 2009; Hyvärinen and Oja, 2000). Yet, despite the success and popularity of ICA methods there are still fundamental limitations with these types of approaches (Haykin, 2009; Vincent et al., 2014). One specific limitation, and the focus of the current paper, is that most ICA techniques are inherently blind, data driven approaches and this can lead to suboptimal or misleading results.

We show that for EEG decomposition there are ways to modify the original ICA approach in order to leverage known features of the underlying source space and better identify target subspaces within the EEG data. With our method, which we refer to as Informed Multidimensional Independent Components Analysis (IMICA), we show that we are able to model both single and multidimensional subspaces. We validate our method on the common problem of EEG artifact removal, as this is where current ICA techniques have had the most success. In our validation tests we show that IMICA extracts target subspaces with a higher SNR than previous methods and, in some cases, performs this extraction using fewer dimensions. We also, briefly, demonstrate the application of our method to the much more difficult problem of modeling specific neural dynamics—namely the alpha and beta spectral changes that have been observed to accompany motor activity (Pfurtscheller and Aranibar, 1977; Pfurtscheller and Lopes da Silva, 1999; Liao et al., 2014).

We begin by providing background on the different decomposition approaches that are typically used to decompose EEG data. In Section 3 we provide further detail about ICA and constrained ICA before introducing our IMICA algorithm. We then describe methods by which one may construct the reference signals that are needed by the IMICA algorithm. We also describe methods to determine the optimal subspace size (i.e. how many dimensions should be used to model the activity of interest), if that size is not known in advance. In Section 4 we use our IMICA approach to model and remove artifacts resulting from electrooculographic (EOG) and electromyographic (EMG) data. We compare the performance of IMICA to other common decomposition methods and show that the subspaces created by IMICA achieve better separation with less crosstalk. For the case of EMG artifacts, we also show that IMICA achieves the improved separation while using fewer components to model the artifact. Next, we provide an example of how IMICA may be used to model specific neural processes, such as the alpha and beta spectral changes commonly associated with left and right finger movements. Finally, in Section 5 we provide an overview of the performance of IMICA as well as outline future directions. We also discuss under what conditions the IMICA methods should be expected to outperform other approaches as well as the limitations that may be encountered.

2. Background

The decomposition methods most commonly applied to EEG data are principal components analysis (PCA) and independent components analysis (ICA). While PCA may be preferred in some cases due to its low computational complexity, ICA has been shown to be the most effective in the broadest range of circumstances ranging from the relatively straightforward problems of artifact removal (Jung et al., 1998a; Niazy et al., 2005; Crespo-Garcia et al., 2008; Gwin et al., 2010; Srivastava et al., 2005) to BCI applications (Kachenoura et al., 2008; Wang and Jung, 2013) as well as the more complex aspects of source modeling and localization (Onton et al., 2006; Eichele et al., 2009; Kobayashi et al., 2001; Groppe et al., 2009).

2.1. Independent components analysis

One of the first ICA algorithms applied to EEG decomposition was the Infomax algorithm developed by (Bell and Sejnowski, 1995). The Infomax algorithm is a gradient algorithm that works by maximizing the entropy of the output variables. This, in effect, reduces the redundancy in the output, or component, vectors compared to the input and ensures maximal, statistical independence. This algorithm, along with its variant the extended Infomax algorithm (Lee et al., 1999a,b), has arguably become one of the more popular methods for decomposition EEG data and has been used for a variety of EEG analysis problems (Jung et al., 1998a; Wang and Jung, 2013; Kachenoura et al., 2008; Groppe et al., 2009).

Another ICA algorithm widely used for EEG analysis is the FastICA algorithm developed by (Hyvärinen, 1999), which works by maximizing the negentropy – a measure of non-Gaussianity – of the projected components. The FastICA algorithm works sequentially and, thus, does not require the costly matrix inversions associated with the Infomax approach. As subsequent components are created they are orthonormalized (using, for example, the Gram-Schmidt method, Haykin, 2009) to ensure that the final decomposition matrix is unique and full rank. There are two main benefits of the FastICA algorithm: (1) an approximation for the computationally demanding negentropy calculation is used which both reduces the computational complexity and speeds up the run time of the algorithm, and (2) components can be learned one at a time, which means the algorithm can be used to create partial decompositions.

Both the Infomax and FastICA algorithms assume each component is independent from the other components in the strong statistical sense, i.e. these methods are designed to account for all higher-order statistics in the data (Hyvärinen and Oja, 2000). There has also been work on decomposition methods that are limited to lower-order statistics. Of these methods, PCA is perhaps the most widely used and it relies purely on second-order statistics. Other popular methods for EEG analysis include SOBI (Belouchrani et al., 1997), which performs joint diagonalization using cross-correlations computed with different time delays and JADE (Cardoso and Souloumiac, 1993), which uses joint diagonalization but focuses on both second- and fourth-cumulants. JADE is particularly unique because it is a block diagonal approach that looks for distinct subspaces within the data that exhibit high intracorrelation and low intercorrelation.

2.2. Independent subspace analysis

ICA methods, such as Infomax and FastICA, assume that the number of underlying independent components is equal to the number of recorded signals. To deal with independent subspaces methods such as Multidimensional Independent Components Analysis (MICA) (Cardoso, 1998) and Independent Subspace Analysis (ISA) (Theis, 2006; Hyvärinen and Oja, 2000) were developed.

MICA was originally proposed as an algorithmic framework that was built around the original ICA problem. With MICA a one-dimensional ICA algorithm, such as Infomax or FastICA, is first applied to the data before a post-processing step is performed that groups output components into internally dependent subspaces. A good example of this is the work by (Ma et al., 2010), which used both Infomax and FastICA variants to decompose fMRI data before hierarchical clustering methods were applied to group the components into hierarchically connected subspaces.

Unlike MICA, which was proposed only as a post-processing approach, ISA methods modify the original ICA update equations in order to create independent subspaces during the decomposition procedure. To do this, ISA methods often require prior knowledge of the data in the form of the size or structure of the subspaces. For example, the work by (Hyvärinen and Köster, 2006) on FastISA, an extension of the FastICA algorithm, assumed preset subspace sizes, while the algorithm developed by (Póczos and Lõrincz, 2005) required preset pruning parameters to auto-divide the subspaces. As will be shown in Section 4, with respect to EEG analysis the correct subspace size is often not known in advance and must be evaluated on a per subject basis.

2.3. Incorporating prior knowledge

One of the more popular methods, as well as the most germane for this article, for incorporating prior information into the ICA procedure is through the use of constrained optimization. There has been much prior work investigating the application of constrained ICA (cICA) to the decomposition of EEG, EOG, and ECG (electrocardiogram) waveforms (Mi, 2014; Lu and Rajapakse, 2003, 2005; James and Gibson, 2003); however, in each of these studies the number of components in the target subspace was known beforehand and a number of parameters were used to limit the influence of the constraints on the decomposition problem. In order to work properly cICA methods require that these parameters be preset by the experimenter (Wang et al., 2006). Often, the experimenter must perform a search over the different parameter settings to find the most optimal solution and this requires a means of evaluating and comparing solutions along the search path. In addition, while the results have been promising the constraint approach has yet to be applied to more complex, multi-dimensional subspaces.

Prior knowledge may also be encoded into the decomposition problem through the use of Bayesian priors that constrain the projection properties and relative statistics for specific sources. Known as informed source separation (ISS), this method is heavily dependent upon the prior model which must be encoded by the researcher (Knuth, 2005). For the case of EEG, prior information related to the location and distribution of sources is not easy to obtain, and the search over possible models can be prohibitively expensive. While prior models can also be encoded by modifying the way in which the statistical measures are computed as in (Machens, 2010; Palmer et al., 2006), where the data related to different processes are used to compute different components, this does not improve the ability to model these processes over the results obtained by simply computing local decompositions.

3. Methods

The IMICA algorithm is an iterative algorithm that builds maximally independent target subspaces from the original data using prior constraints. Before we describe the IMICA algorithm we introduce the basic methods for ICA and cICA. Later, we provide two examples for how to create the necessary reference signals for our approach. We also describe how to determine the size of the target subspaces when that size is not known in advance. Finally, we describe the experimental data that we use in Section 4 to validate and demonstrate our IMICA algorithm.

3.1. Independent components analysis

ICA is a form of blind source separation that assumes the data is a linear mixture of source components (Eq. (1)). Here $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ are the components, which are modeled as functions of time. $\mathbf{A}$ is an $m \times n$ linear mixing matrix and $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_m(t)]^T$ are the recorded output signals. For the remainder of this paper, the dependence on time, t, is dropped for convenience.

$$\mathbf{y} = \mathbf{A} \times \mathbf{x} \tag{1}$$

$$\mathbf{z} = \mathbf{W} \times \mathbf{y} = \mathbf{W} \times \mathbf{A} \times \mathbf{x} \tag{2}$$

The goal of any ICA decomposition is to find a matrix $\mathbf{W}$ that transforms the data into a set of mutually independent signals, $\mathbf{z} = [z_1, z_2, \ldots, z_m]^T$, shown in Eq. (2). This can be done, as in the case of the Infomax and FastICA algorithms, by minimizing mutual information or maximizing the non-Gaussianity, respectively, of the transformed variables $\mathbf{z}$ (Bell and Sejnowski, 1995; Hyvärinen and Oja, 2000). It is difficult to obtain precise mathematical descriptions for the statistical independence of the transformed variables. In practice, approximations are used, which are then encoded into a contrast function, $f(\mathbf{z})$. The weight matrix, $\mathbf{W}$, is then learned by minimizing $f(\mathbf{z})$ through gradient descent.

Due to the gradient nature of most ICA algorithms, it is possible to impart additional information, or constraints, onto the ICA solution through the use of constrained optimization techniques. In the cICA approach the method of Lagrange multipliers (Bertsekas, 1982) is used to solve a constrained version of the ICA problem, given in Eq. (3), where the goal is still to minimize the original contrast function but distinct constraints have been added to the minimization procedure.

$$\begin{aligned} &\text{Minimize} : f(\mathbf{z}) \\ &\text{Subject to} : \boldsymbol{g}(\mathbf{z}) \leq \mathbf{0}, \quad \boldsymbol{h}(\mathbf{z}) = \mathbf{0} \end{aligned} \tag{3}$$

The term $f(\mathbf{z})$ still represents a differentiable ICA contrast function while the terms $\boldsymbol{g}(\mathbf{z}) = [g_1(\mathbf{z}), g_2(\mathbf{z}), \ldots, g_u(\mathbf{z})]^T$ and $\boldsymbol{h}(\mathbf{z}) = [h_1(\mathbf{z}), h_2(\mathbf{z}), \ldots, h_v(\mathbf{z})]^T$ define vectors of u inequality and v equality constraints, respectively. For example, $g(\mathbf{z}) \leq \mathbf{0}$ may encode the constraint that the distance (measured with the L2 norm) between one or more of the output signals and a given set of reference signals, $\mathbf{r}$, be less than some prescribed constants $\boldsymbol{\delta}$, i.e. $g(\mathbf{z}) = ||\mathbf{z}, \mathbf{r}||^2 - \boldsymbol{\delta} \leq 0$. Similarly, $h(\mathbf{z}) = 0$ may encode the desire for the output signals to have a diagonal covariance structure, i.e. $h(\mathbf{z}) = \mathbf{z} \times \mathbf{z}^T - \mathbf{1} = 0$, where $\mathbf{1}$ is a matrix with ones on the diagonal and all other elements set to zero.

