Spatial Filtering for EEG-Based Regression Problems in Brain–Computer Interface (BCI)

Dongrui Wu , *Senior Member, IEEE*, Jung-Tai King, Chun-Hsiang Chuang, Chin-Teng Lin , *Fellow, IEEE*, and Tzyy-Ping Jung , *Fellow, IEEE*

Abstract—Electroencephalogram (EEG) signals are frequently used in brain–computer interfaces (BCIs), but they are easily contaminated by artifacts and noise, so preprocessing must be done before they are fed into a machine learning algorithm for classification or regression. Spatial filters have been widely used to increase the signal-to-noise ratio of EEG for BCI classification problems, but their applications in BCI regression problems have been very limited. This paper proposes two common spatial pattern (CSP) filters for EEG-based regression problems in BCI, which are extended from the CSP filter for classification, by using fuzzy sets. Experimental results on EEG-based response speed estimation from a large-scale study, which collected 143 sessions of sustained-attention psychomotor vigilance task data from 17 subjects during a 5-month period, demonstrate that the two proposed spatial filters can significantly increase the EEG signal quality. When used in LASSO and k-nearest neighbors regression for user response speed estimation, the spatial filters can reduce the root-mean-square estimation error by $10.02-19.77\%$, and at the same time increase the correlation to the true response speed by $19.39-86.47\%$.

Index Terms—Brain–computer interface (BCI), common spatial pattern (CSP), electroencephalogram (EEG), fuzzy sets, psychomotor vigilance task (PVT), response speed (RS) estimation, spatial filtering.

I. INTRODUCTION

ELECTROENCEPHALOGRAM (EEG) is the most widely used signal for brain–computer interfaces (BCIs) [24], [25], [29], [34], [47], [53], mainly due to the convenience to obtain them, compared with magnetoencephalography

D. Wu is with the DataNova, Clifton Park, NY 12065 USA (e-mail: drwu09@gmail.com).

J.-T. King is with the Brain Research Center, National Chiao-Tung University, Hsinchu 300, Taiwan (e-mail: jtchin2@gmail.com).

C.-H. Chuang and C.-T. Ling are with the Centre of Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: cch.chuang@gmail.com; Chin-Teng.Lin@uts.edu.au).

T.-P. Jung is with the Swartz Center for Computational Neuroscience, Institute for Neural Computation and the Center for Advanced Neurological Engineering, Institute of Engineering in Medicine, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: jung@sccn.ucsd.edu).

[32], functional magnetic resonance imaging [44], functional near-infrared spectroscopy [33], and invasive recordings such as electrocorticography [30], [35]. However, EEG signals are often contaminated by ocular, muscular, and cardiac artifacts and other noise (power line, changes in electrode impedances, etc.) [4], [34], [49]. Usually some preprocessing, either manually or automatically [4], [34], is needed to remove the artifacts, and then temporal and spatial filters are applied to further improve the EEG signal quality before feeding the EEG data into a classification or regression algorithm. The most commonly used temporal filters are bandpass filters and notch filters (at 50 or 60 Hz power-line frequency).

This study focuses on spatial filtering for improving the EEG signal quality. Many such approaches have been proposed in the literature [2], [7], [15], [17], [37], [38], [40], [41], [54]. However, almost all of them focus primarily on EEG classification problems in BCI, whereas EEG regression problems have been largely overlooked. Nevertheless, the latter is also very important in BCIs. One example is driver drowsiness (or alertness) estimation from EEG signals, which has been extensively studied in our previous research [26]–[28], [56], [59]–[61]. This is a very important problem because drowsy driving is among the most important causes of road crashes, following only to alcohol, speeding, and inattention [43]. According to the National Highway Traffic Safety Administration [52], 2.5% of fatal motor vehicle crashes (on average 886/year in U.S.) and 2.5% of fatalities (on average 1004/year in U.S.) between 2005 and 2009 involved drowsy driving.

This study proposes two spatial filters for EEG-based regression problems in BCI. We also validate their performance of response speed (RS) estimation from EEG signals measured in a large-scale sustained-attention psychomotor vigilance task (PVT) [21], which collected 143 sessions of data from 17 subjects in a 5-month period.

The remainder of this paper is organized as follows. Section II reviews the state-of-the-art spatial filters for EEG-based classification problems in BCI. Section III introduces our proposed spatial filters for supervised BCI regression problems. Section IV describes the experimental setup, RS, and EEG data preprocessing techniques, and the procedure to evaluate the performances of different spatial filters. Section V presents the results of the comparative studies and parameter sensitivity analysis for the proposed spatial filter. Section VI discusses the limitations of the proposed approaches and outlines several future research directions. Finally, Section VII draws conclusions.

II. Spatial Filters for EEG Classification in BCI

Many spatial filters have been proposed for the EEG classification in BCI. The most basic ones include common average reference [48], Laplacian filters [23], and principal component analysis [19]. Some of the more recent and also more sophisticated ones are as follows.

1) *Independent component analysis (ICA)* [9], [17], [54] decomposes a multivariate signal into independent non-Gaussian signals. ICA has been widely used in the EEG research community to detect and remove stereotyped eye, muscle, and line noise artifacts [20], [26], [49].

   Generally ICA works on an unepoched long block of EEG data, instead of epoched short EEG trials. Let the unepoched EEG data be $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C is the number of EEG channels, and T is the number of time samples. ICA assumes that $\mathbf{X}$ is the linear combination of c independent sources, i.e., $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A} \in \mathbb{R}^{C \times c}$ is the *mixing matrix*, and the source signals, which are the rows of $\mathbf{S} \in \mathbb{R}^{c \times T}$, are supposed to be stationary, independent, and non-Gaussian. ICA can use various different principles [9], [17], [49], [54] to estimate both unknown $\mathbf{A}$ and unknown $\mathbf{S}$ simultaneously from $\mathbf{X}$. Once $\mathbf{S}$ is obtained, cleaner and more representative features may be extracted from it than from the original $\mathbf{X}$ [26].

2) *xDAWN algorithm* [38]–[40] is often used to increase the signal to signal-plus-noise ratio (SNR) in P300-based BCIs.

   Like ICA, xDAWN also works on the unepoched long block of EEG data $\mathbf{X} \in \mathbb{R}^{C \times T}$. It assumes that $\mathbf{X} = \mathbf{PD}^T + \mathbf{N}$, where $\mathbf{P} \in \mathbb{R}^{C \times S}$ represents the P300 signal in an EEG epoch, and $\mathbf{D} \in \mathbb{R}^{T \times S}$ is a Toeplitz matrix whose first column is defined as

$$\mathbf{D}_{\tau_k,1} = \begin{cases} 1, & \tau_k \text{ is the onset of the} \\ & k\text{th target stimulus} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

and $\mathbf{N} \in \mathbb{R}^{C \times T}$ represents the ongoing background brain activity as well as the artifacts and noise. xDAWN then designs a spatial filtering matrix $\mathbf{W}^* \in \mathbb{R}^{C \times F}$, where F is the number of spatial filters, to maximize the signal to SNR, i.e.,

$$\mathbf{W}^* = \arg \max_{\mathbf{W} \in \mathbb{R}^{C \times F}} \frac{\mathrm{Tr}(\mathbf{W}^T \mathbf{PD}^T \mathbf{DP}^T \mathbf{W})}{\mathrm{Tr}(\mathbf{W}^T \mathbf{XX}^T \mathbf{W})} \tag{2}$$

where $\mathrm{Tr}(\cdot)$ is the trace of a matrix. Equation (2) is a generalized Rayleigh quotient [14], and its solution $\mathbf{W}^*$ is the concatenation of the F eigenvectors associated with the F largest eigenvalues of the matrix $(\mathbf{XX}^T)^{-1} \mathbf{PD}^T \mathbf{DP}^T$. The spatially filtered trial for $\mathbf{X}_n$ is then computed as

$$\mathbf{X}_n' = \mathbf{W}^{*T} \mathbf{X}_n, \quad n = 1, \dots, N. \tag{3}$$

3) *Canonical correlation analysis (CCA)* [15], [41] finds linear transformations to maximize the correlations between two datasets. It has been used to improve BCI performance in code-modulated visual evoked potentials [5], steady-state visual evoked potentials [6], and event-related potentials like P300 and error-related potentials [45].

