

# STRUCTURED DICTIONARY LEARNING FOR SPARSE COMMON COMPONENT AND INNOVATION MODEL

*Arash Golibagh Mahyari and Selin Aviyente*

Department of Electrical and Computer Engineering  
Michigan State University  
East Lansing, MI 48823 USA  
Email: mahyari@egr.msu.edu and aviyente@egr.msu.edu

## ABSTRACT

Event-related potentials (ERPs) are electrophysiological responses that are commonly used for detecting the brain response to external stimuli. In this paper, we propose to use the sparse common component and innovations model (SCCI) to extract ERPs from multiple EEG signals recorded across closely located electrodes. This model finds the sparse representation of the common component of the signals and their innovation components with respect to pre-determined common and innovation dictionaries, where the common component refer to an event captured by adjacent electrodes such as ERPs. However, different stimuli may produce different responses and pre-determining the dictionary may not always be optimal. Therefore, we introduce a structured dictionary learning method to simultaneously learn the two dictionaries from training data. The proposed method is applied to a study of error monitoring where two different types of brain responses are elicited corresponding to the decision made by the subject. The learned dictionaries can discriminate between the response types and extract the ERP corresponding to the two responses.

**Index Terms**— Distributed Compressive Sensing, Dictionary Learning, Jointly Sparse Signal Model, Electroencephalography, Event-related potentials.

## 1. INTRODUCTION

Event-related potentials (ERPs) have been used extensively for the detection of neural diseases and the design of brain computer interfaces (BCI) [1, 2]. The difficulty with extracting ERPs is that their amplitude is small compared to the background activity. Different approaches such as averaging, independent component analysis (ICA), and wavelets [3] have been proposed [4]. Since ERPs are time-locked, it is assumed that averaging will add-up the effect of ERP while attenuating the non time-locked on going EEG activity. ICA is used to separate the evoked potential from the background activity using the signals recorded by nearby electrodes [4] assuming that ERP and ongoing EEG activity are statistically independent, which is not necessarily true. Wavelet based approaches, on the other hand, identify the coefficients associated with ERPs statistically [3]. This method requires selecting the threshold value, and assumes ERP lies in a subspace of wavelet space. ERPs have also been extracted [5] by principle component analysis with the assumption that after averaging the variance of the background activity decreases while the

variance of ERP has a direction which can be found using PCA. This assumption is true only if the background activity is noise, where in practice it corresponds to other brain events that are not of interest.

ERP signals such as P300, and error-related negativity (ERN) are usually similar across neighboring electrodes due to volume conduction [6]. Volume conduction refers to activity that may come from a single source but is observed at multiple scalp electrodes [7]. Therefore, ERP extraction from background EEG activity can be performed by extracting the common component of the set of neighboring electrodes. Sparse common component and innovations (SCCI) has been proposed as a plausible model for describing the activity across neighboring electrodes [8, 9]. This model has been commonly used in distributed compressive sensing problems where sensors distributed across a field capture a jointly sparse common component plus innovation signals and have been recently used for compressing EEG signals [10]. Due to the inter-signal correlation, jointly sparse signals are usually assumed to be composed of a common sparse component which is shared by all signals, and an innovation component which is unique to each signal [11].

In SCCI, the sparsity of the two parts of the signals is usually described through pre-defined dictionaries. However, as the literature in dictionary learning indicates, pre-defined dictionaries are not necessarily optimal in terms of sparsity and reconstruction error [12]. The well-known dictionary learning algorithms such as K-SVD and method of optimal directions (MOD) [13] cannot be directly used for SCCI as they do not guarantee the preservation of the special structure of the SCCI dictionary. Recently, Chen *et al.* [14] proposed Joint Orthogonal Matching Pursuit (JOMP) with dictionary update for SCCI. In each iteration, only one atom for representing either the common or the innovation components is selected based on the correlation between the residue and the atoms of the input dictionary. However, when combined with dictionary learning, this method only learns a single dictionary for the common and innovation components, where some atoms belong to the common component and others belong to the innovation components. As a result, when JOMP recovers the sparse representation of the common component of the test data with respect to this dictionary, it may recover the common component based on the atoms which were learned to represent the innovation component. On the other hand, it may recover the innovation component of the test data based on the atoms learned to represent the common component of the training data. Barthlemy *et al.* [15] proposed a data-driven method which takes into account the inter-channels links in the spatial multivariate model, and shift-invariance is used for the temporal model. The proposed algorithm is outperformed the Gabor dictionary since Gabor dictionary lacks the spatial flexibility for inter-channel links.