In the method of Lagrange multipliers, the solution to (3) is found by minimizing the corresponding augmented Lagrange problem given by Eq. (4), where the functions of $\mathbf{z}$ have been rewritten as functions of $\mathbf{W}$. The terms $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_u]^T$ and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_v]^T$ are the vectors of positive Lagrange multipliers associated with the inequality and equality constraints, respectively. The term, γ, is a penalty term that is used to ensure stable convergence.

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{W}) = f(\boldsymbol{W}) + \frac{1}{2\gamma} \sum_{p=1}^{u} \left((\max[0, \mu_p + \gamma g_p(\boldsymbol{W})])^2 - \mu_p^2 \right) \\ + \boldsymbol{\lambda}^t \boldsymbol{h}(\boldsymbol{W}) + \frac{1}{2} \gamma \left\| \boldsymbol{h}(\boldsymbol{W}) \right\|^2 \end{aligned} \tag{4}$$

This minimization can be performed through a Newton-like update as given in Eq. (5), where the term $\boldsymbol{R}_{\mathbf{yy}}$ is the covariance matrix of the input $\mathbf{y}$, k is the iteration index, η is the learning rate, and $L'_{\boldsymbol{W}}$ is the partial derivative of L with respect to $\mathbf{W}$. The updates

to the Lagrange multipliers are given in Eqs. (6) and (7). For more details describing this technique the interested reader is directed to (Lu and Rajapakse, 2005; Bertsekas, 1982).

$$\boldsymbol{W}\left\{k+1\right\} = \boldsymbol{W}\left\{k\right\} - \eta L'_{\boldsymbol{W}} \boldsymbol{R}_{\boldsymbol{yy}}^{-1} \tag{5}$$

$$\boldsymbol{\mu}\left\{k+1\right\} = \max[0, \boldsymbol{\mu}\left\{k\right\} + \gamma \boldsymbol{g}(\boldsymbol{W}\left\{k\right\})] \tag{6}$$

$$\boldsymbol{\lambda}\left\{k+1\right\} = \boldsymbol{\lambda}\left\{k\right\} + \gamma \boldsymbol{h}(\boldsymbol{W}\left\{k\right\}) \tag{7}$$

What is important to note from these equations is that as long as the constraints, $\boldsymbol{g}(\mathbf{z})$ and $\boldsymbol{h}(\mathbf{z})$, are not satisfied, the Lagrange multipliers will grow without bound (Bertsekas, 1982). In the limit, if the constraints are never satisfied, the weight update will be dominated by the constraint portion of Eq. (4). This will lead to suboptimal performance since the update equation will essentially be tasked with simultaneously satisfying two competing, non-overlapping requirements. In addition, because the individual constraint equations are based on relational operators the solution to (4) will only minimally match the given constraints. We argue that this logic is counterintuitive because it is the constraints that, theoretically, represent what is known about the decomposition problem. Thus, it is the constraints that should be optimized while independence is asserted to the extent allowed by both the data and the constraints.

3.2. Informed multi-dimensional ICA

The key to the IMICA algorithm is that we reformulate the decomposition problem to solve for the constraints first, while iteratively projecting the remaining dimensions into a maximally independent subspace. By reformulating the problem in this way we are able to optimize how well the projected components match the prior information, i.e. constraints, without relying on preset parameters to limit the influence of our constraint equations. We still recover maximally independent subspaces, but whereas with most ICA methods the projected components are maximally independent given the input data, in our solution the final components are maximally independent given both the input data and the prior knowledge encoded in $g(\mathbf{z})$. This is true because at each pass of the algorithm, the target subspace is computed from a reduced version of the original subspace that is the maximally, statistically independent residual remaining after application of the gradient-based ICA update rule. The steps of the general IMICA algorithm are:

Step 1: Initialize the variable $\hat{\mathbf{y}} = \mathbf{y}$ and $\mathbf{W}^0 = empty$.

Step 2: Create the N_j dimensional target subspace, $\mathbf{W}^j = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{\mathrm{Nj}}]^T$, by solving the constrained optimization problem shown in Eq. (7), where $\hat{\boldsymbol{z}} = \boldsymbol{W}^j \times \hat{\boldsymbol{y}}$, $\mathbf{z}' = \mathbf{W}' \times \boldsymbol{y}$, and $\mathbf{W}' = [\boldsymbol{W}^0, \boldsymbol{W}^j]$.

$$\text{Minimize} : \boldsymbol{g}(\hat{\mathbf{z}})$$
$$\text{Subject to} : \boldsymbol{h}(\mathbf{z}') = 0$$

Note: if the size of the jth target subspace, N_j, is not known, then components should be incrementally added to $\mathbf{W}^j$. This creates an array of matrices, $\mathbf{W}'\{q\}$, where q is the iteration index. The resulting, intermediate, decompositions $\mathbf{W}'\{q\}$ should be individually evaluated as discussed later in Section 3 to determine the optimal, and final, subspace size.

Step 3: Compute a new matrix $\mathbf{W}^{\mathrm{ICA}}$ by minimizing the differentiable ICA contrast function $f(\mathbf{z} = \hat{\mathbf{W}} \times \mathbf{y})$, where $\hat{\mathbf{W}} = [\mathbf{W}', \mathbf{W}^{\mathrm{ICA}}]$ and $\mathbf{W}'$ is now considered a constant. In the work that follows we use the Infomax update rule to compute $\boldsymbol{W}$ because we found that it produces slightly better results. FastICA is also well-suited for this process since it learns weight vectors one at a time and can thus be used to build components on top of a pre-existing subspace. FastICA is also computationally simpler.

Step 4: Compute a new value for $\hat{\mathbf{y}}$ by using $\hat{\mathbf{W}}$ to remove all components associated with $\mathbf{W}'$. In other words, remove all components associated with the target subspaces $\mathbf{W}'$ by setting these components to zero and backprojecting the data into the original signal space using $inv\left(\hat{\mathbf{W}}\right)$. Set $\mathbf{W}^0 = \mathbf{W}'$.

Step 5: If there are remaining subspaces to be computed, go to Step 2, otherwise return $\mathbf{W}^{\mathrm{final}} = \hat{\mathbf{W}}$.

As in the previous discussion of cICA, the function $g(\mathbf{z}')$ is a differentiable function that relates the properties of the projected components to the known features of the signal and the function $\boldsymbol{h}(\mathbf{z}')$ is differentiable function designed to enforce orthogonality. We use $h(\mathbf{z}) = \mathbf{z} \times \mathbf{z}^T - \mathbf{1} = 0$.

3.3. Reference construction

Similar to the original cICA work of (Lu and Rajapakse, 2003, 2005; James and Gibson, 2003) we use time-varying reference signals to identify the target subspace. We use the Pearson correlation between the projected components and these reference signals for $\boldsymbol{g}(\mathbf{z}) = \boldsymbol{corr}(\mathbf{z}, \mathbf{r})$. Such a choice is not a requirement; any differentiable contrast function could be used as long as it captures the desired features of the target subspace. In this section we describe two different techniques to create the reference signals. These techniques are relatively simple but we have found that they work quite well in practice. The first technique requires the use of an external, but synchronized, time-series to describe the signals of interest. Fig. 1 shows an example in which there are three synchronized time-series representing horizontal and vertical signals from three sources related to a simple object tracking task. These time series include: (1) position of an object being tracked, (2) output of a video-based eye tracking system, and (3) the recorded EOG. For this task we asked subjects to refrain from moving their heads while they visually tracked a ball as it bounced around the screen. The column on the left shows the horizontal movements while the column on the right shows the vertical movements. The top row shows the position of the ball on the screen. The middle row shows the synchronized eye gaze recorded by an external eye tracker. The bottom row shows the horizontal and vertical EOG recordings. As can be seen from the figure, both the eye tracking and object position information provides a good reference for the EOG behavior.

The second technique for creating reference signals relies on precomputing generalized representations of actual time–frequency changes present in the data. Again, we provide an example based on EEG artifacts since the SNR of most artifact subspaces make them easy to distinguish from the background neural activity. To create the time-frequency reference (Fig. 2 Top-Left) we first time-lock to the onset of a muscle artifact. The original data has been bandpass filtered to [1, 50] Hz. We perform an event-related spectral analysis (Fig. 2 Top-Right) across these epochs. By inspecting the data we determine a frequency band whose activity is related to the signal of interest. In the case of EMG activity, it is known (Anderer et al., 1999; Andreassi, 2007) that EMG activity disproportionately affects the EEG frequency bands starting in the upper alpha and beta frequency bands. For our example this is confirmed by visual inspection of the Top-Right image in Fig. 2 Once a frequency band of interest is identified, we then apply a bandpass filter to isolate the information in that band (Fig. 2 Middle-Left). The data is squared and smoothed using a Gaussian convolution kernel (Fig. 2 Middle-Right). This step makes it easier to identify power changes across different epochs. Next, we average over the different epochs to create a generalized

