

Mental State Estimation for Brain–Computer Interfaces

Koel Das, Daniel S. Rizzuto, and Zoran Nenadic*, *Member, IEEE*

Abstract—Mental state estimation is potentially useful for the development of asynchronous brain–computer interfaces. In this study, four mental states have been identified and decoded from the electrocorticograms (ECoGs) of six epileptic patients, engaged in a memory reach task. A novel signal analysis technique has been applied to high-dimensional, statistically sparse ECoGs recorded by a large number of electrodes. The strength of the proposed technique lies in its ability to jointly extract spatial and temporal patterns, responsible for encoding mental state differences. As such, the technique offers a systematic way of analyzing the spatiotemporal aspects of brain information processing and may be applicable to a wide range of spatiotemporal neurophysiological signals.

Index Terms—Brain–computer interfaces (BCIs), classification, curse of dimensionality, electrocorticograms (ECoGs), feature extraction, mental states, small sample size problem.

I. INTRODUCTION

BRAIN–COMPUTER interfaces (BCIs) have received considerable attention in recent years. The primary function of BCIs is to generate control signals for external devices based on real-time analysis of measured brain signals. Since this can be accomplished while bypassing the spinal cord and periphery, an obvious potential of BCI technology lies in providing assistance to individuals with severe motor disabilities, often resulting from amyotrophic lateral sclerosis, brainstem stroke, and high cervical spinal cord injury [1]. While these patients have little or no muscular control, their cognitive functions may be unaffected by the paralysis. The goal of BCI is to convert these thoughts into actions. Potential applications include computer interfaces (e.g., virtual keyboard [2] or virtual reality environments [3]), mobile robots [4], functional muscle stimulation [5], and many more. The clinical success of BCI technology may, therefore, significantly change the quality of life of severely paralyzed.

Successful demonstration of BCI control based on invasive recording technologies has been documented in both animal [6]–[8] and human [9], [10] studies. Likewise, encouraging

results have been reported in humans with noninvasive brain recording technologies such as surface EEG [1], [2], [11], [12]. Despite these advances, a widespread clinical use of BCI technology has not been achieved [13]. For invasive BCI systems based on single-neuron recordings, the biggest obstacle to their clinical application is inability to record stable signals over long periods of time [10]. Noninvasive BCIs, on the other hand, exhibit highly variable subject-specific performances, and a significant number of subjects are not able to achieve a purposeful control of the device [14]. In addition, certain noninvasive BCIs require a relatively long training process [12].

Real-world and clinical BCI applications will likely require asynchronous operation, wherein the subject intentionally engages the interface at his/her own pace. This subject-paced control is a more natural control mode than the commonly used synchronous control, where the subject engages the interface in response to externally supplied cues [15]. Birch *et al.* [16] recorded EEG signals in response to subjects' self-paced thumb movements and were able to accurately classify single EEG trials into idling and movement states. These results spurred an interest in the development of asynchronous BCIs by several groups [4], [17]–[22].

In a typical asynchronous BCI system, it is assumed that the brain undergoes a series of transitions over a discrete set of “brain states,” as the subject is engaged in mental tasks. These brain states and corresponding transitions are defined by changes in brain activity patterns. A classical example of a distinct neural pattern is the readiness potential (RP) [23], which underlies the planning of voluntary movements. Perceptual and cognitive states have also been identified with functional neuroimaging techniques [24]. To support communication and control, an asynchronous BCI system must be able to infer at least two brain states from brain data. The finite-state machine (FSM) diagram of one such system is shown in Fig. 1. At any time, the system resides in one of the two states, *idle* or *move*¹ and the state transitions are initiated based on real-time analysis of continuously acquired brain data. Note that the present definition of *move* state is ignorant of movement directions. If the system is in *move* state, the direction of movement such as *move left* or *move right* can be inferred subsequently through a separate decoding process. Also note that a general BCI system can have more than two states.

In this study, four mental states are estimated based on electrocorticograms (ECoGs) of six patients with medically intractable epilepsy. ECoGs have been considered as an alternative source of control for BCI applications [9], [25]–[27], since they have

Manuscript received March 6, 2008; revised September 16, 2008 and January 20, 2009. First published May 19, 2009; current version published July 15, 2009. Asterisk indicates corresponding author.

K. Das was with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697 USA. She is now with the Department of Psychology, University of California, Santa Barbara, CA 93106 USA (e-mail: kdas@uci.edu).

D. S. Rizzuto was with the Division of Biology, California Institute of Technology, Pasadena, CA 91125 USA. He is now with the Swedish Neuroscience Institute, Seattle, WA 98122 USA (e-mail: dan.rizzuto@gmail.com).

*Z. Nenadic is with the Department of Biomedical Engineering and the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697 USA (e-mail: znenadic@uci.edu).