This work was in part supported by the National Science Foundation under Grant No. CCF-1422262.

In this paper, we propose a two-stage dictionary learning algorithm which addresses these problems by learning two separate dictionaries for the common and innovation components simultaneously. In the proposed structured K-SVD method, first the common component dictionary is updated the coefficient vector and the training data, and then the innovation component dictionary. The procedure to update the common and innovation dictionaries are similar to K-SVD, where the updated atom is the singular vector of the error matrix. Unlike K-SVD, the structured K-SVD decomposes the main dictionary to the common and innovation dictionaries and learns these subictionaries. The structured K-SVD is used to learn the common and innovation dictionaries for error-related negativity (ERN) and the correct-related negativity (CRN) from training data. ERN and CRN are ERPs that are commonly used to index cognitive control [16]. Finally, the learned dictionaries are used to extract the two ERPs corresponding to different responses.

## 2. BACKGROUND

### 2.1. Sparse Common Component and Innovations Model

Distributed compressive sensing assumes that signals acquired across multiple sensors such as EEG signals are jointly sparse in a dictionary. Due to the inter-signal correlation, jointly sparse signals are usually assumed to be composed of a common sparse component which is shared by all signals, and an innovation component which is unique to each signal [11].

Given a set of signals  $\Lambda = \{\mathbf{x}_j \in \mathbb{R}^N; \forall j \in \{1, 2, \dots, J\}\}$  that are jointly sparse, and using the joint sparsity model (JSM) [17, 9] with a common component  $\mathbf{z}_c \in \mathbb{R}^N$  and an innovation component  $\mathbf{z}_j \in \mathbb{R}^N$ , the signal  $\mathbf{x}_j$  can be written as:  $\mathbf{x}_j = \mathbf{z}_c + \mathbf{z}_j$ ,  $j = 1, 2, \dots, J$ . The common and innovation components of the set of signals  $\Lambda$  are sparse with respect to two different sets of bases,  $\phi_c$  and  $\phi_j$ , respectively, as:  $\mathbf{z}_c = \phi_c \theta_c$  and  $\mathbf{z}_j = \phi_j \theta_j$ ,  $j \in \{1, 2, \dots, J\}$ , where  $\theta_c$  and  $\theta_j$  are the coefficient vectors. Since the signal  $\mathbf{x}_j$  is sparse in the transform domain, the coefficient vectors have a small number of nonzero entities,  $\|\theta_c\|_0 = K_c \ll N$  and  $\|\theta_j\|_0 = K_j \ll N$ .

In order to recover the sparse representation of the set of signals  $\Lambda$ , all signals are stacked to form a single optimization problem.

$$\mathbf{X} = \Phi \Theta, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_J \end{bmatrix}, \Phi = \begin{bmatrix} \phi_c & \phi_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \phi_c & \mathbf{0} & \phi_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \phi_c & \mathbf{0} & \cdots & \mathbf{0} & \phi_J \end{bmatrix}, \Theta = \begin{bmatrix} \theta_c \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_J \end{bmatrix}, \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{JN \times 1}$ ,  $\Phi \in \mathbb{R}^{JN \times (J+1)N}$ , and  $\Theta \in \mathbb{R}^{JN \times 1}$ .

The goal is to find the coefficient vector  $\Theta$  such that the error between the signal  $\mathbf{X}$  and the sparse representation  $\Phi \Theta$  is minimized:

$$\hat{\Theta} = \arg \min \|\Theta\|_0 \text{ s.t. } \|\mathbf{X} - \Phi \Theta\|_2^2 \leq \epsilon, \quad (2)$$

where  $\epsilon$  is the bound on the recovery error. Since the optimization problem in Eq. 2 is NP-hard and non-convex, the  $l_1$ -norm minimization is commonly used [17, 9].