   Unlike ICA and xDAWN, CCA works on epoched EEG trials. Consider a binary classification problem with N_1 training examples in Class 1 and N_2 training examples in Class 2. Let $(\mathbf{X}_n, y_n)$ be the nth training example, where $\mathbf{X}_n \in \mathbb{R}^{C \times S}$ (C is the number of channels, and S is the number of time samples in each trial), and $y_n \in \{1, 2\}$. Let $\bar{\mathbf{X}}_k \in \mathbb{R}^{C \times S}$ be the average of $\mathbf{X}_n$ in Class k ($k = 1, 2$). We then construct $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1\ \tilde{\mathbf{X}}_2]$ and $\tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}_1\ \tilde{\mathbf{Z}}_2]$, where $\tilde{\mathbf{X}}_k$ is the concatenation of all N_k $\mathbf{X}_n$ in Class k, and $\tilde{\mathbf{Z}}_k$ is the concatenation of N_k $\bar{\mathbf{X}}_k$. CCA first finds two vector filters $\mathbf{w}_{\tilde{\mathbf{X}}}$ and $\mathbf{w}_{\tilde{\mathbf{Z}}}$ such that the correlation between $\mathbf{w}_{\tilde{\mathbf{X}}}^T \tilde{\mathbf{X}}$ and $\mathbf{w}_{\tilde{\mathbf{Z}}}^T \tilde{\mathbf{Z}}$ is maximized. $\mathbf{w}_{\tilde{\mathbf{X}}}^T \tilde{\mathbf{X}}$ and $\mathbf{w}_{\tilde{\mathbf{Z}}}^T \tilde{\mathbf{Z}}$ are called the first pair of canonical variables. CCA then finds the second pair of canonical variables in a similar way, subject to the constraint that they are uncorrelated with the first pair of canonical variables. This procedure can be continued up to C times.

   Finally, the spatial filtering matrix is the concatenation of all $\mathbf{w}_{\tilde{\mathbf{X}}}$, which can be applied to each $\mathbf{X}_n$ to increase its SNR.

4) *Common spatial patterns (CSP)* [7], [37] is a supervised technique frequently used to enhance the binary classification performance of EEG data. The basic idea is to separate the EEG signal into additive subcomponents that have maximum differences in variance between the two classes. In the following, we introduce the one-versus-the-rest (OVR) CSP [11], which extends the traditional CSP from binary classification to K classes.

   Like CCA, OVR CSP also works on epoched EEG trials. Let $(\mathbf{X}_n, y_n)$ be the nth training example, as defined above. Assume the mean of $\mathbf{X}_n$ has been removed, e.g., by high-pass or bandpass filtering. Then, for Class k, OVR CSP finds a spatial filter matrix $\mathbf{W}_k^* \in \mathbb{R}^{C \times F}$, where F is the number of spatial filters, to maximize the variance difference between Class k and the rest:

$$\mathbf{W}_k^* = \arg \max_{\mathbf{W} \in \mathbb{R}^{C \times F}} \frac{\mathrm{Tr}(\mathbf{W}^T \bar{\boldsymbol{\Sigma}}_k \mathbf{W})}{\mathrm{Tr}[\mathbf{W}^T (\sum_{i \neq k} \bar{\boldsymbol{\Sigma}}_i) \mathbf{W}]} \tag{4}$$

where $\bar{\boldsymbol{\Sigma}}_k$ is the mean covariance matrix of trials in Class k. Equation (4) is also a generalized Rayleigh quotient [14], and the solution $\mathbf{W}_k^*$ is the concatenation of the F eigenvectors associated with the F largest eigenvalues of the matrix $(\sum_{i \neq k} \bar{\boldsymbol{\Sigma}}_i)^{-1} \bar{\boldsymbol{\Sigma}}_k$.

   Finally, we concatenate the K individual OVR CSP spatial filters to obtain the complete filter:

$$\mathbf{W}^* = [\mathbf{W}_1^*, \dots \mathbf{W}_K^*] \in \mathbb{R}^{C \times KF} \tag{5}$$

and compute the spatially filtered trial for $\mathbf{X}_n$ by (3).

III. Spatial Filters for Supervised BCI Regression Problems

In this section, we propose two CSP for regression (CSPR) filters, which extend the multiclass CSP filters from classification to regression by using fuzzy sets [63], as we have done in [60].
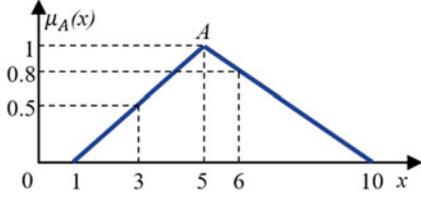
Fig. 1.   Examples of a fuzzy set.
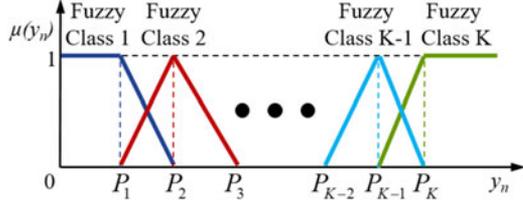
Fig. 2. K fuzzy classes for y_n, when triangular fuzzy sets are used.

First, a brief introduction of fuzzy sets is given below.

A. Fuzzy Sets

A fuzzy set A is comprised of a *universe of discourse D_A* of real numbers together with a *membership function $\mu_A : D_A \to [0, 1]$*, i.e.,

$$A = \int_{D_A} \mu_A(x)/x. \tag{6}$$

Here, $\int$ denotes the collection of all points $x \in D_A$ with associated *membership degree $\mu_A(x)$*. An example of a fuzzy set is shown in Fig. 1. The membership degrees are $\mu_A(1) = 0$, $\mu_A(3) = 0.5$, $\mu_A(5) = 1$, $\mu_A(6) = 0.8$, and $\mu_A(10) = 0$. Observe that this is different from traditional (binary) sets, where each element can only belong to a set completely (i.e., with membership degree 1), or does not belong to it at all (i.e., with membership degree 0); there is nothing in between (i.e., with membership degree 0.5). Fuzzy sets are frequently used in modeling concepts in natural language [22], [36], [55], which may not have clear boundaries.

B. Common Spatial Patterns for Regression One-Versus-the-Rest

Let $\mathbf{X}_n \in \mathbb{R}^{C \times S}$ ($n = 1, \ldots N$) be the nth EEG trial, where C is the number of channels and S is the number of time samples in each trial. We assume that the mean of each channel measurement has been removed, which is usually performed by bandpass filtering. Let $y_n \in \{1, \ldots, K\}$ be the RS of $\mathbf{X}_n$.

With the help of fuzzy sets, we can define "fuzzy" classes to connect regression problems and classification problems. Assume K fuzzy classes are used. First, we partition the interval $[0, 100]$ into $K + 1$ equal intervals, and denote the partition points as $\{p_k\}_{k=1,\ldots,K}$. It is easy to obtain that

$$p_k = \frac{100 \cdot k}{K + 1}, \qquad k = 1, \ldots, K. \tag{7}$$

Input: EEG training examples $(\mathbf{X}_n, y_n)$, where
    $\mathbf{X}_n \in \mathbb{R}^{C \times S}$, $n = 1, \ldots, N$;
    K, the number of fuzzy classes for y_n;
    F, the number of spatial filters for each
      fuzzy class.
Output: Spatially filtered EEG trials $\mathbf{X}'_n \in \mathbb{R}^{KF \times S}$.
Band-pass filter each $\mathbf{X}_n$ to remove the mean of each
  channel;
Compute $\{p_k\}_{k=1,\ldots,K}$ in (7);
Compute the corresponding percentile values
  $\{P_k\}_{k=1,\ldots,K}$ for y_n;
Construct the K fuzzy classes as shown in Fig. 2;
Compute $\bar{\mathbf{\Sigma}}_k$ by (8);
Compute $\mathbf{W}^*_k$ by (4);
Construct $\mathbf{W}^*$ by (5);
Return $\mathbf{X}'_n$ by (3)

For each p_k, we then find the corresponding p_k percentile value of all training y_n and denote it as P_k. Next, we define K fuzzy classes from them, as shown in Fig. 2. In this way, we can "classify" the training y_n into K fuzzy classes, corresponding to the K crisp classes in the CSP for classification. However, note that in the CSP for classification, a y_n belongs to a crisp class either completely or not at all. For a fuzzy class here, a y_n can belong to it at a membership degree in $[0, 1]$.