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

Digital Object Identifier 10.1109/TBME.2009.2022948

¹These states are referred to as *no control* and *intentional control* by some researchers [15], while state transitions are referred to as “brain switch.”

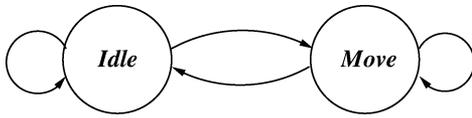


Fig. 1. FSM diagram of the minimal asynchronous BCI system.

better spatial resolution than EEG signals and require less invasive surgical procedures than the implantation of microarrays, commonly used in invasive BCIs [8], [10], [28]. Failure to reliably decode mental states may present a serious obstacle to the development of real-world BCIs.

The paper is organized as follows. In Section II, technical challenges associated with the analysis of high-dimensional, statistically sparse brain data are discussed. To handle these challenges while minimizing the number of assumptions, a novel pattern recognition algorithm [29] is applied. Section III is concerned with experimental setup and the details of the proposed algorithm. The validation of the technique on high-dimensional ECoG data is presented in Section IV, followed by discussion and conclusion in Section V.

II. TECHNICAL CHALLENGES IN ANALYSIS OF MULTISENSOR BRAIN DATA

Translation of brain signals into control variables for external devices typically entails the selection of one or more variables from a discrete set of controls [1], [30], [31].² This is achieved by compiling data over multiple trials into a *training database*. The statistical differences of data under various stimulus or cognitive conditions are then exploited so that novel brain data can be classified into one of the existing classes (e.g., *idle* versus *move*), and appropriate control signals are generated. The classification of brain signals in BCI applications, however, is not straightforward, since data are typically high dimensional. For example, 500 ms of EEG data, sampled at 200 Hz and recorded with 19 electrodes of the 10-20 international standard, represent a vector in 1900-dimensional (1900-D) space. Such a high-dimensional space presents many challenges, most notably the *small sample size problem*, which arises when the dimension of data exceeds the size of the training database. Practically, this is always the case in BCI applications, since the training time is limited.

In pattern recognition theory, these challenges are addressed by reducing the dimension of data through extraction of a small set of useful features. Numerous feature-extraction techniques have been developed over time, ranging from classical discriminant analysis tools [33] and their various extensions [34] to the use of probabilistic measures [35], [36]. Ironically, a majority of these techniques require data dimension be reduced before features can be extracted.³

Common approaches to feature extraction in BCI applications rely on separation of spatial and temporal domains. For example, EEG data are first processed spatially by the

²Exceptions are regression-based BCIs (see [32], for example); however, signal processing challenges are quite similar.

³A notable exception is feature-extraction techniques used in face recognition (see Section III-B for more details).

Laplacian filter [11], [12], [37], followed by temporal processing such as filtering in various frequency bands (e.g., μ -band or β -band [11], [12]). A common strategy used in conjunction with the aforementioned approaches is to handpick a small subset of channels (typically in the vicinity of the Rolandic C3 and C4 areas [12], [30]) according to their individual usefulness. While intuitive, this approach ignores joint statistical properties of channels, which may lead to suboptimal performances. In particular, based on classical results [38]–[40] from pattern recognition theory, it is conceivable that combining a “bad” channel with a good one may boost the performance. A more adaptive approach is to use data-driven spatial processing such as *common spatial pattern* (CSP) filtering [41], [42]. This method allows the number of channels to be systematically reduced (see [43] for review); however, spatial and temporal processing remain separated [41]–[45].

A direct approach to joint extraction of spatial and temporal features is to treat space and time simultaneously. As mentioned earlier, this normally gives rise to a severe small sample size problem, where a direct implementation of classical discriminant tools such as linear discriminant analysis (LDA) [33] is not possible. The problem is ultimately caused by singular covariance matrices. While regularization algorithms exist [46], they do not solve the *curse of dimensionality problem*⁴ *per se*. It should be noted that similar challenges are faced in the field of image recognition, although image recognition techniques have not been extensively tested outside of their primary application domain.

In this paper, a computationally efficient, locally adaptive feature extraction and classification algorithm [29], [48] is used for mental state decoding. The algorithm does not invoke the space-time separability assumption and avoids the pitfalls of heuristic approaches such as channel handpicking. In addition, the method provides an insight into the spatiotemporal aspects of neural information processing.

III. MATERIALS AND METHODS

A. Experimental Setup

Experimental results reported in this paper are based on ECoG recordings from the brains of six epileptic patients while they were making directed arm movements to various locations on a touch screen interface (see Fig. 2). A brief account of the experiment is presented here. For original reference, see [49].