### 2.2. K-SVD Algorithm

In traditional sparse signal recovery, predefined basis sets such as Fourier, wavelets or their combinations are commonly used to find the sparse representation of signals. However, in many applications,

the dictionary representing the sparse representation of the signal is not known. Dictionary learning (DL) has been developed to address this problem [18, 12]. Current dictionary learning algorithms such as K-SVD [12] and MOD [13] focus on signals that have a common sparse representation, and not on SCCI model described in Section II.A.

Dictionary learning algorithms try to find the sparse coefficient vectors  $\mathbf{A} = [\alpha_1, \alpha_2, \dots, \alpha_J]$  and the dictionary  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_L]$  for a set of training signals  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J]$  in the model  $\mathbf{Y} = \mathbf{D}\mathbf{A}$  such that:

$$\min_{\mathbf{A}, \mathbf{D}} \{\|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2\} \quad \text{s. t. } \forall i \|\alpha_i\|_0 \leq K \quad (3)$$

where  $K$  is the sparsity level of the coefficient vectors,  $\mathbf{y}_j \in \mathbb{R}^N$ ,  $\alpha_j \in \mathbb{R}^L$ ,  $\mathbf{Y} \in \mathbb{R}^{N \times J}$ ,  $\mathbf{D} \in \mathbb{R}^{N \times L}$ ,  $\mathbf{A} \in \mathbb{R}^{L \times J}$ , and  $L \geq N$ .

K-SVD minimizes Eq. 3 iteratively by first finding the coefficient vectors,  $\mathbf{A}$ , while the dictionary,  $\mathbf{D}$ , is fixed and then updating  $\mathbf{D}$ . After initializing the dictionary  $\mathbf{D}^{(0)}$ , in the first stage, K-SVD uses a pursuit algorithm to find the sparse coefficient vector at each iteration  $k$ ,  $\alpha_j^{(k)}$ , for each signal  $\mathbf{y}_j, j \in \{1, 2, \dots, J\}$  through solving the following optimization problem:

$$\min_{\alpha_j^{(k)}} \|\alpha_j^{(k)}\|_0 \quad \text{s. t. } \|\mathbf{y}_j - \mathbf{D}^{(k)} \alpha_j^{(k)}\|_2^2 \leq \epsilon \quad \forall j \in \{1, 2, \dots, J\}. \quad (4)$$

Once the sparse coefficients are computed, K-SVD updates one column of  $\mathbf{D}^{(k)}$  at a time, fixing all columns in  $\mathbf{D}^{(k)}$  except  $\mathbf{d}_l^{(k)}$ , and updating the  $l$ th column vector as  $\mathbf{d}_l^{(k+1)}$ . In order to update the  $l$ th column  $\mathbf{d}_l$ ,  $\omega_l$  is defined as the group of indices pointing to the training signals  $\mathbf{y}_j$ s that use the atom  $\mathbf{d}_l$ , which is equivalent to the nonzero entries of  $\alpha_l^T$ , where  $\alpha_l^T$  is the  $l$ th row of  $\mathbf{A}$ , as:  $\omega_l = \{j | 1 \leq j \leq J, \alpha_l^T(j) \neq 0\}$ . The overall representation error matrix  $\mathbf{E}_l$  is computed as:

$$\mathbf{E}_l = \mathbf{Y} - \sum_{f \neq l} \mathbf{d}_f \alpha_f^T. \quad (5)$$

The matrix  $\mathbf{E}_l^R$  is obtained by selecting the columns of  $\mathbf{E}_l$  with the columns identified in  $\omega_l$ . After applying SVD to the matrix  $\mathbf{E}_l^R = \mathbf{U}\Delta\mathbf{V}^T$ , the updated dictionary column  $\mathbf{d}_l^{k+1}$  is equal to the first column of  $\mathbf{U}$  whereas the updated coefficient vector  $\alpha_l^R$  is the first column of  $\mathbf{V}$  multiplied by  $\Delta(1, 1)$ . This process is repeated until the stopping criteria are met.