Next, for each fuzzy class, we compute its mean spatial covariance matrix as

$$\bar{\mathbf{\Sigma}}_k = \frac{\sum_{n=1}^{N} \mu_k(y_n) \mathbf{X}_n \mathbf{X}_n^T}{\sum_{n=1}^{N} \mu_k(y_n)}, \qquad k = 1, \ldots, K \tag{8}$$

where $\mu_k(y_n)$ is the membership degree of y_n in fuzzy class k. Substituting (8) into (4), we can solve for the spatial filtering matrix $\mathbf{W}^*_k$ for fuzzy class k. Essentially, this $\mathbf{W}^*_k$ makes those $\mathbf{X}_n$ in fuzzy class k different from those not in fuzzy class k, which will help the regression performance, as we will demonstrate in Section V.

Next, we construct a concatenated spatial filtering matrix $\mathbf{W}^*$ by (5), and finally perform the spatial filtering for each EEG trial $\mathbf{X}_n$ by (3). The complete CSPR-OVR spatial filter for supervised BCI regression problems is summarized in Algorithm 1.

C. CSPR One-Versus-All (OVA)

In (4), we construct the multiclass CSP using an OVR approach, but it can also be constructed using the following OVA approach:

$$\mathbf{W}^*_k = \arg \max_{\mathbf{W} \in \mathbb{R}^{C \times F}} \frac{\mathrm{Tr}(\mathbf{W}^T \bar{\mathbf{\Sigma}}_k \mathbf{W})}{\mathrm{Tr}[\mathbf{W}^T (\sum_{i=1}^{K} \bar{\mathbf{\Sigma}}_i) \mathbf{W}]}. \tag{9}$$

The only difference between (9) and (4) is that the numerator of (9) also includes the contribution from Class k itself. If we view Class k as the signal of interest, and all other classes as noise,

then (9) maximizes the signal to SNR, as (2) in the xDAWN algorithm.

Equation (9) is also a generalized Rayleigh quotient [14], and the solution $\mathbf{W}_k^*$ is the concatenation of the F eigenvectors associated with the F largest eigenvalues of the matrix $(\sum_{i=1}^{K} \bar{\mathbf{\Sigma}}_i)^{-1} \bar{\mathbf{\Sigma}}_k$. The OVA CSP for classification still uses (5) to construct the final spatial filter, and (3) to perform the filtering.

Using the technique introduced in the previous section, we can easily develop the CSPR-OVA spatial filter for BCI regression problems. Its procedure is almost identical to that in Algorithm 1. The only difference is that $\mathbf{W}_k^*$ is computed by (9) instead of (4).

IV. EXPERIMENTS AND DATA

This section introduces a PVT experiment that was used to evaluate the performances of the proposed spatial filtering algorithms, the corresponding RS and EEG data preprocessing procedures, and the feature sets.

A. Experiment Setup

A total of 17 university students (13 males; average age 22.4, standard deviation 1.6) from the National Chiao Tung University (NCTU) in Taiwan volunteered to support the data-collection efforts over a five-month period to study EEG correlates of attention and performance changes under specific conditions of real-world fatigue [21], as determined by the effectiveness score of Readiband [42]. The voluntary, fully informed consent of the participants of this research was obtained as required by federal and army regulations [50], [51]. The Institutional Review Board of NCTU approved the experimental protocol.

All participants registered their fatigue levels through a smartphone daily, and received notifications to report for laboratory experiments when the effectiveness score deemed their conditions fitted the experimental requirement (low fatigue: > 90; normal: $[70, 90]$; and high fatigue: < 70). Upon completion of the related questionnaires [Karolinska Sleepiness Scale (KSS) [1], and electronically-adapted visual analog scale for fatigue (VAS-F) and stress (VAS-S)] and the informed consent form, subjects performed a PVT, a dynamic attention-shifting task, a lane-keeping task, and selected surveys (KSS, VAS-F, VAS-S, state-trait anxiety inventory, and mind wandering) before each task. EEG data were recorded at 1000 Hz using a 64-channel NeuroScan system. Most participants performed the laboratory experiment thrice in each of the three fatigue states.

This study focuses on the PVT [10], which is a sustained-attention task that uses RS to measure the speed with which a subject responds to a visual stimulus. It is widely used, particularly by NASA, for its ease of scoring, simple metrics, convergent validity, and free of learning effects. In our experiment, the PVT was presented on a smartphone with each trial initiated as an empty solid white circle centered on the touchscreen that began to fill in red displayed as a clockwise sweeping motion like the hand of a clock. The sweeping motion was programmed to turn solid red in 1 s or terminate upon a response by the participants, which required them to tap the touchscreen with the thumb of their dominant hand. The RS was computed as the inverse of the elapsed time between the appearance of the empty solid white circle and the participant's response. Following completion of each trial, the circle went back to solid white until the next trial. Intertrial intervals consisted of random intervals between 2 and 10 s.

A total of 143 sessions of PVT data were collected from the 17 subjects, and each session lasted 10 min. Our goal is to predict the RS using a three-s EEG trial immediately before it.

B. Performance Evaluation Process

The following procedure was performed to evaluate the performances of different spatial filters:
1) EEG data preprocessing to suppress artifacts and noise;
2) RS data preprocessing to suppress outliers;
3) fivefold crossvalidation to compute the regression performance for each combination of spatial filters and regression method: First, randomly partition the trials into five equal folds; then, use four folds for supervised spatial filtering and regression model training, and the remaining fold for testing; repeat this five times so that every fold is used in testing; finally, compute the regression performances in terms of root-mean-square error (RMSE) and correlation coefficient (CC). Two regression methods were used: LASSO, whose adjustable parameter λ was optimized by an inner fivefold crossvalidation on the training dataset, and k-nearest neighbors (kNN) regression, where $k = 5$.
4) Repeat step 3) ten times and compute the average regression performance.

More details about the first two steps are given in the next two sections.

C. EEG Data Preprocessing

We first downsampled the EEG data to 256 Hz, then epoched them to 3-s trials according to the onset of the PVTs. Let the onset time of the nth PVT be t_n. Then, the 62-channel EEG trial in $[t_n - 3, t_n]$ s was used to predict the RS, i.e., $\mathbf{X}_n \in \mathbb{R}^{62 \times 768}$. Each trial was then individually filtered by a $[1, 20]$ Hz finite-impulse response bandpass filter to make each channel zero-mean and to remove unuseful high-frequency components.

Because the intertrial intervals consisted of random intervals between 2 and 10 s, it is possible that a 3-s EEG trial covers part of data from the previous trial. Additionally, a trial may also contain the EEG oscillations related to motor reaction (tapping the touchscreen) in the previous trial. To remedy these problems, we removed overlapping trials: Let the RS of the nth trial be y_n (the corresponding response time is $1/y_n$); then, the nth trial is removed if $t_n - t_{n-1} < 1/y_{n-1} + 3$, i.e., when the 3-s EEG data for Trial n overlap with the data and response for the previous trial.