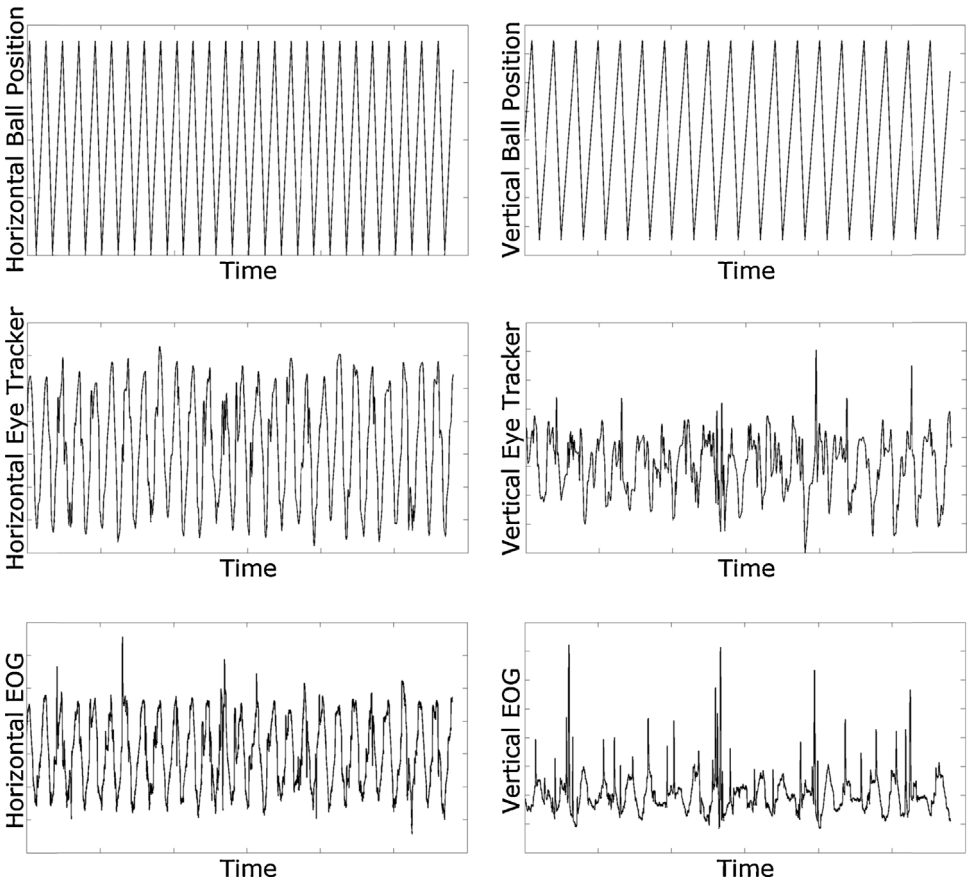

Fig. 1. Reference pairs used to capture EOG artifacts. (Top-Left) Horizontal reference based on the position of a bouncing ball the subjects were instructed to visually track. (Top-Right) Vertical reference based on the position of a bouncing ball the subjects were instructed to visually track. (Middle-Left) Horizontal reference based on eye tracker recordings during the bouncing ball task. (Middle-Right) Vertical reference based on eye tracker recordings during the bouncing ball task. (Bottom-Left) Horizontal EOG during the bouncing ball task. (Bottom-Right) Vertical EOG during both the bouncing ball task.

representation of the band-specific power changes associated with the EMG signal (Fig. 2 Bottom-Left). This overall process is a well-known method, described in Pfurtscheller and Lopes da Silva (1999) for the purposes of assessing neural-specific event-related synchronizations and desynchronizations. Using this method we create an averaged representation of the signal dynamics that are not specific to any single epoch or single channel. To create a continuous representation we replicate the averaged signal for each of the original, labeled epochs (Fig. 2 bottom-right). It should be noted, that there is going to be some variability across individual "instances" of the signal of interest. The goal is to create a reference that captures the general properties of the signal without specifically modeling the activity of any single channel or any single time point.

3.4. Model size estimation

In some cases (e.g. modeling EOG activity) the size of the target subspace may be relatively well-known, but most of the time the optimal size will vary from subject to subject and will not be known in advance. Incorrectly guessing the size of the target subspace can have substantial negative effects on the quality of the solution. If too many components are used then the SNR will be adversely affected by including too much additional information. If too few components are used then the SNR will suffer because the model did not capture a sufficient amount of the signal of interest. When the size of the target subspace is unknown we propose incrementally adding components while evaluating the performance of the model at each step.

The most straightforward and, arguably most popular, way to evaluate EEG decompositions is through visual inspection. However, this can be very time consuming for the researcher and can vary from one analysis to the next. An alternative approach is to evaluate the properties of the decomposed data against a control condition. If the experimenter has access to data in which the target subspace is not active but the remaining subspaces are active then this evaluation can be performed by comparing the effects of selectively removing the subspace of interest. In Section 4 we demonstrate the use of this technique but acknowledge that acquiring the necessary control data can be difficult in practice.

Another, more practical, evaluation technique is to use a model selection approach, such as the techniques commonly used in information theory (Akaike, 1981; Schwarz, 1978), and have been previously used for a variety of different EEG analyses (Gordon et al., 2013; Lawhern et al., 2012). These methods often require the use of a likelihood function and a penalty term. The model selection criterion then balances the explanatory power of the model determined by the likelihood function against the complexity, or size of the model, encoded in the penalty term. For the EEG decomposition problems that we describe in this article, while the size of the individual target subspaces may vary, the total number of parameters estimated, i.e. the size of the decomposition matrix $\mathbf{W}$, is constant. For this reason, we propose to simply use the likelihood curve component, and not the penalty term, for our model selection criterion. To compute the likelihood curve (LC) we assume a Gaussian mixture model and use $\boldsymbol{W}\{q\}$, the decomposition matrix

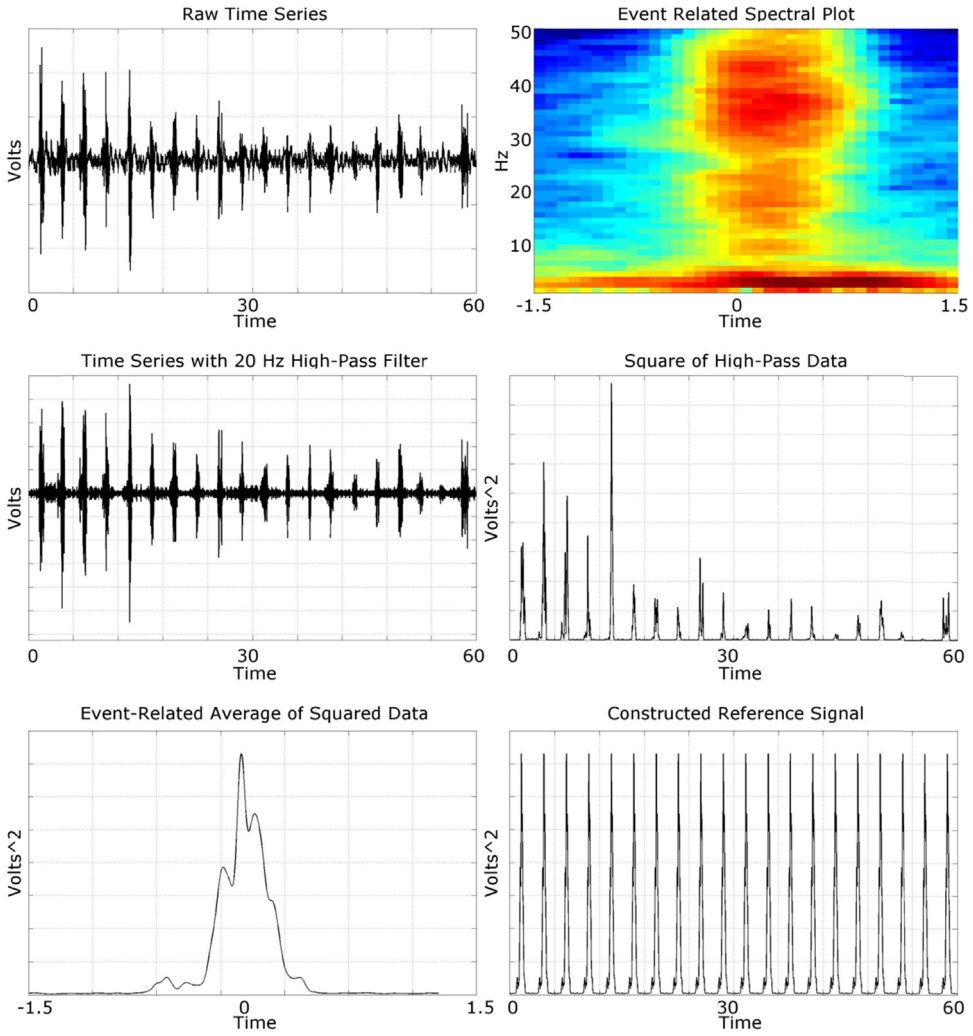

Fig. 2. Creation of reference signal for EMG activity. (Top-Left) Original EEG signal recorded at electrode F3 for multiple EMG artifacts generated by jaw clenching. (Top-Right) Event-related, i.e. averaged, spectral analysis of artifact instances with window taken from −1.5 s to 1.0 s. The *y*-axis is frequency and the *x*-axis is time. (Middle-Left) Original EEG signal after 25 Hz high-pass filter. (Middle-Right) Squared amplitudes of high-pass filtered data after smoothing with Gaussian window of width 100 ms. (Bottom-Left) Averaged power change across all instances of the artifact. (Bottom-Right) Reference signal created by replicating the averaged power change for each artifact instance.

computed at iteration q, to separate the data into target, $\mathbf{y_1}$, and non-target, $\mathbf{y_2}$, subspaces as shown in the following equation:

$$\mathrm{LC}\{q\} = \log\left(P_1\left(\boldsymbol{y}_1\left|\boldsymbol{W}\{q\}\right.\right) + P_2\left(\boldsymbol{y}_2\left|\boldsymbol{W}\{q\}\right.\right)\right) \tag{8}$$

The final model size is determined by referring to the LC. In the results we provide an example of this approach and compare the model size estimates using our control condition to the model size estimates predicted by the LC.

Finally, we would like to point out that it is not typically clear how one should compare different decompositions. There have been several approaches to this problem including the work of (Delorme et al., 2012; Glass et al., 2004; Karvanen, 2005; Cashero and Anderson, 2011), but it is acknowledged that often the metric for comparison tends to be biased towards one type of decomposition or one type of analysis. For example, the work of (Cashero and Anderson, 2011) compares the quality of different decompositions for the purposes of building brain-computer interfaces, which represents only one type of EEG analysis. The work of (Delorme et al., 2012) uses mutual information reduction, which is biased towards Infomax ICA. Similarly, one would expect decorrelation to be biased toward PCA. In the work that follows we have attempted to select a variety of metrics for comparison, but note that some bias will inherently be present.

3.5. Experimental data

One of the most common applications of ICA to EEG data is for the task of modeling and removing artifact signals related to EOG, ECG, and EMG activity (Jung et al., 2000; Parra et al., 2005; Mennes et al., 2010; Milanesi et al., 2008; Matic et al., 2009; De Clercq et al., 2006; Jung et al., 1998a,b; Iriarte et al., 2003). In this paper we demonstrate the utility of IMICA by comparing its performance on the task of artifact removal to other well-accepted methods. We chose artifact removal because it is one of the most common uses of ICA methods for EEG analysis. Artifact removal also allows us to better analyze the separation properties of the different methods in question, since EEG artifacts tend to have much higher amplitude than the neural data. We focus on two tasks: (1) modeling and removing horizontal and vertical eye movements and blinks, and (2) modeling and removing high power EMG artifacts resulting from jaw clenches.