## 3. COMMON AND INNOVATION DICTIONARIES LEARNING

The challenge with using dictionary learning algorithms such as K-SVD in SCCI is that they cannot be applied directly to Eq. 2. The dictionary in these equations,  $\Phi \in \mathbb{R}^{JN \times (J+1)L}$ , is constructed from different sets of basis,  $\phi_c \in \mathbb{R}^{N \times L}$  and  $\phi_j \in \mathbb{R}^{N \times L}$ <sup>1</sup>. Thus, some entities of the dictionary  $\Phi$  are constrained to zero (as depicted in Eq. 1). However, no existing dictionary learning algorithm guarantees that this structure of  $\Phi$  will be preserved during the update step of the dictionary. Therefore, we propose a structured dictionary learning algorithm based on K-SVD to fulfill this constraint and learn the dictionary  $\Phi$  from the training samples. To learn the common dictionary, we need multiple observations of the common component. However, the whole set of signals share one common component. Therefore, we randomly select  $J$  training signals from the whole set of training signals to learn the dictionaries, and repeat the process to update  $\Phi$ .

<sup>1</sup>Note that we assume one unique  $\phi_j$  for all  $J$  innovation components

After initializing the bases  $\phi_c$  and  $\phi_j$  and constructing  $\Phi^{(k)}$  at the  $k$ th iteration, the coefficient vector  $\Theta^{(k)} \in \mathbb{R}^{(J+1)L \times 1}$  is estimated, as:

$$\hat{\Theta}^{(k)} = \arg \min \|\Theta^{(k)}\|_0 \text{ s.t. } \|\mathbf{X} - \Phi^{(k)}\Theta^{(k)}\|_2^2 \leq \epsilon. \quad (6)$$

where  $\mathbf{X} \in \mathbb{R}^{JN \times 1}$  is the input set of training signals as in Eq. 1.

After updating the coefficient vector, dictionaries  $\phi_c$  and  $\phi_j$  are updated separately in similar manner to K-SVD. In order to update  $\phi_c^{(k)}$ ,  $\phi_j^{(k)}$  is assumed to be fixed and  $\phi_c^{(k)} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_L]$ . The vector of input signals  $\mathbf{X} \in \mathbb{R}^{N \times J \times 1}$  is rearranged as a matrix  $\mathbf{Y} = [\mathbf{x}_1, \dots, \mathbf{x}_J] \in \mathbb{R}^{N \times J}$ , and the coefficient vector  $\hat{\theta}_c \in \mathbb{R}^{L \times 1}$  for the common part is concatenated to form the matrix  $\mathbf{A}_c^{(k)} = [\hat{\theta}_c, \dots, \hat{\theta}_c] \in \mathbb{R}^{L \times J}$ . Since all training signals have the same common component, they all either use the  $l$ th atom or not. Thus,  $\omega_{c,l}$  is defined as a row vector of ones with size  $J$ ,  $\mathbf{1} \in \mathbb{R}^{1 \times J}$  if the common component includes the  $l$ th atom of the common dictionary, or a row vector of zeros with size  $J$ ,  $\mathbf{0} \in \mathbb{R}^{1 \times J}$ . The error corresponding to the  $l$ th atom,  $\mathbf{E}_{c,l} \in \mathbb{R}^{N \times J}$ , is defined as:

$$\mathbf{E}_{c,l} = \mathbf{Y} - \sum_{f \neq l} \mathbf{d}_f \theta_{c,T}^f. \quad (7)$$

The matrix  $\mathbf{E}_{c,l}^R$  is obtained by selecting the columns of  $\mathbf{E}_{c,l}$  with the columns identified in  $\omega_{c,l}$ .  $\mathbf{d}_l^{(k+1)}$  is the first eigenvector of  $\mathbf{E}_{c,l}^R = \mathbf{U}_c \mathbf{\Delta}_c \mathbf{V}_c^T$  while its corresponding coefficient vectors are updated by replacing them with the first column of  $\mathbf{V}_c$  multiplied by  $\Delta_c(1, 1)$ .