D. RS Data Preprocessing

The raw response times for two subjects are shown in Fig. 3. The top panel is from a typical subject, whose response times were mostly shorter than 1 s. The lower panel is from a subject
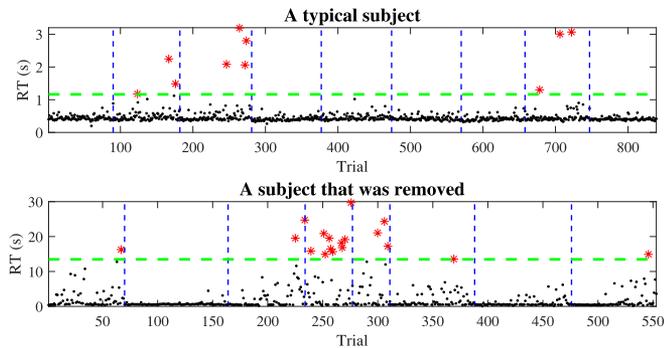
Fig. 3. Response times for a typical subject (top panel) and a subject with possible data recording issues (bottom panel). The green line is the threshold, and the red stars represent response times above the threshold, which will be brought to the threshold.

Fig. 4. Distributions of the preprocessed RSs for the 16 subjects.

with possible data recording issues, because lots of response times were longer than 5 s, which are highly unlikely in practice. So, we excluded that subject from the consideration in this paper, and only used the remaining 16 subjects.

As shown in Fig. 3, the response times were very noisy, and there were obvious outliers. It is very important to suppress the outliers and noise so that the performances of different algorithms can be more accurately compared. In addition to the step in the previous section to remove overlapping trials, we also employed the following two-step procedure for response time preprocessing.

1) *Outlier thresholding* aimed to suppress abnormally large response times. First, a threshold $\theta = m_y + 3\sigma_y$ was computed for each subject, where m_y is the mean response time from all sessions of that subject, and σ_y is the corresponding standard deviation. Then, all response times larger than θ were replaced by θ. Note that the threshold was different for different subjects.

2) *Moving average smoothing* replaced each response time by the average response time during a 60 s moving window centered at the onset of the corresponding PVT to suppress the noise.

We then computed the RS as the inverse of the RT. The RSs for the 16 subjects are shown in Fig. 4. Observe that they are roughly in the same range, and many of them are approximately Gaussian.

E. Feature Extraction

We extracted the following four feature sets for each preprocessed EEG trial:

1) `Raw`: *Theta and Alpha powerband features from the bandpass filtered EEG trials.* We computed the average power spectral density in the Theta band (4–8 Hz) and Alpha band (8–13 Hz) for each channel using Welch's method [57], and converted these $62 \times 2 = 124$ band powers to decibel as our features.

2) `CAR`: *Theta and Alpha powerband features from EEG trials filtered by CAR.* This procedure was almost identical to `Raw`, except that the bandpass filtered EEG trials were
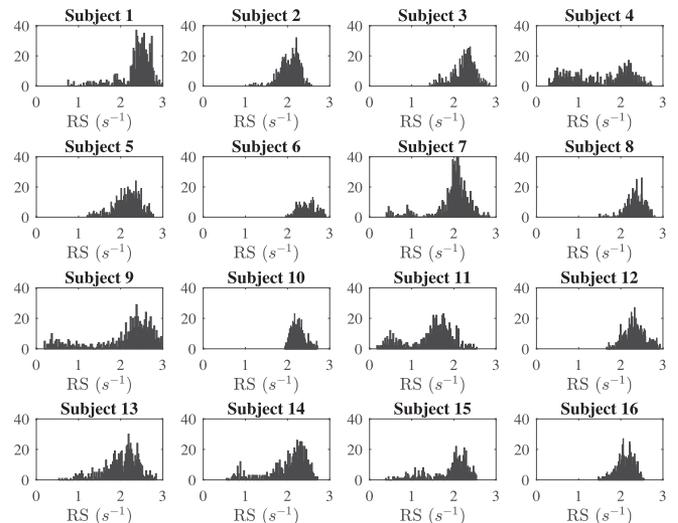
also spatially filtered by CAR before the $62 \times 2 = 124$ powerband features were computed. CAR is one of the most commonly used spatial filters for EEG, and in [31], it has been showed that it helped improve EEG classification performance. It simply removes the mean of all channels from each channel.

3) `OVR`: *Theta and Alpha powerband features from EEG trials filtered by CSPR-OVR.* This procedure was almost identical to `CAR`, except that the CAR filter was replaced by CSPR-OVR. We used three fuzzy classes for the RSs, and 21 spatial filters[1] for each fuzzy class, so that the spatially filtered signals had dimensionality of 63×1280, roughly the same as the dimensionality of the original signals. We then extracted $63 \times 2 = 126$ band power features for each trial.

4) `OVA`: *Theta and Alpha powerband features from EEG trials filtered by CSPR-OVA.* This procedure was also almost identical to `CAR`, except that the spatial filtering was performed by CSPR-OVA instead of CAR. There were also $63 \times 2 = 126$ band power features for each trial.

V. EXPERIMENTAL RESULTS

This section compares the informativeness of the features in `Raw`, `CAR`, `OVR`, and `OVA`, presents the regression performances, and also performs parameter sensitivity analysis for Algorithm 1.

A. Informativeness of the Features

Before studying the regression performances, it is important to check if the extracted features in `Raw`, `CAR`, `OVR` and `OVA` are indeed meaningful. We picked a typical subject, partitioned his

---

[1]We used 21 spatial filters here so that the filtered signals had roughly the same dimensionality as the original signals, which ensured a fair performance comparison. In Section V-C, we also performed sensitivity analysis on the number of spatial filters.
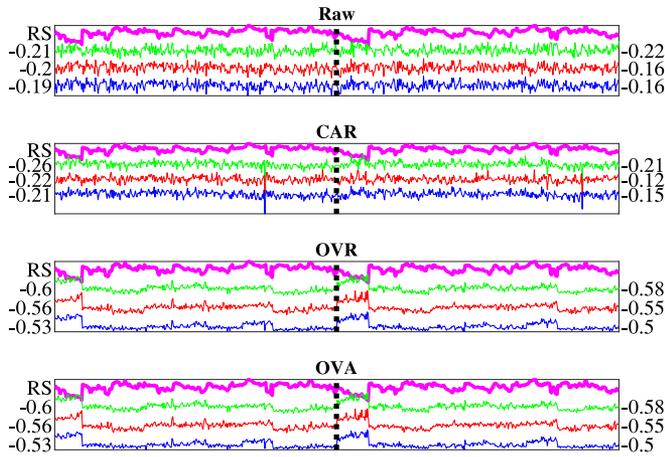
Fig. 5. Powerband features from different feature extraction methods, and the corresponding training and testing CCs with the RS.

TABLE I
p-VALUES OF TWO-WAY ANOVA TESTS FOR {Raw, CAR, OVR, OVA}

	LASSO		kNN	
	RMSE	CC	RMSE	CC
p	0.0061	0.0000	0.0000	0.0000

TABLE II
p-VALUES OF NONPARAMETRIC MULTIPLE COMPARISON FOR {Raw, CAR, OVR, OVA}

	LASSO						kNN					
	RMSE			CC			RMSE			CC		
	Raw	CAR	OVR	Raw	CAR	OVR	Raw	CAR	OVR	Raw	CAR	OVR
CAR	.5883			.3374			.1437			**.0009**		
OVR	**.0063**	**.0034**		**.0000**	**.0000**		**.0000**	**.0001**		**.0000**	**.0000**	
OVA	**.0122**	**.0044**	.4960	**.0000**	**.0000**	.4970	**.0000**	**.0001**	.4937	**.0000**	**.0000**	.4741

data random into 50% training and 50% testing, and extracted Raw and CAR. We then designed the spatial filters using CSPR-OVR and CSPR-OVA on the training data, and extracted the corresponding OVR and OVA. For each feature set, we identified the top three channels that had the maximum correlation with the RS using the training data, and also computed the corresponding CCs for the testing data.