We analyzed data from 10 right-handed, male subjects aged 20–40. The data was obtained with the approval of the Institutional Review Board of the Army Research Laboratory's Human Research and Engineering Directorate. EEG data was collected with a 64 channel BioSemi Active II system. Eye tracking data was collected using a FaceLAB eye tracking system version 4.2 (www.seeingmachines.com). The subjects were asked to perform a battery

of artifact-generating tasks. The tasks included: (1) tracking a ball using smooth pursuit eye movements as the ball slowly bounced around the screen, (2) performing a series of quick saccades in order to follow a changing pattern, (3) blinking the eyes, and (4) clenching the jaws. The subjects had to perform 20 eye blinks and 20 jaw clenches. Each subject tracked the bouncing ball for 2 min and performed the saccade task for approximately 1 min at the rate of one saccade approximately every 2 s. After the artifact battery, the subjects then participated in another study, which took approximately 30 min to complete. In this second study, which was designed to analyze neural and behavioral dynamics of target discrimination, the subjects were required to observe visual stimuli in a moving environment and to discriminate the type of stimuli as either target (an image of a man with a gun) or non-target (an image of a man). After discrimination the subjects responded by pushing a button with the left or right hand to indicate target type.

The second data set that we analyzed was a basic motor tapping study. We analyzed data from 8 right-handed males and 2 right-handed females. The data was obtained with the approval of the Institutional Review Board of the Army Research Laboratory's Human Research and Engineering Directorate. EEG data was collected with a 256 channel BioSemi Active II system. The subjects performed a self-paced finger tapping task, in which they were required to use the left middle, left index, right middle, or right index finger to press a button. The subjects were asked to maintain approximately four to 5 s between movements. Movements were recorded in blocks of 2 min. For each block the subject was instructed which finger to use. There were a total of 32 blocks, or eight blocks for each finger.

4. Results

4.1. Removal of eye movements and blinks

The first validation test that we performed focused on modeling and removing horizontal and vertical eye movements and blinks. This is a common problem for EEG analysis and is the first test of many decomposition methods. For this analysis we preset the model size to one component for horizontal movements and one component for the combination of vertical movements and blinks.

We used three different synchronized time-series to generate the references. First, we used the horizontal and vertical positions of the bouncing ball on the screen (Fig. 1) and then supplemented the vertical reference with a square-wave that was time-locked to the blink task. The use of a square wave to represent blink activity is based on the approach of (Lu and Rajapakse, 2005). Second, we used the horizontal and vertical measurements obtained from the synchronized eye tracker. Again, we supplemented the vertical reference with the time-locked square wave to capture the blink activity. Third, we used the horizontal and vertical EOG channels directly. Since, in each case, the vertical reference included the blink task this meant that the vertical reference was trained using slightly more data than the horizontal reference.

We then compared the performance of IMICA using these reference methods with Infomax ICA, FastICA, JADE, SOBI, and PCA. We selected these methods for comparison because based on our experience they are the most common decomposition methods used for EEG analysis. Each of the non-reference methods were trained using the same training data that was used for IMICA, i.e. the approximately two and a half minutes from the combined tracking and blinking tasks. All of the algorithms were tested against both the horizontal and vertical EOG channels from the separate saccade task, and against a separate set of manually-labeled blink events that occurred within the data, but not within the training period. Finally, we determined which components to remove using one of the following two criteria: (1) for the IMICA decompositions we removed only those components that had been created to model the reference signals, (2) for the non-reference methods we removed the components that had maximal correlation with the horizontal and vertical EOG signals during the training period (Plöchl et al., 2012).

For the initial assessment of performance we computed a number of measures including:

1. Maximum amplitude of time-locked blink events after artifact removal divided by the maximum amplitude of time-locked blink events prior to artifact removal.
2. R^2 of the modeled blink ERP component, i.e. using only the two identified EOG components, to the original blink ERP, i.e. using all components, averaged across all scalp channels.
3. Correlation of the best horizontal component with the horizontal EOG.
4. Correlation of the best vertical component with the vertical EOG.
5. Correlation of the best horizontal component with the synchronized horizontal eye tracking data.
6. Correlation of the best vertical component with the synchronized vertical eye tracking data.

To compute all of the metrics we used a separate set of testing data. For the blink metrics the testing data was derived from at least 16 manually selected blink events per subject. These events were taken from a baseline period in which no other tasks were performed. For the remaining metrics we used the data from the separate saccade task. The results are summarized in Table 1.

Based on these results it is clear to see that IMICA, regardless of the reference signal used, performed as well as, if not better than, the non-reference methods for each of the chosen criteria with the exception of correlation with the VEOG, in which both SOBI and PCA were able to identify strong VEOG components. There are slight performance differences between the different reference methods. This is largely a function of the quality and type of reference signals used. Using a clean EOG provides the most accurate reference signal for capturing the electrophysiological artifacts arising from horizontal and vertical eye movement and blink activity.

To assess whether we biased the results in favor of IMICA by using only a small amount of data and to determine whether methods such as Infomax, FastICA, and JADE would benefit from having more training data we performed a second analysis in which additional training data was incrementally included. We only performed this computation for the non-reference methods and compared the results to those achieved by the reference methods using the original training data. For each step of this analysis the non-reference methods were always given the original training data that had been used to generate the decompositions presented in Table 1. Additional data was then included in increments of 2, 3, 4, 5, 10, and 15 min. These results are presented in Table 2 for the metrics (1, 3, & 4).

Even with the inclusion of an additional 15 min of training data the performance of the non-reference methods did not reach that of IMICA. The last comparison we performed was to look at the variability of the different methods across subjects. We focused only on the first metric: the amount of residual blink activity left after artifact removal. We chose to include only this metric both to conserve space and because the variability observed with the remaining metrics followed this same pattern. Fig. 3 shows the mean and standard deviation for the three IMICA approaches, the non-reference methods using only the original 2.5 min of training data, and the non-reference methods using the additional 15 min of data. As can be seen, the results produced by IMICA tend to have lower means and exhibit less variability than the non-reference

Table 1
Comparison of performance across different methods using horizontal and vertical eye movements and blinks using 2.5 min of data.

Projection method	(1) Residual blink amplitude remaining after removal (%)	(2) R^2 using modeled blink activity	(3) Corr. w/HEOG channel	(4) Corr. w/VEOG channel	(5) Corr. w/horiz. eye tracker	(6) Corr. w/vert. eye tracker
IMICA: EOG	**0.1567**	0.9408	**0.9034**	0.6660	0.6040	**0.2470**
IMICA: EyeTracker + blink	0.1859	**0.9529**	0.8423	0.5228	0.5916	0.1316
IMICA: SmoothPursuit + blink	0.1864	**0.9529**	0.8754	0.5381	**0.6063**	0.1345
Infomax	0.5157	0.5676	0.4466	0.4621	0.2932	0.1483
FastICA	0.6567	0.4388	0.4532	0.4500	0.3066	0.1527
JADE	0.6940	0.4128	0.4484	0.4297	0.2758	0.1498
SOBI	1.3346	−1.0658	0.7966	**0.7046**	0.5461	0.2125
PCA	0.7345	0.2534	0.8539	0.6359	0.5864	0.1736

Table 2
Comparison of performance at removing eye artifacts after adding training data to the original 2.5 min. Results shown are for metrics 1, 3, and 4.

Time (min)	Infomax	FastICA	JADE	SOBI	PCA
2	**0.47**, 0.45, 0.26	0.69, 0.45, 0.28	0.57, 0.45, 0.24	1.12, 0.78, 0.36	0.72, **0.88, 0.60**
3	**0.47**, 0.48, 0.24	0.56, 0.43, 0.20	0.51, 0.45, 0.24	1.03, 0.77, 0.30	0.72, **0.89, 0.62**
4	**0.38**, 0.48, 0.26	0.54, 0.48, 0.22	0.50, 0.48, 0.25	1.23, 0.73, 0.35	0.73, **0.89, 0.59**
5	**0.31**, 0.48, 0.17	0.41, 0.46, 0.20	0.42, 0.49, 0.21	1.23, 0.75, 0.30	0.83, **0.89, 0.61**
10	**0.29**, 0.67, 0.17	0.42, 0.53, 0.24	0.38, 0.57, 0.23	1.05, 0.77, 0.37	0.68, **0.89, 0.58**
15	0.30, 0.77, 0.21	0.41, 0.59, 0.29	**0.22**, 0.62, 0.16	1.09, 0.79, 0.41	0.84, **0.88, 0.58**

methods. Using a two-sample *F*-test we determined that the variability for the IMICA results produced using the EYE tracker and SMOOTH pursuit data was significantly ($p < 0.05$) less than the results produced by any other non-reference method, with the exception of SOBI ALL.

To summarize the results: with respect to EOG removal IMICA consistently outperformed the blind decompositions. The best performing blind methods appear to be Infomax, FastICA, and JADE with Infomax performing better with less training data while JADE performed best once more data was included. Methods such as SOBI and PCA did a very good job at finding components that matched the horizontal and vertical eye movements; however, these methods were consistently unable to isolate this information and thus removing the strongest components either had little effect upon or, at times, even increased the blink amplitude. While these methods may have performed better if more components had been removed, one of the goals of the current analysis was to assess performance with a fixed model-size. Learning the number of components needed to model the artifact is addressed in the next section.

4.2. Removal of muscle activity

The second validation test that we performed involved the much more difficult task of removing high-amplitude muscle activity. For this test subjects were asked to repeatedly clench their jaws. This generated large muscle artifacts across the entire scalp. We used IMICA to model and remove these artifacts. Since it was not known how many components should be used for each subject we incrementally grew the target subspace and analyzed the results to determine the appropriate cutoff point. For each subject we limited the size of the target subspace to 45 distinct components. To determine the final number of components we used both a control condition and the LC.

We created reference signals using the event-related spectral analysis method outlined previously. To create the artifact subspace, we analyzed data around each jaw clench using a window size of 1.75 s, starting half a second before the initiation of muscle activity. We then computed the remaining components using all the data from the calibration period, i.e. the other artifact generating tasks and the rest periods. Since the other decomposition

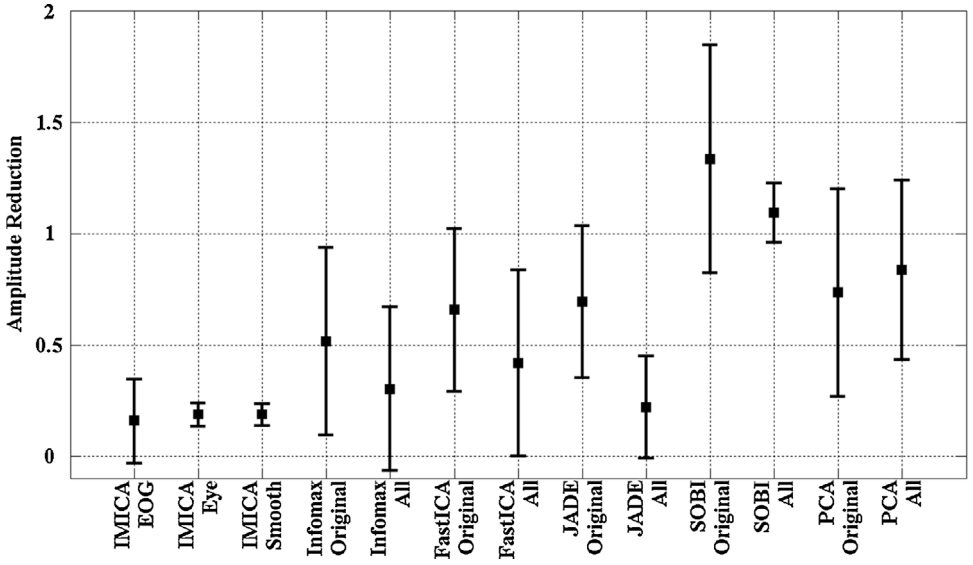

Fig. 3. Means and standard deviations for residual blink amplitude using only the original 2.5 min of training data and the original training data plus 15 min.