Once all the columns of  $\phi_c^{(k+1)}$  are updated, the second dictionary  $\phi_j^{(k)}$  is updated in a similar manner assuming  $\phi_c^{(k+1)}$  is fixed. The dictionary  $\phi_j^{(k)} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L]$ , and the vector of coefficients  $\{\hat{\theta}_j^{(k)}\}_{j=1}^J$  from  $\Theta^{(k)}$  is rearranged to  $\mathbf{A}_j^{(k)} = [\hat{\theta}_1, \dots, \hat{\theta}_J] \in \mathbb{R}^{L \times J}$ . The set  $\omega_{j,l} = \{i | 1 \leq i \leq J, \theta_{j,T}^i(i) \neq 0\}$  is formed to indicate which signals use the  $l$ th atom of the innovation dictionary. The error matrix  $\mathbf{E}_{j,l}$  is then computed based on the signals identified in  $\omega_{j,l}$ , as:

$$\mathbf{E}_{j,l} = \mathbf{Y} - \sum_{f=1, f \neq l}^N \mathbf{c}_f \theta_{j,T}^f, \quad (8)$$

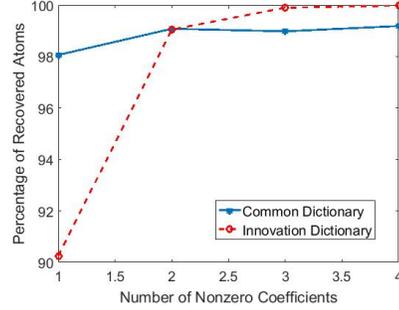
where  $\theta_{j,T}^f$  is the  $f$ th row of  $\mathbf{A}$ . The matrix  $\mathbf{E}_{j,l}^R$  is obtained by selecting the columns of  $\mathbf{E}_{j,l}$  with the columns identified in  $\omega_{j,l}$ . The first eigenvector of  $\mathbf{E}_{j,l}^R$  is considered as  $\mathbf{c}_l^{(k+1)}$ .

Once  $\phi_c^{(k+1)}$  and  $\phi_j^{(k+1)}$  are updated completely, the main dictionary  $\Phi^{(k+1)}$  is constructed again and the coefficient vector  $\hat{\Theta}^{(k+1)}$  is updated using Eq. 6. This process is repeated until convergence. The pseudocode of the structured dictionary learning for SCCI is presented in Algorithm 1.

## 4. EXPERIMENTAL RESULTS

### 4.1. Synthetic Data

The proposed structured dictionary learning is first evaluated on a set of synthesized signals. Two random matrices  $\phi_c \in \mathbb{R}^{32 \times 32}$  and  $\phi_j \in \mathbb{R}^{32 \times 32}$  are generated with i.i.d. uniformly distributed entries from  $U(0, 1)$  and  $U(5, 6)$ , respectively. Each column is normalized to a unit  $l_2$ -norm. 150 signals  $\{\mathbf{x}_i\}_{i=1}^{150}$ , where  $\mathbf{x}_i \in \mathbb{R}^{32 \times 1}$ , are produced by adding the common component  $\mathbf{z}_c \in \mathbb{R}^{32 \times 1}$  and the innovation components  $\mathbf{z}_i \in \mathbb{R}^{32 \times 1}$ . The common component  $\mathbf{z}_c$  is a linear combination of  $K_c$  atoms of  $\phi_j$ , with uniformly distributed i.i.d. coefficients in random and independent locations,  $U(0, 1)$ . Similarly, the innovation components  $\mathbf{z}_i$  are different linear combinations of  $K_j$  atoms of  $\phi_j$ .



**Fig. 1.** Percentage of successfully recovered atoms for the common and innovation dictionaries versus the number of sparse coefficients.