The results are shown in Fig. 5, where in each subfigure the data on the left of the black dotted line were used for training, and the right for testing. The top thick curve is the RS, and the bottom three curves are the maximally correlated features (note that good features are negatively correlated with the RS) identified from the training data. The training and testing CCs are shown on the left and right of the corresponding channel, respectively. Observe that the features from CAR had slightly better correlations with the RS in training than those from Raw, but not necessarily in testing. However, the features from OVR and OVA had much higher training and testing correlations to the RS than those from Raw and CAR, suggesting that CSPR-OVR and CSPR-OVA can indeed increase the signal quality. The reason is if we view Class k as the signal of interest, and all other classes as noise, then CSPR-OVR in (4) enhances the SNR of the EEG signal, and CSPR-OVA in (9) enhances the signal to SNR.

B. Regression Performance Comparison

The RMSEs and CCs of LASSO and kNN using the four feature sets are shown in Fig. 6 for the 16 subjects. Recall that for each subject, the feature extraction methods were run ten times, each with randomly partitioned training and testing data, and the average regression performances are shown here. The average RMSEs and CCs across all subjects are also shown in the last group of each panel. Observe that CAR had comparable or slightly better performance than Raw. Regardless of which regression algorithm was used, generally OVR and OVA had similar performance, and both of them achieved much smaller RMSEs and much larger CCs than Raw and CAR, suggesting

that our extension of CSP from a supervised classification to a supervised regression can indeed improve the regression performance. Finally, LASSO had better performance than kNN on Raw and CAR, but kNN became better on OVR and OVA.

The corresponding percentage performance improvements of LASSO and kNN using the four feature sets are shown in Fig. 7, where the legend "LASSO,OVR/Raw" means the percentage performance improvement of LASSO on OVR over LASSO on Raw, and other legends should be interpreted in a similar manner. For both LASSO and kNN, OVR and OVA achieved similar performance improvements over Raw, and also over CAR. For LASSO, on average OVR had 10.02% smaller RMSE than Raw, and 19.39% larger CC. For kNN, on average OVR had 19.77% smaller RMSE than Raw, and 86.47% larger CC.

We also performed a two-way analysis of variance (ANOVA) for different regression algorithms to check if the RMSE and CC differences among the four feature sets were statistically significant, by setting the subjects as a random effect. The results are shown in Table I, which indicated that there were statistically significant differences in both RMSEs and CCs among different feature sets for both LASSO and kNN.

Then, nonparametric multiple comparison tests based on Dunn's procedure [12], [13] were used to determine if the difference between any pair of algorithms was statistically significant, with a p-value correction using the false discovery rate method [3]. The p-values are shown in Table II, where the statistically significant ones are marked in bold. Table II shows that, except for the CC of kNN, generally there was no statistically significant difference between Raw and CAR. However, for both LASSO and kNN, the RMSE and CC differences between {OVR, OVA} and {Raw, CAR} were always statistically significant. In all cases, there were no statistically significant differences between OVR and OVA.

C. Parameter Sensitivity Analysis

There are two adjustable parameters in CSPR-OVR: K, the number of fuzzy classes for the RSs, and F, the number of spatial filters for each fuzzy class. In this section, we
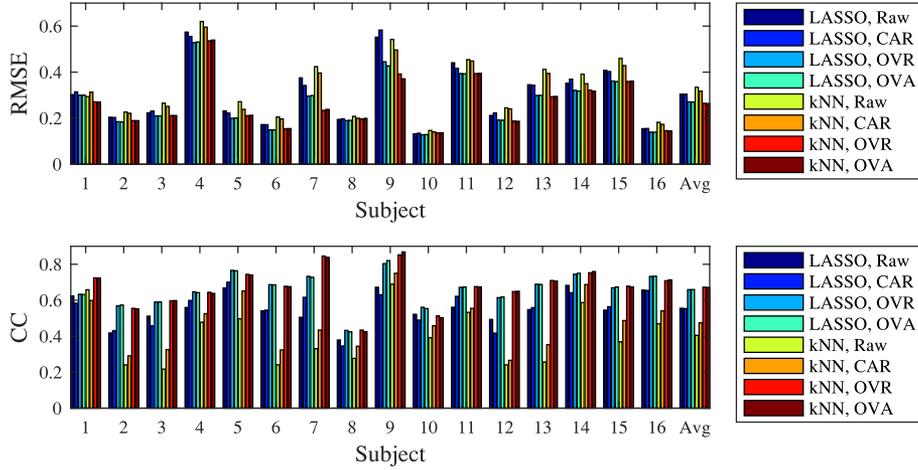
Fig. 6.   RMSEs and CCs of the 8 approaches on the 16 subjects.
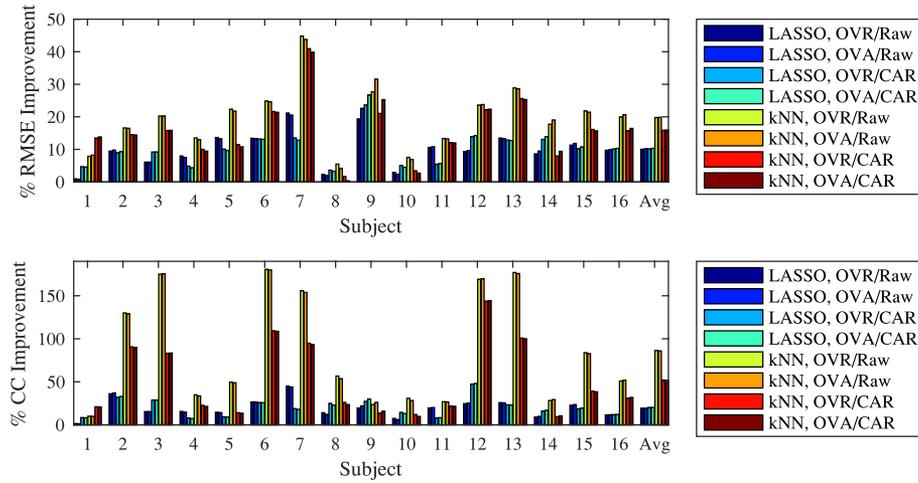
Fig. 7.   Pairwise percentage performance improvement of the algorithms on the 16 subjects.

study the sensitivity of the regression performance to these two parameters.

The regression performances for $K = \{2, 3, 4, 5, 6, 7\}$ (F was fixed to be 21) are shown in Fig. 8. Algorithm 1 was repeated five times, each with a random partition of training and testing data, and the average regression results are shown. For both LASSO and kNN, on average $K = 2$ gave the worst performance, but $K = \{3, 4, 5, 6, 7\}$ resulted in roughly the same RMSE and CC. Hence, $K = 3$ seems to be a good compromise between performance and computational cost.

The regression performances for $F = \{5, 10, 20, 30, 40, 50, 60\}$ (K was fixed to be 3) are shown in Fig. 9. Algorithm 1 was again repeated five times, and the average regression results are shown. For both LASSO and kNN, generally a larger F resulted in a smaller RMSE and a larger CC, but the performance may reach a plateau at a certain F. Also, a larger F means heavier computational cost, which should be taken into consideration in choosing F. For the PVT experiment, $F \in [20, 30]$ seemed to achieve a good compromise between performance and computational cost.

D. Different Fuzzy Set Shapes

In Section III, we used triangular fuzzy sets for simplicity, but other shapes can also be used. Fig. 10 illustrates how Gaussian fuzzy sets can be designed here: The center of the kth Gaussian fuzzy class is at P_k [computed from (7)], and the spread is specially designed so that two adjacent fuzzy sets intersect at the midpoint with membership grade 0.5. As a result, generally the Gaussian fuzzy classes are not symmetric.

When the Gaussian fuzzy classes in Fig. 10 are used in CSPR-OVR and CSPR-OVA, the results are shown in Fig. 11, which are almost identical to those obtained from triangular fuzzy sets (see Fig. 6).