Table 3
Average number of components removed to model the jaw clench artifact based on matching the ratio of beta (gamma) power to the control condition. The values in parenthesis in each row indicate the ratio of beta (gamma) power in the artifact condition before and after artifact removal and the ratio of beta (gamma) power in the control condition before and after artifact removal, respectively.

		IMICA	Infomax	FastICA	JADE	SOBI	PCA
Average using artifact data	Beta	19 (0.87, 0.97)	26 (0.88, 0.98)	26 (0.88, 0.99)	33 (0.89, 0.99)	2 (1.00, 1.00)	2 (0.86, 0.89)
	Gamma	20 (0.81, 0.98)	27 (0.83, 0.98)	30 (0.83, 0.98)	35 (0.83, 0.99)	2 (1.00, 1.00)	5 (0.80, 0.85)
Average using all data	Beta	19 (0.87, 0.97)	24 (0.88, 0.89)	38 (0.89, 0.94)	31 (0.90, 0.90)	13 (1.00, 0.99)	4 (0.85, 0.85)
	Gamma	20 (0.81, 0.98)	31 (0.76, 0.81)	42 (0.77, 0.92)	40 (0.79, 0.84)	12 (1.00, 0.99)	11 (0.72, 0.76)

methods do not selectively use the data in this way we performed two comparisons. In the first we compared results obtained using IMICA to results obtained from the non-reference methods trained on only the jaw clench epochs. Second, we compared the results obtained using IMICA to the results obtained from the non-reference methods trained on all of the data for each subject.

To perform the control condition comparison we used the alpha (7–12 Hz), beta (13–30 Hz), and gamma (31–50 Hz) activity from both the artifact and non-artifact subspaces. For our control we used the blink task, as described previously. Both the blink and jaw clench tasks required the subject to watch for a visual cue and generate an artifact on command. The only differences between the two tasks were the type of artifact generated and the specific visual cue. For the blink task the visual cue was the word "blink" and for the jaw clench task the cue was the word "clench". Visual inspection of the data revealed that the blink artifacts were isolated to the lower frequency bands, i.e. delta and theta, while jaw clench artifacts tended to mostly affect the upper frequency bands, i.e. upper alpha, beta, and gamma. The goal for this decomposition was to remove the high frequency artifact activity while preserving the high frequency, non-artifact activity.

We used two simple criteria to determine the appropriate size of the artifact subspace:

1. The ratio of beta power of the cleaned EEG data from the jaw clench condition to the beta power in the control condition.
2. The ratio of gamma power of the cleaned EEG data from the jaw clench condition to the gamma power in the control condition.

We incrementally removed components until these ratios were as close to unity as possible and compared the number of components needed by each method to model the artifact. We also analyzed the effect of removing the jaw clench subspace on the remaining neural data.

Finally, before presenting these results, we need to describe how we determined the jaw clench subspace for the non-reference methods. One benefit of IMICA is that the target subspace is ordered within the decomposition. For the non-reference methods, however, we had to sort the components so that we could remove them based on their contribution to the artifact signal. We did this using a full search over all components in which we ranked the amount of variance in the EEG signal that was explained by each component. We removed the highest ranking component before repeating the search over the remaining components. This is a computationally intensive step that is not often employed in ICA analyses but was used here because empirical tests showed that it far surpassed manual inspection and ranking.

Table 3 shows the average number of components removed across subjects for each decomposition method using the two different selection criteria and the two different sets of training data. The two values in parentheses are (1) the ratio of beta (gamma) activity from the jaw clench condition after and before artifact removal, with values closer to unity indicating that the artifact was not removed and (2) the ratio of beta activity from the control condition after and before removal of the artifact subspace, with values closer to unity indicating the signal power in the control condition was preserved. Inspecting these two values from Table 3 it can be seen that IMICA, Infomax, FastICA, and JADE are equally able to reduce the beta (gamma) power for the jaw clench artifact (first value in parenthesis) while preserving the beta (gamma) power in the control condition (second value in parenthesis), yet IMICA achieves this performance by removing fewer components. Fig. 4 shows an example of removing the artifact subspace for one representative subject. To generate this figure components were removed as determined by the beta criterion, we also only present a subset of the channels due to space constraints. The results for IMICA, Infomax, FastICA, and JADE are very similar, with little-to-no visible jaw clench artifact remaining after subspace removal. For SOBI and PCA, however, there is still clear muscle activity that has not been properly removed.

Inspecting Table 3 and Fig. 4 one can see that SOBI and PCA did not perform well on this task. For the case of SOBI, we could not identify a usable artifact subspace with either our algorithmic approach or subsequent visual inspection that both removed the artifact but did not distort the non-artifact data. For the case of PCA, the artifact was simply not isolated very well and the removal of artifact components caused a general attenuation of the entire signal. As with SOBI, we attempted visual inspection of the components produced by PCA but were not able to achieve significantly better results.

From the bottom two rows of Table 3 one observes that including all of the experiment data did not improve the performance of the non-reference methods. In many cases the number of components increased, which caused more attenuation in the control condition before and after subspace removal. For the remainder of this section we will only focus on the performance of the non-reference methods when trained with just the artifact epochs.

In addition to the results shown in Table 3 we also assessed the extent to which artifact subspace removal impacted an ERP and time-frequency analysis. Since it is difficult to obtain time-locked neural signatures that precisely co-occur with artifacts and yet are not a part of the artifact, we chose to focus on non-artifact data taken from the second study that the subjects completed. As mentioned previously, after the subjects performed the battery of artifact generating tasks, each subject participated in a second study in which the requirement was to visually discriminate between two different target types and then to respond appropriately by pushing a button with either the left or right hand as quickly as possible.

We analyzed each of the artifact removal methods across all subjects. We first focused on the evoked responses from visual target presentation and we only analyzed correct responses. We also limited our analysis to only clean trials, i.e. trials without identifiable muscle or blink artifacts based on visual inspection of the data. It is important to note that by removing the jaw clench components when a neural event was occurring but no jaw clench was occurring we were better able to measure the effects that removing the jaw clench subspace has on the true neural signal. To assess performance we analyzed the correlation of the time-locked ERP after artifact removal with the time-locked ERP prior to artifact removal.

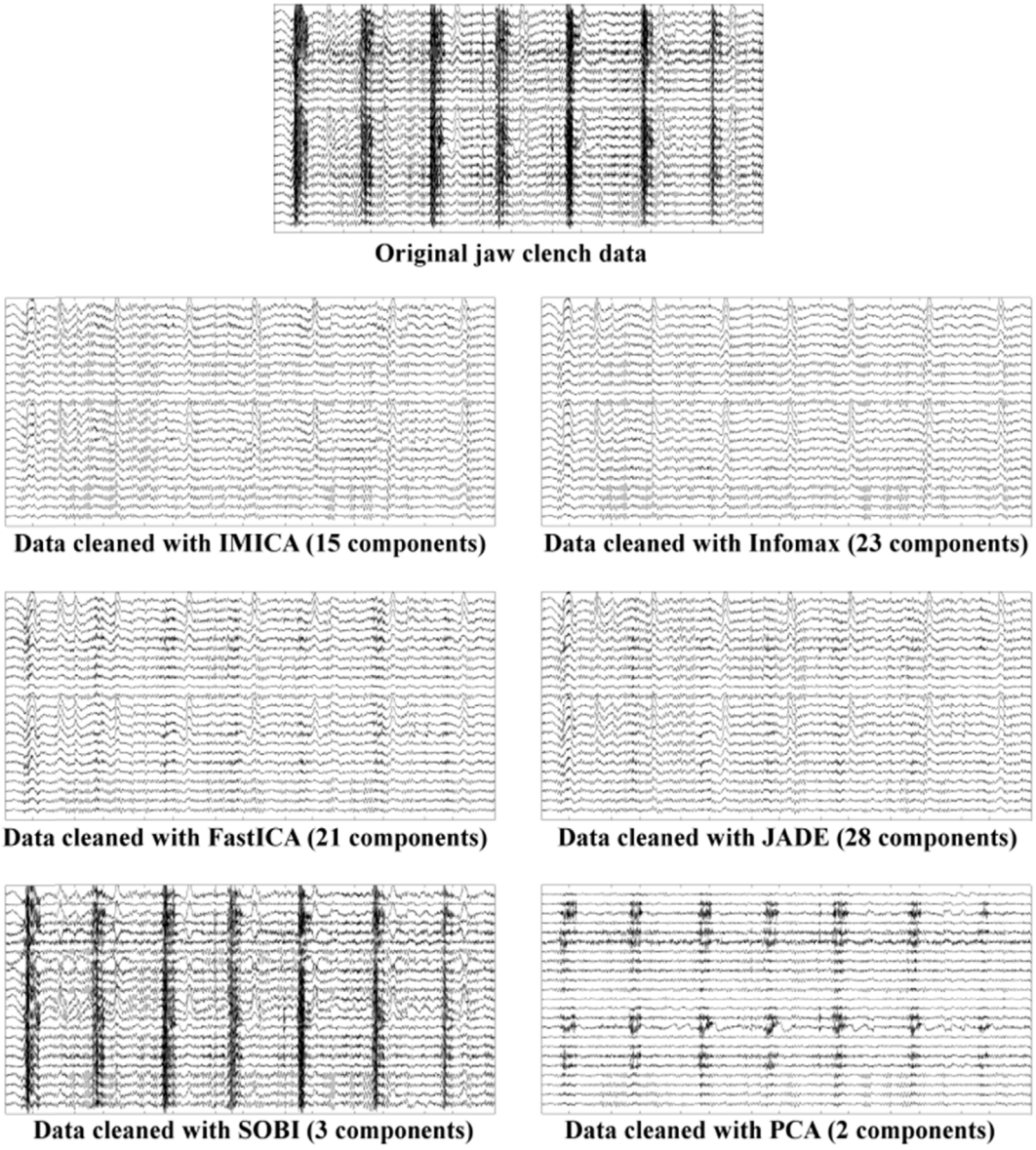

Fig. 4. Removal of jaw clench artifacts. (A) Original data. (B–G) data cleaned with (B) IMICA, (C) Infomax, (D) FastICA, (E) JADE, (F) SOBI, and (G) PCA. The number of components to remove was selected using the beta criterion.