In order to apply K-SVD to synthetic signals, the dictionaries  $\phi_c$  and  $\phi_j$  are initialized with the training data. After constructing  $\Phi$  and  $\mathbf{X}$ , the coefficient vector  $\hat{\Theta}$  is found using orthogonal matching pursuit (OMP) with  $K_c + (K_j \times 150)$  coefficients, where  $K = K_c = K_j$  varying from 2 to 5. The maximum number of iterations was set to 20. The computed dictionaries  $\hat{\phi}_c$  and  $\hat{\phi}_j$  are compared separately against the generating dictionaries  $\phi_c$  and  $\phi_j$ . This comparison is done by sweeping through the columns of the generating dictionary  $\phi_c$  (or  $\phi_j$ ) and finding the closest column (in  $l_2$  distance) in the computed dictionary  $\hat{\phi}_c$  (or  $\hat{\phi}_j$ ). The distance is measured via  $1 - |\mathbf{d}_i^T \hat{\mathbf{d}}_i|$ , where  $\mathbf{d}_i$  is an atom of  $\phi_c$  (or  $\phi_j$ ) and  $\hat{\mathbf{d}}_i$  is the corresponding atom in the learned dictionary. A distance less than 0.1 for  $\phi_c$  and less than 0.05 for  $\phi_j$  are considered a success. The simulations are repeated 500 times, and the number of successes in each simulation is computed. Fig. 1 displays the percentage of successfully recovered atoms for the common and innovation dictionaries. Overall, the percentage of recovered atoms of the common and innovation dictionaries are more than 90%, and as the sparsity level increases, the accuracy increases. The number of the recovered atoms of the common dictionary is less than that of the innovation dictionary.  $J$  different innovation coefficients  $\{\theta_j\}_{j=1}^J$  are used to update the innovation dictionary. However,  $J$  different common coefficients belong to the same common component, which means they all carry redundant information. Thus, the atoms of the common dictionary are updated based on only one common coefficient vector.

### 4.2. EEG Data

The proposed algorithm is applied to a set of EEG signals<sup>2</sup> containing ERN and CRN responses, which are ERPs that are commonly used to index cognitive control [16]. The ERN is a brain potential response that occurs following performance errors in a speeded reaction time task usually 25-75 ms after the response [19]. In this study, we focus on the EEG recordings for 16 female participants. A speeded-response flanker task was employed, and response-locked averages were computed for each subject [20]. The brain activity is recorded during the error and correct processing using electroencephalogram with sampling rate equal to 256Hz with 1 second before the response and 2 seconds after the response. For each subject, EEG data over multiple trials (repetition of the same visual stimulus) is recorded (126 trials for ERN, and 1289 trials for CRN). The signals from 62-channels were collected in accordance with the 10/20 system

<sup>2</sup>We thank Dr. Jason Moser from Michigan State University for sharing his EEG dataset with us.

---

**Algorithm 1** Structured Dictionary Learning for SCCI
 

---

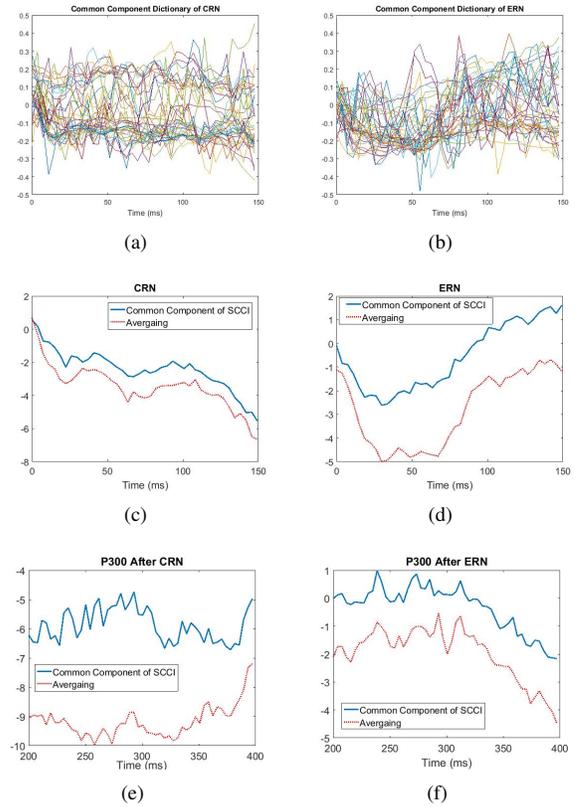
- 1: **Input:** Initialize  $\phi_c^{(0)}$  and  $\phi_j^{(0)}$ . Choose  $\epsilon$ . Set  $k = 0$ .
  - 2: **Output:** The common component dictionary  $\phi_c$ , and the innovation component dictionary  $\phi_j$ .
  - 3: **while** The convergence criteria is not met **do**
  - 4:   Form the basis matrix ( $\Phi^{(k)} \in \mathbb{R}^{NJ \times (J+1)L}$ ) from  $\phi_c^{(k)}$  and  $\phi_j^{(k)}$ . Form the signal vector ( $\mathbf{X} \in \mathbb{R}^{JN \times 1}$ ) in Eq. 1 from the training signals.
  - 5:   Solve  $l1$ -norm minimization in Eq. 6.
  - 6:   **for**  $l = 1:L$  **do**
  - 7:     Compute the error matrix  $\mathbf{E}_{c,l}$  using Eq. 7.
  - 8:     Update  $\mathbf{d}_l^{k+1}$  using the singular vector of  $\mathbf{E}_{c,l}^R$ .
  - 9:   **end for**
  - 10:   **for**  $l = 1:L$  **do**
  - 11:     Compute the error matrix  $\mathbf{E}_{j,l}$  using Eq. 8.
  - 12:     Update  $\mathbf{c}_l^{k+1}$  using the singular vector of  $\mathbf{E}_{j,l}^R$ .
  - 13:   **end for**
  - 14:    $k = k+1$
  - 15: **end while**
- 