E. Robustness to Noise

It is also important to study the robustness of different spatial filters to the noise. According to Zhu and Wu [64], there are two types of noise: *class noise*, which is the noise on the model outputs, and *attribute noise*, which is the noise on the model inputs. In this section, we focus on the attribute noise.
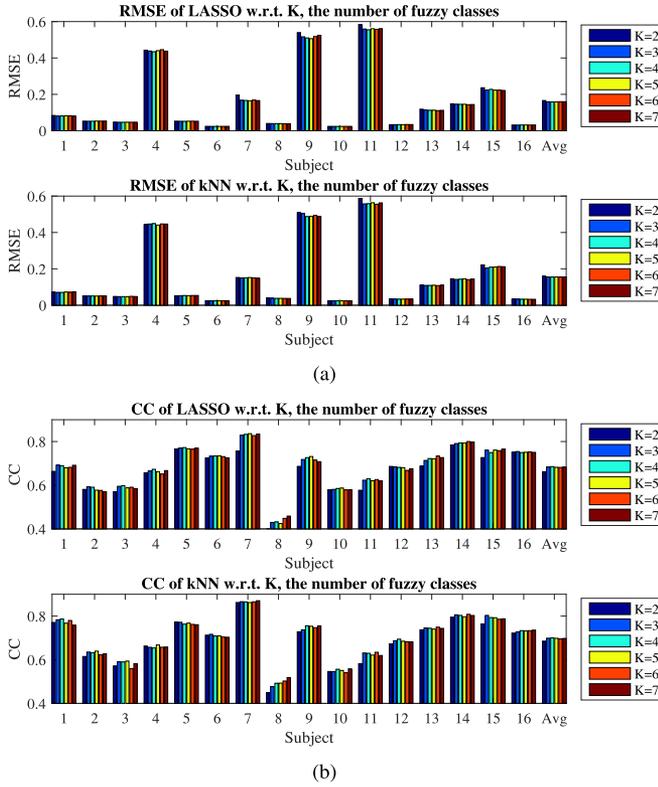
Fig. 8. (a) RMSEs and (b) CCs of LASSO and kNN with respect to K, the number of fuzzy classes in Algorithm 1.
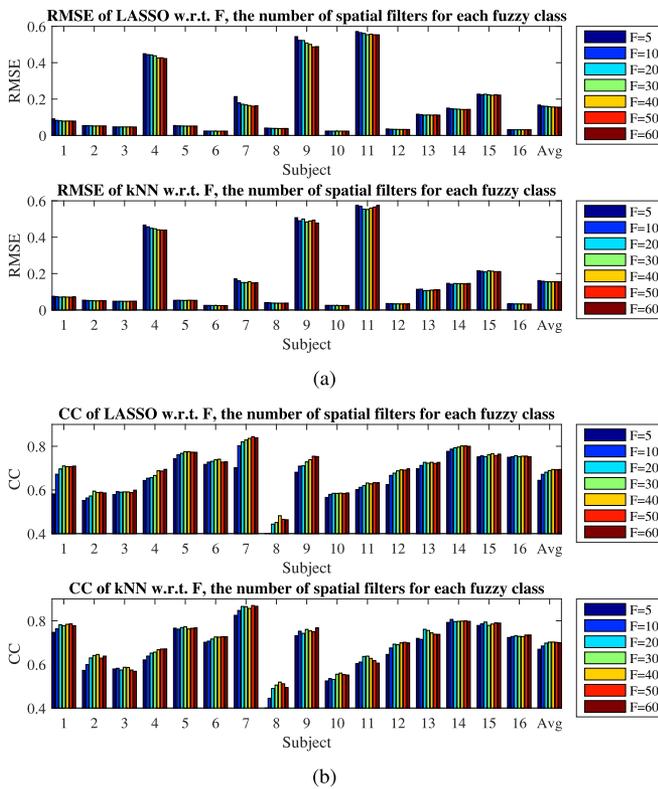
Fig. 9. (a) RMSEs and (b) CCs of LASSO and kNN with respect to F, the number of spatial filters for each fuzzy class in Algorithm 1.
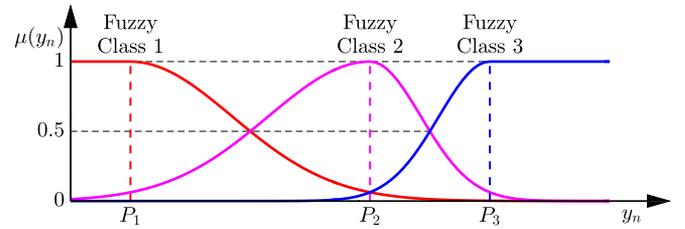
Fig. 10. Three fuzzy classes for y_n, when Gaussian fuzzy sets are used.
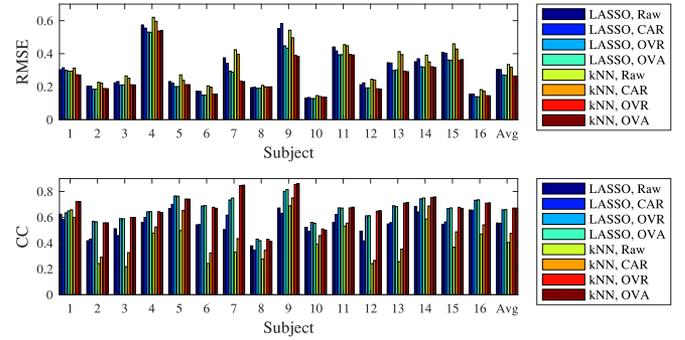
Fig. 11. RMSEs and CCs of the 8 approaches on the 16 subjects, when the three Gaussian fuzzy sets in Fig. 10 are used in CSPR-OVR and CSPR-OVA.
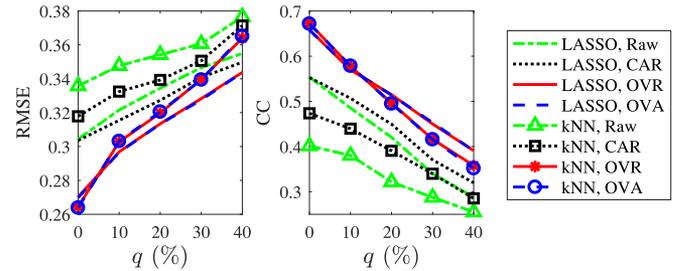
Fig. 12. Average RMSEs and CCs of the eight approaches wrt different attribute noise levels.

As in [64], for each model input, we randomly replaced $q\%$ ($q = 0, 10, \ldots, 40$) of all trials from a subject with a uniform noise between its minimum and maximum values. After this was done for both the training and testing data, we extracted feature sets Raw, CAR, OVR, and OVA, and trained LASSO and kNN, on the corrupted training data. We then tested their performances on the corrupted testing data. The results are shown in Fig. 12. Generally, as the noise level increased, the performances decreased, which is intuitive. However, OVR and OVA achieved better RMSEs and CCs than Raw and CAR at almost all noise levels, suggesting that it is still beneficial to use CSPR-OVR and CSPR-OVA even under high attribute noise.

F. Computational Cost

Observe from Algorithm 1 that in training, CSPR-OVR needs to perform a matrix inversion and an eigen decomposition to compute $\mathbf{W}^*$; however, once the training is done, the filtering of new EEG trials can be conducted very efficiently by a simple
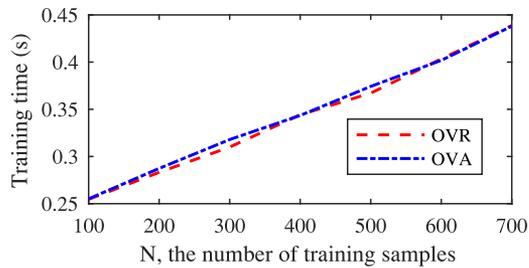
Fig. 13.   Training time of CSPR-OVR and CSPR-OVA wrt N.

matrix multiplication [see (3)]. Let N be the number of training samples. Then, the actual training time of CSPR-OVR and CSPR-OVA increased linearly with N, as shown in Fig. 13. The platform was a Dell XPS15 laptop (Intel i7-6700HQ CPU @2.60 GHz, 16 GB memory) running Windows 10 Pro 64-bit and MATLAB 2016b. A least-squares curve fit shows that the training time is $0.2216 + 0.0003\,N$ s, which should not be a problem for a practical N.