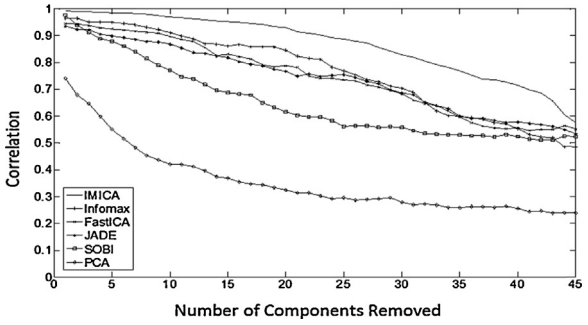

Fig. 5. Reduction in visual ERP as more components are used to isolate and remove the jaw clench artifact. Performance is measured as correlation values between the original ERP and the ERP obtained after removal of the artifact subspace.

The results for the ERP analysis are presented in Fig. 5 averaged across subjects and electrodes. These results show that IMICA removes less of the ERP than the non-reference methods during artifact removal regardless of the number of components removed. This is because IMICA creates a more focal decomposition that is designed to model specific phenomena and thus has less overlap with the other components in the data.

Next, we focused on the time–frequency dynamics when each subject pressed the response button. Prior research indicates that finger movements, such as those associated with button pressing, elicit changes in spectral power in the alpha, beta, and gamma bands (Pfurtscheller and Lopes da Silva, 1999). We visually identified clean trials for each subject. We computed the event-related spectral perturbation (ERSP) across all trials for each subject using both the original data as well as data in which the jaw clench subspace had been removed. We then computed the average correlation between the original ERSP and the new ERSP for the alpha, beta, and gamma bands. The results are presented in Fig. 6.

As with the ERP analysis IMICA preserves more of the original signal as the size of the artifact subspace grows than any of the other methods with the exception of SOBI. When few components are removed the performance of SOBI is equivalent to the other non-reference methods. However, as more components are removed the performance of SOBI tends to flatten out quicker, while

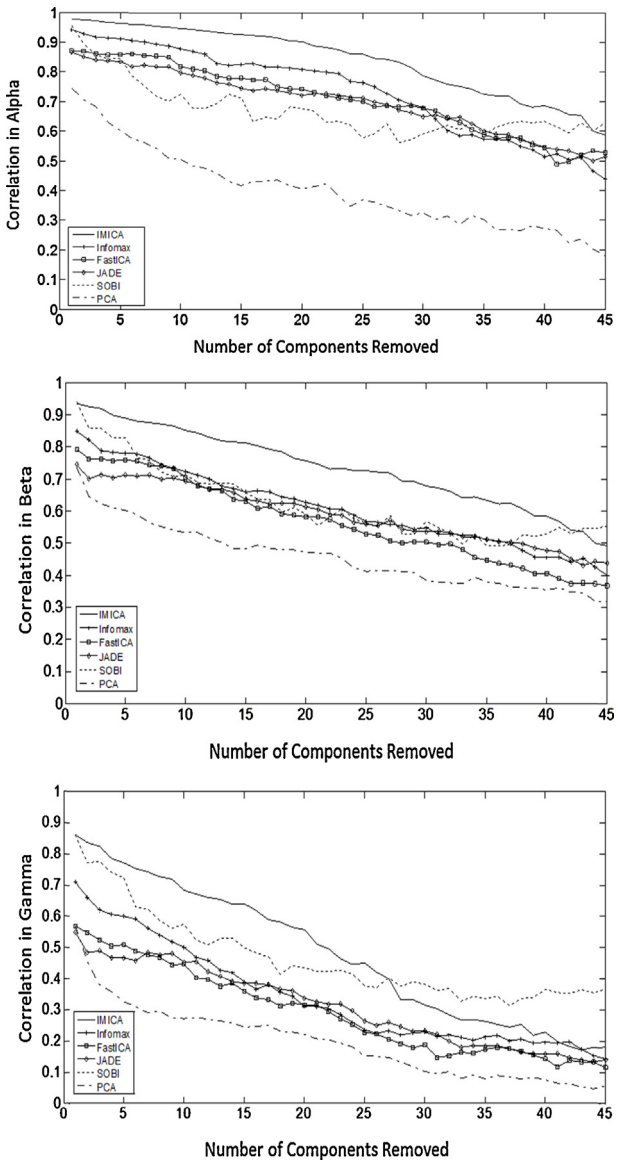

Fig. 6. Reduction in ERSP related to motor movements (button presses) as more components are removed. (Top) alpha band changes, (Middle) beta band changes, and (Bottom) gamma band changes.

the performance of both IMICA and the other non-reference methods continues to decrease. As mentioned earlier, SOBI is also the method that performs the worst at isolating and removing the target subspace.

Finally, we address the issue of how to determine the size of the target subspace when there is not a control condition available. We computed the LC for each subject, which are shown in Fig. 7. The dashed lines indicate the cut-off value predicted by our beta criterion and the dotted lines indicate the cut-off value predicted by our gamma criterion. A single line indicates that the two criteria agreed on the cut-off value.

As expected, a precise cut-off point is not obvious from the LC for any subject but rather the LC values predict a range of model sizes from which the experimenter must choose one solution. However, there is a strong relationship between the LCs and the model size estimates produced by our control condition. For each subject the control the control condition estimates are located at, or near, the point of inflection in the LC. But, as with any model selection process the final decision for the model size, based on the LC, is a function of both (1) the goals of the analysis and (2) the tradeoffs associated with different models.

4.3. Modeling alpha and beta activity

To demonstrate the use of our method beyond the scope of artifact modeling we analyzed data from a motor tapping study. In this study, subjects performed self-paced finger movements using the left index, left middle, right index, or right middle fingers. This produced the well-known alpha and beta synchronizations (i.e. increases in power) and desynchronizations (i.e. decreases in power), which could be most clearly observed over the contralateral motor cortex (Pfurtscheller and Aranibar, 1977; Pfurtschellar et al., 1999; Liao et al., 2014). An example is shown in Fig. 8, for one representative subject.

We used IMICA to model the alpha and beta activity for each subject. We first identified the unique alpha and beta bands in which the activity occurred. We computed the reference signal using the same approach that we used to model the jaw clench artifacts. We determined the optimal number of components using the LC method. Finally, for each subject we removed the background activity (non-alpha and non-beta subspace) by backprojecting only the alpha and beta subspaces to the scalp. Fig. 9 shows the LC for the alpha and beta subspaces (Top Row), the spectral responses using only the alpha and beta subspaces (Middle Row) and the residual spectral activity not included in the alpha and beta subspaces (Bottom Row) for the subject originally shown in Fig. 8. From the LC for this subject we chose to use four components to model the alpha activity and four components to model the beta activity.

From Fig. 9 it can be seen that IMICA has done an effective job at isolating and separating the alpha and beta activity from the rest of the EEG data for this subject. In the bottom row of Fig. 9 there is little-to-no visible event-related alpha and beta activity remaining after removal of the alpha and beta subspaces. In the middle row, which represents the backprojection of the alpha and beta subspaces, the neural activity is well preserved and, potentially, magnified. This indicates that IMICA was correctly able to model and separate the observed alpha and beta activity for this subject.

Finally, to compare performance across subjects we identified the point of inflection for each subject's alpha and beta time-spectral plots, i.e. the approximate time point in which the observed event-related desynchronization becomes an event-related synchronization (Figs. 8 and 9: alpha time = 1 s, beta time = 0.5 s). We computed the power ratio between the windows immediately before and after this inflection point, using a window size of 0.5 s, by subtracting the average power in the pre-inflection window from the average power in the post-inflection window. We then divided this power difference by the average power in the pre-inflection window. The results of the individually identified alpha and beta bands are presented in Figs. 10 and 11, respectively, using (1) the original scalp data, (2) only the backprojected alpha and beta subspaces, and (3) the residual, non-alpha and non-beta subspace. The subject from Figs. 8 and 9 is Subject 2 in Figs. 10 and 11.

For all subjects, except Subject 7, the application of IMICA resulted in the creation of independent subspaces that captured, and in most cases greatly improved the relative strength, of the event-related alpha and beta activity. For all subjects, except Subject 7, the residual activity had little-to-no measurable event-related power changes in the alpha and beta bands. Again, we argue that this is because the activity in these bands was cleanly separated by our IMICA algorithm. With respect to Subject 7, we found that there was no discernable alpha and beta activity in the original data. We do not know if this is because of poor signal quality or a unique aspect of this subject's neural dynamics, or both. When estimating

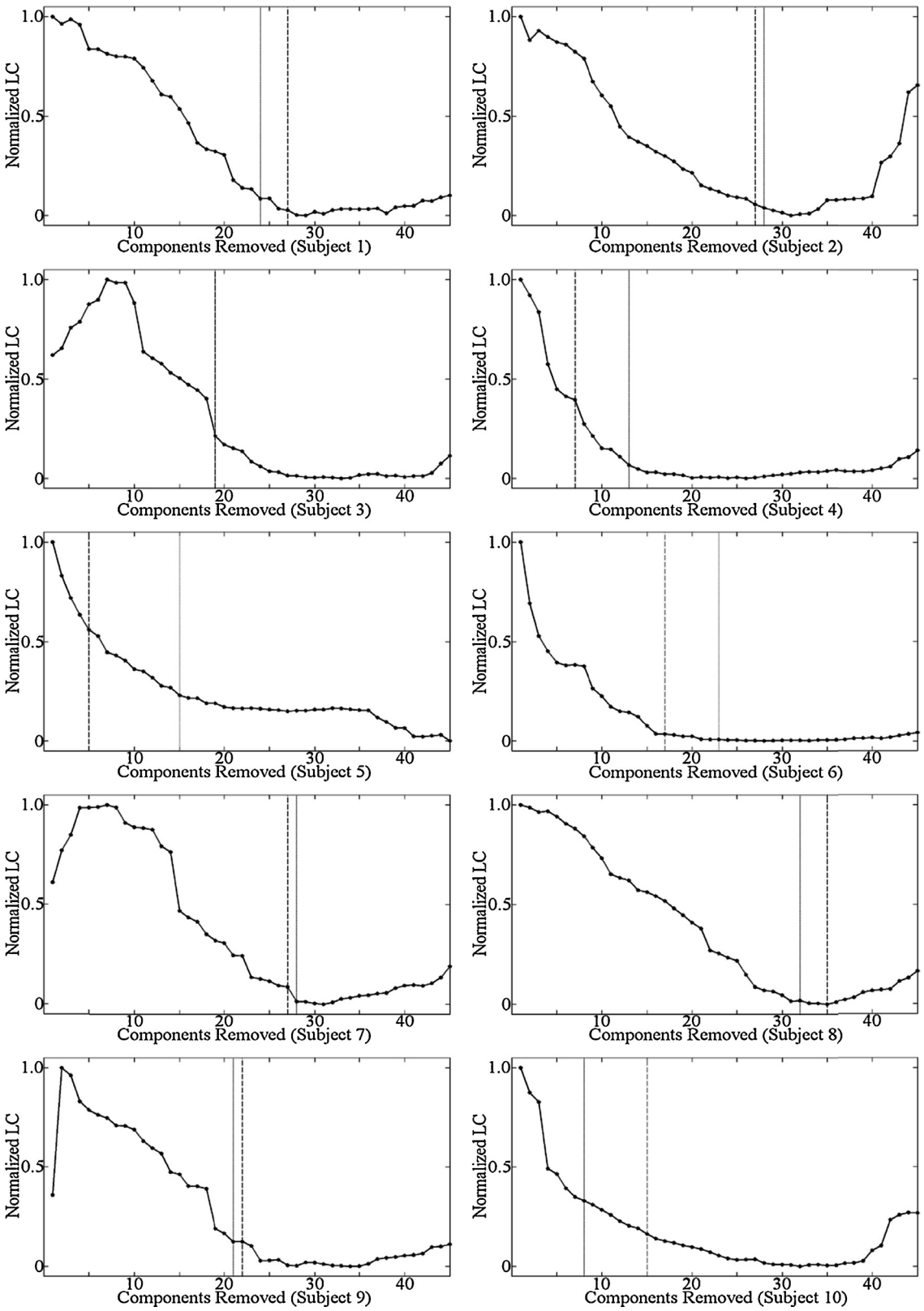

Fig. 7. LC curves for each subject. The dashed line indicates the model size predicted by the beta criterion from the control condition, while the dotted line indicates the model size predicted by the gamma criterion from the control condition.