on a Neuroscan Synamps2 system (Neuroscan, Inc.).

We assume that the electrodes that are close to each other will follow SCCI model due to volume conduction [6]. Since we know most of the ERN and CRN activities to be localized around frontal-central brain regions [21], we select a time frame of (0, 150)ms corresponding to the ERP from electrodes F2, F4, F3, F5, and FCZ. We randomly select 10 trials from each electrode for CRN and average them to obtain the training dataset, and use another 10 and average them to get the test dataset. Similarly, the trials of each electrode of ERN are divided into two sets of training and test dataset. The number of training and test signals after averaging over trials for 16 subjects and 5 electrodes is 80. The training signals are used to learn the common and innovation dictionaries for ERN and CRN, separately. Fig. 2 (a) and (b) show the learned common component dictionaries for CRN and ERN, respectively. In order to evaluate how similar the learned dictionaries for CRN and ERN are, the distance  $1 - |\mathbf{d}_i^T \hat{\mathbf{d}}_i|$  is computed. 12 atoms out of 41 atoms of the common components dictionary of ERN are close to that of CRN, which justifies the need to learn the dictionaries for the two response types separately.

The learned dictionaries are used to find the common components of the test CRN and ERN data. Fig. 2 (c) and (d) show the common component of CRN and ERN, respectively. The error processing waveform has a negative peak right after the response and becomes positive around 100ms after the response. However, the waveform for the correct response does not show the same negative peak (Fig. 2 (c)). This is expected as the ERN is known to have a larger negative peak immediately after the response as opposed to the CRN. We also compared the recovered ERPs to the ones obtained by averaging (Fig. 2 (c) and (d)). The recovered ERN using averaging remains negative after 100ms (Fig. 2 (d)), and detecting the ERN boundaries becomes difficult.

Similarly, the same electrodes are used to extract P300, a well-known evoked potential to visual stimulus, which starts around 200ms after the response and ends around 400ms after the response. P300 after CRN and ERN are expected to be different. Thus, the common and innovation dictionaries for P300 of CRN and ERN are learned from the set of training signals, separately. The learned dictionaries are used to extract the common components from the test data. Fig. 2 (e) and (f) show P300 extracted from the test data using the learned dictionaries. In comparison to averaging, DL can extract



**Fig. 2.** (a) The dictionary learned to represent the common component of CRN, (b) and ERN; (c) The common components of the CRN, (d) and ERN of the test data recovered using the learned dictionary and averaging; (e) The common components of P300 after CRN and (f) ERN of the test data recovered using the learned dictionary and averaging.

positive peaks in P300 after ERN which is aligned with literature and referred to as P3e [22].