VI. DISCUSSIONS AND FUTURE RESEARCH

Recall that fivefold crossvalidation was used in the performance evaluation in the previous section, i.e., we concatenated the nine-session data from the same subject, randomly partitioned them into five equal-length folds, and then used fourfolds for training and the remaining one for testing. So, the training and testing folds contained data from the same sessions. This is equivalent to the case that we label some session-specific data in offline regression. Our results showed that in this case, CSPR-OVR and CSPR-OVA can significantly improve the regression performance.

To avoid the use of session-specific data, we also investigated a different validation method: leave-one-session-out validation, in which for each subject, we trained the spatial filters using eight sessions and tested them on the remaining session. Interestingly, all four feature sets and both regression models achieved very poor performance here. The reasons are as follows: 1) We need a proper way to normalize the RSs from different sessions, as done for the response times in [16]; and, 2) there is large intra-subject variation, meaning that the EEG responses for the same subject vary at different times (recall that these nine sessions were collected at different days); so, the patterns learned from the previous sessions become obsolete for the new session, and hence spatial filtering alone does not help. However, our previous research [58], [60], [62] has shown that transfer learning can cope well with the intersubject variation (individual differences) in both classification and regression problems, and we conjecture that it can also handle the intrasubject variation. One of our future research directions is to demonstrate the performance of CSPR-OVR and CSPR-OVA in a transfer learning framework to individualize a generalized model for regression problems, as done in [18] and [46] for EEG-based cognitive performance classification.

Another direction of our future research will apply CSPR-OVR and CSPR-OVA to other important EEG-based regression problems, e.g., drowsiness (or alertness) estimation during driving, and integrate it with more sophisticated feature extraction approaches, e.g., Riemannian geometry [8], for better regression performance.

VII. CONCLUSION

EEG signals are easily contaminated by artifacts and noise, so preprocessing is needed before they can be used by a machine learning algorithm in BCI. Spatial filters, e.g., ICA, xDAWN, CSP, and CCA, have been widely used to increase the EEG signal quality for classification problems, but their applications in BCI regression problems have been very limited. In this paper, we have proposed two CSP filters for EEG-based regression problems in BCI, which were extended from the CSP filter for classification, by making use of fuzzy sets. Extensive experimental results on EEG-based RS estimation from a large-scale study, which collected 143 sessions of PVT data from 17 subjects during a 5-month period, demonstrated that the proposed spatial filters can significantly increase the EEG signal quality. When used in LASSO and kNN, the spatial filters can reduce the estimation RMSE by 10.02–19.77%, and at the same time increase the CC by 19.39–86.47%.

REFERENCES

[1] T. Akerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci*, vol. 52, no. 1/2, pp. 29–37, 1990.

[2] A. Barachant, "MEG decoding using Riemannian geometry and unsupervised classification," 2014. [Online]. Available: http://alexandre.barachant.org/wp-content/uploads/2014/08/documentation.pdf. Accessed: August 17, 2016.

[3] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc., Ser. B*, vol. 57, pp. 289–300, 1995.

[4] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Frontiers Neuroinformat.*, vol. 9, 2015.

[5] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao, "A high-speed BCI based on code modulation VEP," *J. Neural Eng.*, vol. 8, no. 2, 2011.

[6] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method," *J. Neural Eng.*, vol. 6, no. 4, 2009.

[7] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[8] M. Congedo, A. Barachant, and A. Andreev, "A new generation of brain-computer interface based on Riemannian geometry," arXiv: 1310.8115, Jan. 2013.

[9] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, pp. 9–21, 2004.

[10] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behavior Res. Methods, Instrum. Comput.*, vol. 17, no. 6, pp. 652–655, 1985.

[11] G. Dornhege, G. C. B. Blankertz, and K.-R. Muller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[12] O. Dunn, "Multiple comparisons among means," *J. Amer. Statist. Assoc.*, vol. 56, pp. 62–64, 1961.

[13] O. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, pp. 214–252, 1964.

[14] G. H. Golub and C. F. V. Loan, *Matrix Computation*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[15] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[16] Z. Hu, Y. Sun, J. Lim, N. Thakor, and A. Bezerianos, "Investigating the correlation between the neural activity and task performance in a psychomotor vigilance test," in *Proc. 37th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Milan, Italy, Aug. 2015, pp. 4725–4728.

[17] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw.*, vol. 13, no. 4, pp. 411–430, 2000.

[18] R. R. Johnson, D. P. Popovic, R. E. O. M. Stikic, D. J. Levendowski, and C. Berka, "Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model," *Biol. Psychol.*, vol. 87, 2011, Art. no. 241250.

[19] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.

[20] T.-P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[21] S. Kerick *et al.*, "Inter- and intra-individual variations in sleep, subjective fatigue, and vigilance task performance of students in their real-world environments over extended periods," to be published.

[22] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.

[23] T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition," *J. Clin. Neurophysiol.*, vol. 14, no. 1, pp. 73–82, 1997.

[24] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain-computer interface technologies in the coming decades," *Proc. IEEE*, vol. 100, no. 3, pp. 1585–1599, May 2012.

[25] L.-D. Liao *et al.*, "Biosensor technologies for augmented brain-computer interfaces in the next decades," *Proc. IEEE*, vol. 100, no. 2, pp. 1553–1566, May 2012.

[26] C. T. Lin *et al.*, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst.*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.

[27] C.-T. Lin *et al.*, "Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1582–1591, May 2008.

[28] C.-T. Lin, L.-W. Ko, I.-F. Chung, T.-Y. Huang, Y.-C. Chen, T.-P. Jung, and S.-F. Liang, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits Syst., I*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.

[29] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain-computer interfaces," *Proc. IEEE, Special Centennial Issue*, vol. 100, pp. 1567–1584, May 2012.

[30] E. M. Maynard, C. T. Nordhausen, and R. A. Normann, "The Utah intracortical electrode array: A recording structure for potential brain-computer interfaces," *Electroencephalography Clin. Neurophysiol.*, vol. 102, no. 3, pp. 228–239, 1997.

[31] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalography Clin. Neurophysiol.*, vol. 103, pp. 386–394, 1997.

[32] J. Mellinger *et al.*, "An MEG-based brain-computer interface (BCI)," *Neuroimage*, vol. 36, no. 3, pp. 581–593, 2007.

[33] N. Naseer and K.-S. Hong, "fNIRS-based brain-computer interfaces: A review," *Frontiers Human Neurosci.*, vol. 9, no. 3, 2015.

[34] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces: A review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[35] X. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 8, no. 4, 2011, Art. no. 046028.

[36] C. C. Ragin, *Fuzzy-Set Social Science*. Chicago, IL, USA: Univ of Chicago Press, 2000.

[37] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[38] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.

[39] B. Rivet, H. Cecotti, A. Souloumiac, E. Maby, and J. Mattout, "Theoretical analysis of xDAWN algorithm: Application to an efficient sensor selection in a P300 BCI," in *Proc. 19th Eur. Signal Process. Conf.*, Barcelona, Spain, Aug. 2011, pp. 1382–1386.

[40] B. Rivet and A. Souloumiac, "Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis," *Signal Process.*, vol. 93, no. 2, pp. 387–398, 2013.

[41] R. N. Roy, S. Bonnet, S. Charbonnier, P. Jallon, and A. Campagne, "A comparison of ERP spatial filtering methods for optimal mental workload estimation," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 7254–7257.

[42] C. Russell, J. Caldwell, D. Arand, L. Myers, P. Wubbels, and H. Downs, "Validation of the fatigue science readiband actigraph and associated sleep/wake classification algorithms," 2015. [Online]. Available: http://static1.squarespace.com/static/550af02ae4b0cf85628d981a/t/5526c99ee4b019412c323758/1428605342303/Readiband_Validation.pdf. Accessed on: Aug. 8, 2016.

[43] F. Sagberg, P. Jackson, H.-P. Kruger, A. Muzer, and A. Williams, "Fatigue, sleepiness and reduced alertness as risk factors in driving," Inst. Transp. Econ., Oslo, Norway, Tech. Rep. TOI Rep. 739/2004, 2004.