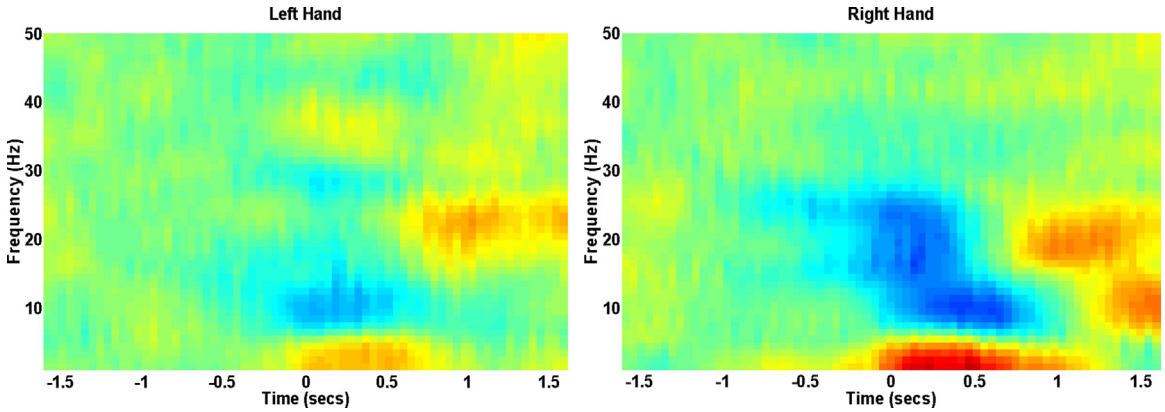

Fig. 8. Event-related spectral activity for the Left and Right hand over the contralateral motor cortex for one subject.

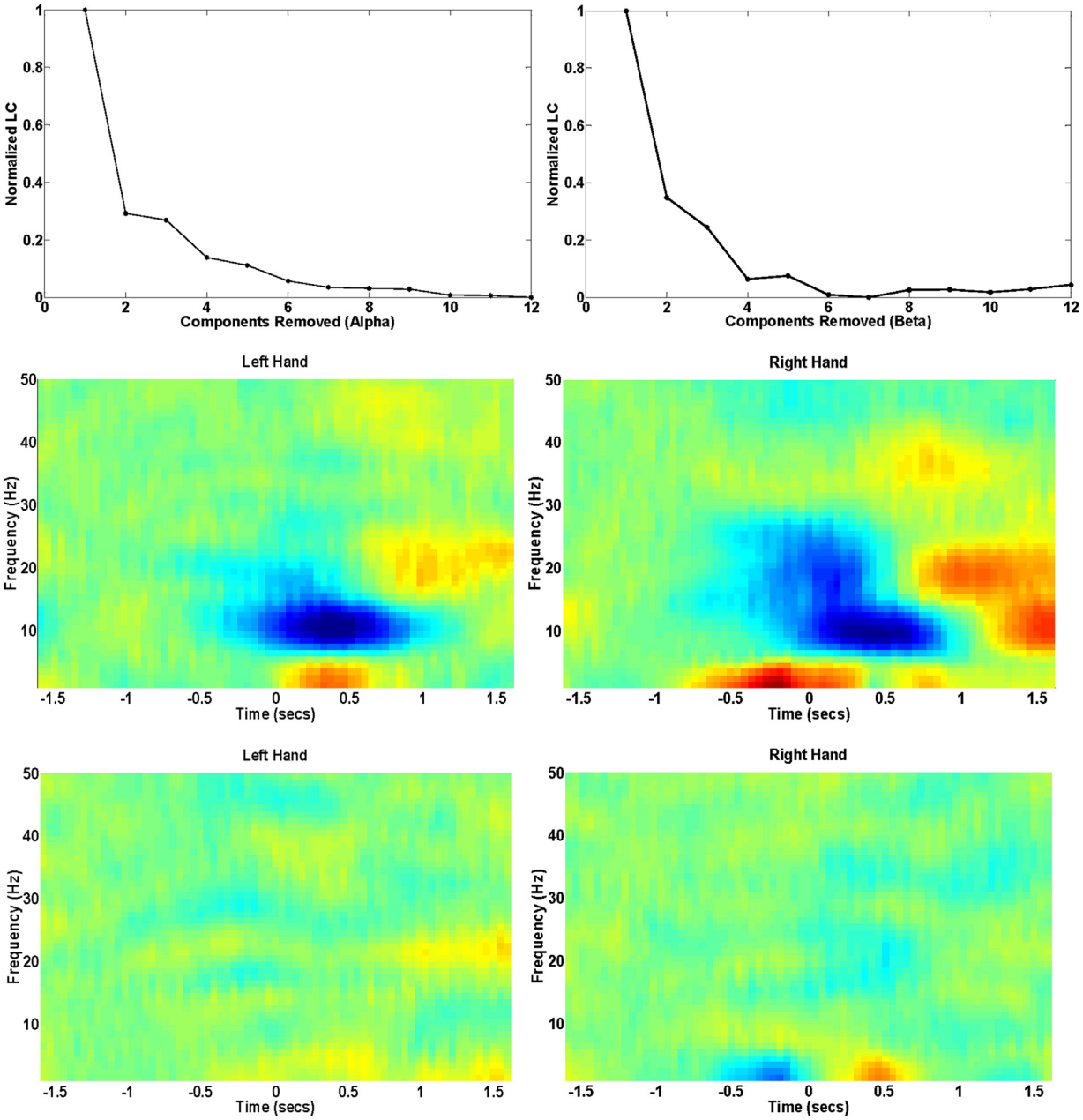

Fig. 9. The LC used to compute alpha (Top-Left) and beta (Top-Right) subspace size for one subject. Event-related alpha and beta activity for the Left (Middle-Left) and Right (Middle-Right) hand using only the identified alpha and beta subspaces. This activity is shown from the contralateral hemisphere, electrodes C4 and C3, respectively. Residual activity for the Left (Bottom-Left) and Right (Bottom-Right) hand after removal of the alpha and beta subspaces.

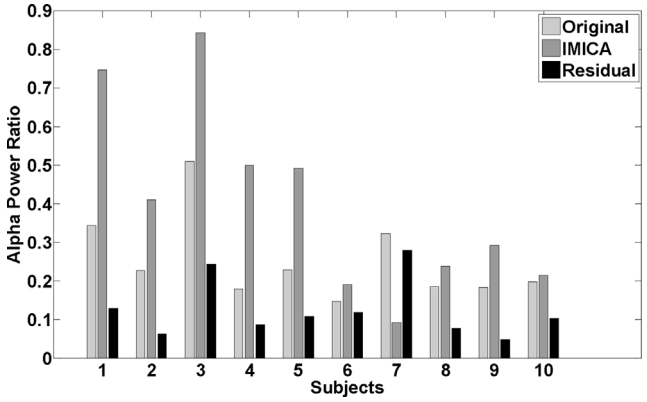

Fig. 10. Alpha power ratio for each subject using the original scalp data (left), alpha and beta subspaces identified with our IMICA algorithm (middle), and the residual non-alpha and non-beta activity (right).

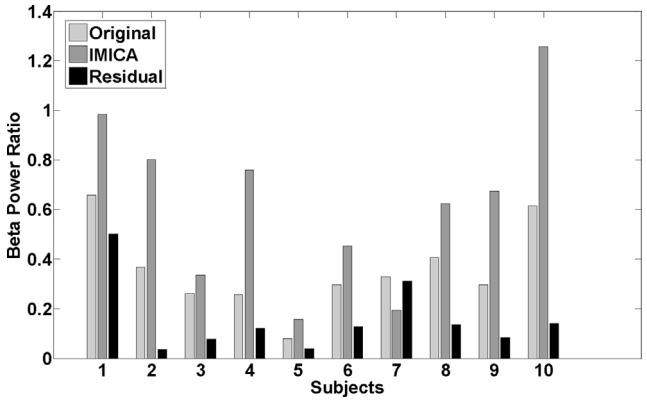

Fig. 11. Beta power ratio for each subject using the original scalp data (left), alpha and beta subspaces identified with our IMICA algorithm (middle), and the residual non-alpha and non-beta activity (right).

the alpha and beta bands for this subject, we were forced to make our best guess, which clearly did not improve the quality of the decomposition.

5. Conclusion

The goal for this work was to develop a better method for decomposing and separating information within EEG data. Current decomposition methods tend to suffer from the limitation that they are inherently blind and thus cannot leverage domain-specific knowledge to improve the overall quality of their results. In addition, these methods are not designed to selectively use data to create different components. We have shown that by incorporating additional information about the signals of interest we are able to create decompositions that better isolate that information, thus providing a cleaner separation of that information from the remaining data.

However, our IMICA algorithm is not necessarily applicable, or ideal, for all analyses. This method is best suited for decomposition problems in which the experimenter has a reasonable understanding of the target subspaces and can represent the features of those subspaces in some quantifiable way. This is exemplified by the poor performance of our algorithm on Subject 7 in the analysis of event-related alpha and beta activity. In this paper, and building on previous constrained ICA approaches, we created reference signals that captured the time-frequency properties of the target subspaces before subsequently creating components to match these reference signals. While this approach was sufficient for our current task, other criteria could be used so long as they could be encoded within a differentiable constraint function.