## 5. CONCLUSIONS

In this paper, we proposed to use the sparse common component and innovations model to extract ERPs from EEG recordings of closely located electrodes. Since responses to different stimuli have different characteristics, predefined dictionaries are not suitable to represent and distinguish between them. We introduced a structured K-SVD algorithm to simultaneously learn the common and innovation dictionaries. The proposed structured dictionary learning algorithm is used to separately learn the dictionaries representing the common components of ERN and CRN. The comparison of the learned dictionaries shows the differences between the dictionaries needed to represent ERN and CRN. Finally, the learned dictionaries are used to extract the ERPs corresponding to different responses. The results are consistent with previous findings about the waveforms of ERN and CRN. Future work will consider the extension of the structured K-SVD algorithm to hierarchical implementation of SCCI for time and memory efficiency.

## 6. REFERENCES

- [1] O.I. Khan, F. Farooq, F. Akram, M. Choi, S.M. Han, and T. Kim, "Robust extraction of p300 using constrained ica for bci applications," *Medical & biological engineering & computing*, vol. 50, no. 3, pp. 231–241, 2012.
- [2] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [3] M. Fatourehchi, S.G. Mason, G. E. Birch, and R.K. Ward, "A wavelet-based approach for the extraction of event related potentials from eeg," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on. IEEE*, 2004, vol. 2, pp. ii–737.
- [4] H. Serby, E. Yom-Tov, and G.F. Inbar, "An improved p300-based brain-computer interface," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 13, no. 1, pp. 89–98, 2005.
- [5] A. Ivannikov, T. Karkkainen, T. Ristaniemi, and H. Lyytinen, "Extraction of erp from eeg data," in *2007 9th International Symposium on Signal Processing and Its Applications*, 2007.
- [6] P.L. Nunez, R. Srinivasan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, R.B. Silberstein, and P.J. Cadusch, "EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 5, pp. 499–515, 1997.
- [7] S. Aviyente, E.M. Bernat, W.S. Evans, and S.R. Sponheim, "A phase synchrony measure for quantifying dynamic functional integration in the brain," *Human Brain Mapping*, vol. 32, no. 1, pp. 80–93, 2011.
- [8] D. Baron, M. B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk, "Distributed compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [9] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R.G. Baraniuk, "Distributed compressive sensing," *arXiv preprint arXiv:0901.3403*, 2009.
- [10] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process*, 2007, pp. 181–184.
- [11] M. F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, and R.G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conf. Signals, Sys., Comput*, 2005, pp. 1537–1541.
- [12] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [13] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on. IEEE*, 1999, vol. 5, pp. 2443–2446.
- [14] W. Chen, I.J. Wassell, and M.R.D. Rodrigues, "Dictionary design for distributed compressive sensing," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 95–99, 2015.
- [15] Q. Barthélemy, C. Gouy-Pailler, Y. Isaac, A. Souloumiac, A. Larue, and J. Mars, "Multivariate temporal dictionary learning for eeg," *Journal of neuroscience methods*, vol. 215, no. 1, pp. 19–28, 2013.
- [16] J.R. Hall, E.M. Bernat, and C.J. Patrick, "Externalizing psychopathology and the error-related negativity," *Psychological Science*, vol. 18, no. 4, pp. 326–333, 2007.
- [17] D. Baron, M.F. Duarte, S. Sarvotham, M. B. Wakin, and R.G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. 45rd Conference on Communication, Control, and Computing*, 2005.
- [18] M.S. Lewicki and T.J. Sejnowski, "Learning overcomplete representations," *Neural computation*, vol. 12, no. 2, pp. 337–365, 2000.
- [19] J. R. Hall, E. M. Bernat, and C. J. Patrick, "Externalizing psychopathology and the error-related negativity," *Psychological Science*, vol. 18, no. 4, pp. 326–333, 2007.
- [20] B.A. Eriksen and C.W. Eriksen, "Effects of noise letters upon the identification of a target letter in a nonsearch task," *Perception & psychophysics*, vol. 16, no. 1, pp. 143–149, 1974.
- [21] A.G. Mahyari, D. Zoltowski, E. Bernat, and S. Aviyente, "A tensor decomposition based approach for detecting dynamic network states from EEG," 2016.
- [22] D.S. Ruchkin, S. Sutton, and D. Mahaffey, "Functional differences between members of the p300 complex: P3e and p3b," *Psychophysiology*, vol. 24, no. 1, pp. 87–103, 1987.