[44] R. Sitaram *et al.*, "fMRI brain-computer interface: A tool for neuroscientific research and treatment," *Comput. Intell. Neurosci.*, 2007, Art. no. 25487.

[45] M. Spuler, A. Walter, W. Rosenstiel, and M. Bogdan, "Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1097–1103, Nov. 2014.

[46] M. Stikic, R. R. Johnson, D. J. Levendowski, D. P. Popovic, R. E. Olmstead, and C. Berka, "EEG-derived estimators of present and future cognitive performance," *Frontiers Human Neurosci.*, vol. 5, 2011.

[47] D. S. Tan and A. Nijholt, Eds., *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. London, U.K.: Springer, 2010.

[48] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1–11, 2002.

[49] J. A. Uriguen and B. Garcia-Zapirain, "EEG artifact removal— State-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, 2015.

[50] US Department of Defense Office of the Secretary of Defense, "Code of federal regulations protection of human subjects," Government Printing Office, no. 32 CFR 19, 1999.

[51] US Department of the Army, "Use of volunteers as subjects of research," Government Printing Office, no. AR 70-25, 1990.

[52] "Traffic safety facts crash stats: Drowsy driving," US Dept. Transp., Nat. Highway Traffic Safety Admin., Washington, DC, USA. [Online]. Available: http://www-nrd.nhtsa.dot.gov/pubs/811449.pdf

[53] J. van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.

[54] R. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589–593, May 2000.

[55] L.-X. Wang, *A Course in Fuzzy Systems and Control*. Upper Saddle River, NJ, USA: Prentice Hall, 1997.

[56] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, T.-P. Jung, N. Bigdely-Shamlo, and C.-T. Lin, "Selective transfer learning for EEG-based drowsiness detection," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, pp. 3229-3232, Hong Kong, Oct. 2015.

[57] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, no. 2, pp. 70–73, Jun. 1967.

[58] D. Wu, "Online and offline domain adaptation for reducing BCI calibration effort," *IEEE Trans. Human-Mach. Syst.*, 2017, to be published.

[59] D. Wu, C.-H. Chuang, and C.-T. Lin, "Online driver's drowsiness estimation using domain adaptation with model fusion," in *Proc. Int. Conf. Affective Comput. Intell. Interaction*, Xi'an, China, Sep. 2015, pp. 904–910.

[60] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR)," *IEEE Trans. Fuzzy Syst.*, to be published.

[61] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Budapest, Hungary, Oct. 2016, pp. 730–736.

[62] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Spectral meta-learner for regression (SMLR) model aggregation: Towards calibrationless brain-computer interface (BCI)," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Budapest, Hungary, Oct. 2016, pp. 743–749.

[63] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.

[64] X. Zhu and X. Wu, "Class noise vs. attribute noise: A quantitative study of their impacts," *Artif. Intell. Rev.*, vol. 22, pp. 177–210, 2004.
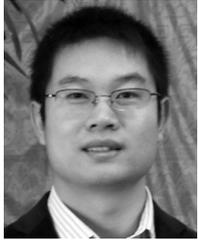
Dongrui Wu (S'05–M'09–SM'14) received the B.E. degree in automatic control from the University of Science and Technology of China, Hefei, China, in 2003, the M.Eng. degree in electrical engineering from the National University of Singapore, Singapore, in 2005, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2009.

He was a Lead Research Engineer with GE Global Research from 2010 to 2015. He is currently a Chief Scientist with DataNova, Clifton Park, NY, USA. He has more than 90 publications, including a book entitled *Perceptual Computing* (Wiley-IEEE, 2010). His research interests include affective computing, brain–computer interface, computational intelligence, and machine learning.

Dr. Wu received the IEEE International Conference on Fuzzy Systems Best Student Paper Award in 2005, the IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award in 2012, the IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2014, and the North American Fuzzy Information Processing Society Early Career Award in 2014. He was a finalist of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING Most Influential Paper Award in 2015, and IEEE Brain Initiative Best Paper Award in 2016. He is an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON HUMAN–MACHINE SYSTEMS, and the *IEEE Computational Intelligence Magazine*.

Jung-Tai King received the B.S. degree in psychology from the National Cheng-Chi University, Taipei, Taiwan, in 1998, the M.S. degree in criminology from the National Chung-Cheng University, Chiayi County, Taiwan, in 2001, and the Ph.D. degree in neuroscience from the National Yang-Ming University, Taipei, Taiwan, in 2010.

He is currently an Assistant Research Fellow in the Brain Research Center, National Chiao-Tung University, Hsinchu, Taiwan. His research interests include psychophysiology, cognitive and social neuroscience, and neuromarketing.

Chun-Hsiang Chuang received the B.S. degree in mathematics education from Taipei Municipal Teachers College, Taipei, Taiwan, in 2004, the M.S. degree in educational measurement and statistics from the National Taichung University, Taichung, Taiwan, in 2009, and the Ph.D. degree in electrical engineering from the National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 2014.

He was a Visiting Scholar with the Swartz Center for Computational Neuroscience, University of California at San Diego, La Jolla, CA, USA, from 2012 to 2013. During 2014–2016, he was a Postdoctoral Researcher and an Assistant Researcher with the Brain Research Center, NCTU. He is currently a Lecturer (Assistant Professor) in the University of Technology, Sydney, Australia. His current research interests include machine learning, computational neuroscience, biomedical signal processing, and the brain–computer interface.

Chin-Teng Lin (S'88–M'91–SM'99–F'05) received the B.S. degree in control engineering from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1986, and the Master's and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively.

He is currently a Distinguished Professor of the Faculty of Engineering and Information Technology, University of Technology Sydney. He also holds an Honorary Chair Professorship of Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego, La Jolla, CA, USA, and an Honorary Professorship of the University of Nottingham, Nottingham, U.K. He is the coauthor of *Neural Fuzzy Systems* (Prentice-Hall, 1996), and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (World Scientific, 1994). He has published more than 220 journal papers (Total Citation: 18 236, H-index: 57, i10-index: 295) and holds 97 patents in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain–computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal papers.

Dr. Lin was elevated to an IEEE Fellow for his contributions to biologically inspired information systems in 2005 and was elevated to an International Fuzzy Systems Association Fellow in 2012. He received the Merit National Science Council Research Fellow Award, Taiwan in 2009, the Outstanding Electrical and Computer Engineer, Purdue University in 2011, the Outstanding Achievement Award by Asia Pacific Neural Network Assembly in 2013, and the IEEE Fuzzy Systems Pioneer Award in 2016. He served as the Deputy Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II in 2006–2008 and as the Editor in-Chief of the IEEE Transactions on Fuzzy Systems from 2011 to 2016. He also served on the Board of Governors of the IEEE Systems, Man, Cybernetics Society during 2003–2005, the IEEE Circuits and Systems (CAS) Society during 2005–2008, and the IEEE Computational Intelligence Society (CIS) during 2008–2010. He was the Chair of the IEEE Taipei Section during 2009–2010. He was a Distinguished Lecturer of the IEEE CAS Society from 2003 to 2005 and the CIS Society from 2015 to 2017. He was the Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics in 2005 and the General Chair of the 2011 IEEE International Conference on Fuzzy Systems.

Tzyy-Ping Jung (S'91–M'92–SM'06–F'15) received the B.S. degree in electronics engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the Ohio State University, Columbus, OH, USA, in 1989 and 1993, respectively.

He is currently a Research Scientist and the Co-Director of the Center for Advanced Neurological Engineering, Institute of Engineering in Medicine, University of California-San Diego (UCSD), La Jolla, CA, USA. He is also an Associate Director of the Swartz Center for Computational Neuroscience, Institute for Neural Computation, and an Adjunct Professor of bioengineering in UCSD. In addition, he is an Adjunct Professor of computer science in the National Chiao Tung University, and an an Adjunct Professor in the School of Precision Instrument and Opto-Electronic Engineering, Tianjin University, Tianjin, China. His research interests include the areas of biomedical signal processing, cognitive neuroscience, machine learning, EEG, functional neuroimaging, and brain–computer interfaces and interactions.