As the results here have shown, for complex subspaces such as EMG artifact activity there does not tend to be a single optimal model size. There is a tradeoff that must be made when determining the model size and the experimenter should tailor the process to the goals of the subsequent analysis. By overfitting the IMICA models the experimenter is sure to capture more of the signal of interest, at the cost of including more of the background activity. By underfitting, the experimenter will eliminate more of this noise, but will also lose more of the signal of interest.

In this paper we primarily focused on EEG artifact signals since artifact removal is one of the more common applications of ICA to EEG and because the SNR of the artifact signals simplified the task of comparing different approaches. We demonstrated the use of IMICA to the decomposition and analysis of neural data but we acknowledge there are still open questions with respect to this type of approach. In particular, what are the best ways to capture subtle neural responses and how accurately does one need to represent these responses to produce accurate subspaces? Also, how does reducing the dimensionality of the signal affect different subsequent analyses (e.g. time-based topography), and to what extent do the individual dimensions of the target subspaces approximate source components? Given the complexity of most modern analysis techniques, we feel these are questions that must be addressed in future work.

Finally, we have yet to focus on optimizing the computational complexity of IMICA. This complexity is especially problematic for large dimensional data sets (such as 256 channel EEG) in which the size of the target subspace is unknown. A benefit of IMICA is that the researcher can selectively use different epochs of the training data to perform different aspects of the decomposition, which can improve computation time. But, of course, when the size of the target subspace is not known beforehand it may be necessary to perform multiple decompositions to find the most optimal model size. We believe approximations to the independence projection may be useful, but have yet to explore these methods in detail.

Acknowledgements

The authors would like to thank Dr. Kelvin Oie of the Army Research Laboratory and Dr. Jason Metcalfe of DCS Corporation for helpful discussions and directions with respect to the experimental design. The authors would also like to thank Dr. Piotr Franaszczuk of the Army Research Laboratory for his overall guidance and instruction. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

Akaike H. Likelihood of a model and information criteria. J Econometrics 1981;16(1):3–14.
Anderer P, Roberts S, Schlögl A, Gruber G, Klösch G, Herrmann W, et al. Artifact processing in computerized analysis of sleep EEG—a review. Neuropsychobiology 1999;40(3):150–7.
Andreassi JL. Psychophysiology: human behavior and physiological response. Hillsdale, NJ: Lawrence Erlbaum; 2007.
Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comput 1995;7(6):1129–59.
Belouchrani A, Abed-Meraim K, Cardoso JF, Moulines E. A blind source separation technique using second-order statistics. IEEE Trans Signal Process 1997;45(2):434–44.
Bertsekas DP. Constrained optimization and Lagrange multiplier methods. Boston: Academic Press; 1982.

Cardoso JF, Souloumiac A. Blind beamforming for non-Gaussian signals. In: IEE proceedings F (radar and signal processing); 1993, December. p. 362–70.
Cardoso JF. Multidimensional independent component analysis. In: Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, IEEE; 1998, May. p. 1941–4, Vol. 4.
Cashero Z, Anderson C. Comparison of EEG blind source separation techniques to improve the classification of P300 trials. In: Engineering in Medicine and Biology Society, EMBC, 2011 annual international conference of the IEEE. IEEE; 2011, Augus). p. 7183–6.
Comon P. Independent component analysis, a new concept? Signal Process 1994;36(3):287–314.
Crespo-Garcia M, Atienza M, Cantero JL. Muscle artifact removal from human sleep EEG by using independent component analysis. Ann Biomed Eng 2008;36(3):467–75.
De Clercq W, Vergult A, Vanrumste B, Van Hees J, Palmini A, Van Paesschen W, et al. A new muscle artifact removal technique to improve the interpretation of the ictal scalp electroencephalogram. In: IEEE-EMBS 2005: 27th annual international conference of the engineering in medicine and biology society, 2005, IEEE; 2006, January. p. 944–7.
Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S. Independent EEG sources are dipolar. PLoS ONE 2012;7(2):e30135.
Eichele T, Calhoun VD, Debener S. Mining EEG–fMRI using independent component analysis. Int J Psychophysiol 2009;73(1):53–61.
Glass KA, Frishkoff GA, Frank RM, Davey C, Dien J, Malony AD, et al. A framework for evaluating ICA methods of artifact removal from multichannel EEG. In: Independent component analysis and blind signal separation. Berlin, Heidelberg: Springer; 2004. p. 1033–40.
Gordon SM, Franaszczuk PJ, Hairston WD, Vindiola M, McDowell K. Comparing parametric and nonparametric methods for detecting phase synchronization in EEG. J Neurosci Methods 2013;212(2):247–58.
Groppe DM, Makeig S, Kutas M. Identifying reliable independent components via split-half comparisons. NeuroImage 2009;45(4):1199–211.
Gwin JT, Gramann K, Makeig S, Ferris DP. Removal of movement artifact from high-density EEG recorded during walking and running. J Neurophysiol 2010;103(6):3526–34.
Haykin SS. Neural networks and learning machines, vol. 3. Upper Saddle River: Pearson Education; 2009.
Hyvärinen A, Köster U. FastISA: a fast fixed-point algorithm for independent subspace analysis. In: European symposium on artificial neural networks; 2006, April. p. 371–6.
Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Networks 2000;13(4):411–30.
Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Networks 1999;10(3):626–34.
Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C, et al. Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. J Clin Neurophysiol 2003;20(4):249–57.
James CJ, Gibson OJ. Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis. IEEE Trans Biomed Eng 2003;50(9):1108–16.
Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, Sejnowski TJ. Extended ICA removes artifacts from electroencephalographic recordings. Adv Neural Inf Process Syst 1998a;10:894–900.
Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, et al. Removing electroencephalographic artifacts: comparison between ICA and PCA. In: Neural networks for signal processing VIII: proceedings of the 1998 IEEE signal processing society workshop, IEEE; 1998, August. p. 63–72.
Jung TP, Makeig S, Humphries C, Lee TW, Mckeown MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 2000;37(02):163–78.
Kachenoura A, Albera L, Senhadji L, Comon P. ICA: a potential tool for BCI systems. IEEE Signal Process Mag 2008;25(1):57–68.
Karvanen J. A resampling test for the total independence of stationary time series: application to the performance evaluation of ica algorithms. Neural Process Lett 2005;22(3):311–24.
Knuth KH. Informed source separation: a Bayesian tutorial. In: 2005 13th European signal processing conference. IEEE; 2005, September. p. 1–8.
Kobayashi K, Merlet I, Gotman J. Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. Clin Neurophysiol 2001;112(3):405–13.
Lawhern V, Hairston WD, McDowell K, Westerfield M, Robbins K. Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. J Neurosci Methods 2012;208(2):181–9.
Lee TW, Lewicki MS, Girolami M, Sejnowski TJ. Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Process Lett 1999a;6(4):87–90.
Lee TW, Girolami M, Sejnowski TJ. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput 1999b;11(2):417–41.
Liao K, Xiao R, Gonzalez J, Ding L. Decoding individual finger movements from one hand using human EEG signals. PLoS ONE 2014;9(1):e85192.
Lu W, Rajapakse JC. Eliminating indeterminacy in ICA. Neurocomputing 2003;50:271–90.
Lu W, Rajapakse JC. Approach and applications of constrained ICA. IEEE Trans Neural Networks 2005;16(1):203–12.
Ma S, Li XL, Correa NM, Adali T, Calhoun VD. Independent subspace analysis with prior information for fMRI data. In: 2010 IEEE international conference on acoustics speech and signal processing (ICASSP), IEEE; March 2010. p. 1922–5.
Machens CK. Demixing population activity in higher cortical areas. Front Comput Neurosci 2010;4.
Matic V, Deburchgraeve W, Van Huffel S. Comparison of ICA algorithms for ECG artifact removal from EEG signals. In: Proc. of the 4th annual symposium of the IEEE-EMBS Benelux Chapter (IEEE-EMBS); 2009, November. p. 137–40.
Mennes M, Wouters H, Vanrumste B, Lagae L, Stiers P. Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. Psychophysiology 2010;47(6):1142–50.
Mi J-X. A novel algorithm for independent component analysis with reference and methods for its applications. PLoS ONE 2014;9(5):e93984, http://dx.doi.org/10.1371/journal.pone.0093984.
Milanesi M, Martini N, Vanello N, Positano V, Santarelli MF, Landini L. Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals. Med Biol Eng Comput 2008;46(3):251–61.
Niazy RK, Beckmann CF, Iannetti GD, Brady JM, Smith SM. Removal of FMRI environment artifacts from EEG data using optimal basis sets. NeuroImage 2005;28(3):720–37.
Nunez PL, Srinivasan R. Electric fields of the brain: the neurophysics of EEG. New York, NY: Oxford University Press; 2006.
Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev 2006;30(6):808–22.
Palmer JA, Kreutz-Delgado K, Makeig S. Super-Gaussian mixture source model for ICA. In: Rosca J, Erdogmus D, Principe JC, Haykin S, editors. Lecture notes in computer science. Berlin: Springer; 2006. p. 854–61.
Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. NeuroImage 2005;28(2):326–41.
Pfurtscheller G, Aranibar A. Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalogr Clin Neurophysiol 1977;42(6):817–26.
Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 1999;110(11):1842–57.
Plöchl M, Ossandón JP, König P. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. Front Hum Neurosci 2012;6.
Póczos B, Lõrincz A. Independent subspace analysis using geodesic spanning trees. In: Proceedings of the 22nd international conference on machine learning. ACM; 2005, August. p. 673–80.
Safieddine D, Kachenoura A, Albera L, Birot G, Karfoul A, Pasnicu A, et al. Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. EURASIP J Adv Signal Process 2012;2012(1):1–15.
Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6(2):461–4.
Shao SY, Shen KQ, Ong CJ, Wilder-Smith E, Li XP. Automatic EEG artifact removal: a weighted support vector machine approach with error correction. IEEE Trans Biomed Eng 2009;56(2):336–44.
Srivastava G, Crottaz-Herbette S, Lau KM, Glover GH, Menon V. ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner. NeuroImage 2005;24(1):50–60.
Theis FJ. Towards a general independent subspace analysis. In: Advances in neural information processing systems; 2006. p. 1361–8.
Vincent E, Bertin N, Gribonval R, Bimbot F. From blind to guided audio source separation: how models and side information can improve the separation of sound. IEEE Signal Process Mag 2014;31(3):107–15.
Vos DM, Riès S, Vanderperren K, Vanrumste B, Alario FX, Huffel VS, et al. Removal of muscle artifacts from EEG recordings of spoken language production. Neuroinformatics 2010;8(2):135–50.
Wang Y, Jung TP. Improving brain–computer interfaces using independent component analysis. In: Towards practical brain–computer interfaces. Berlin, Heidelberg: Springer; 2013. p. 67–83.
Wang G, Qiu L, Wang G, Wang Z. Constraint setting in constrained ICA. In: 2006 eighth international conference on signal processing, IEEE; 2006, Vol. 3.