At the beginning of each trial, a fixation stimulus was presented in the center of the touch screen and the subject initiated the trial by placing his right hand on the stimulus. This marked the onset of a *fixation* period. After a short period, a peripheral target was flashed on the screen at one of six or eight locations (see Table I for details), which marked the onset of a *target* period. The subjects were instructed not to move their eyes, but no quantitative assessment of their eye movements was available. The *target* period was followed by a *memory* period, whose onset

⁴Inversion and spectral decomposition of large matrices may not be feasible with standard computer architectures. An efficient solution to this problem has recently been proposed in [47].

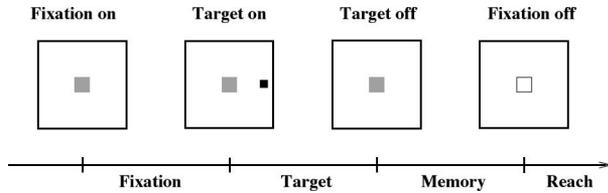


Fig. 2. Timeline of a single experimental trial.

TABLE I
SUBJECT-SPECIFIC INFORMATION: ELECTRODE TYPE, NUMBER OF ELECTRODES N_e , DIMENSION OF 1 S OF DATA n , (SEE THE APPENDIX), NUMBER OF TRIALS n_t , AND NUMBER OF TARGET LOCATIONS N_t

Subject	Electrode Type	N_e	n	n_t	N_t
S008	bilateral depths	91	18200	1752	6
S007	bilateral depths	61	12200	1948	6
S006	L grids	126	25200	1792	6
S005	R depths/grids	76	15200	1572	6
S004	R grids	123	24600	1532	8
S003	bilateral depths	75	15000	1737	8

was defined by extinguishing the peripheral target. The fixation stimulus was extinguished after the *memory* period, which acted as a cue for the participants to reach to the memorized location and marked the onset of a *reach* period. Upon completion of the reach, a beep indicated whether it was correct or not and a new trial was initiated after a brief delay period. To prevent phase-locked response, the durations of *fixation*, *target*, and *memory* periods were randomized (uniform distribution between 1 and 1.3 s). The position of the peripheral target was parametrized by an angle, e.g., 0° (right), 45° (top-right), 180° (left), etc., and was randomly interleaved on a trial-by-trial basis. The number of trials varied per target location and per subject. The ECoG signals were amplified, sampled at 200 Hz, and bandpass filtered (0.3–70 Hz).

Table I lists experimental details for each participant. Electrode placement was determined by the neurologist so as to best localize epileptogenic regions. A combination of grid and depth electrodes was used for the study. Note that bilateral depth electrodes were placed on both right (R) and left (L) hemispheres, whereas grid electrodes covered a single hemisphere. Since the seizure focus was often in the temporal regions, grid electrodes were typically placed over these areas. The patients were kept off antiepileptic medications for the duration of the study.

The state data were defined by pooling all data from a single period, regardless of the target location. Since the durations of individual periods were randomized, a 1-s-long segment of data from each period was used in the process. This yielded training datasets containing between 1532 and 1948 trials per subject (cf., Table II). These datasets were used to decode the four mental states (*fixation*, *target*, *memory*, *reach*) and quantify the classification rates.

B. Classwise Principal Component Analysis

Classification of high-dimensional, statistically sparse data is often encountered in the field of face (image) recognition, where numerous techniques have been proposed (their comprehensive

TABLE II
CLASSIFICATION RATES (IN PERCENT) AND COMPUTATIONAL COST OF CPCA AND COMPETING METHODS

Subject	Method	$m = 1$	$m = 2$	$m = 3$	$m = 4$	Cost
S008	PLDA	35.93	39.74	41.41	-	64.14
	DLDA	57.54	74.20	79.07	-	9.38
	RLDA	66.50	80.83	92.17	-	425.00
	CSP	-	51.70	-	58.59	39.75
	CPCA	75.36	84.16	93.37†	93.23	65.60
S007	PLDA	37.39	36.86	38.03	-	74.31
	DLDA	41.86	54.73	55.39	-	6.80
	RLDA	50.33	59.94	61.05	-	503.00
	CSP	-	46.04	-	50.23	21.35
	CPCA	57.17	61.16	65.40	66.08†	58.51
S006	PLDA	26.16	25.89	26.85	-	79.79
	DLDA	39.08	41.30	41.16	-	12.88
	RLDA	48.86	55.03	59.54†	-	542.00
	CSP	-	43.04	-	46.23	89.94
	CPCA	49.97	56.39	59.23	59.17	93.75
S005	PLDA	25.66	27.57	29.42	-	49.79
	DLDA	50.85	58.34	62.33	-	6.75
	RLDA	49.34	55.94	62.15	-	311.00
	CSP	-	62.41	-	67.28	21.80
	CPCA	55.12	62.68	67.84†	67.83	58.11
S004	PLDA	29.24	31.99	31.83	-	51.95
	DLDA	34.81	36.36	37.58	-	10.57
	RLDA	36.35	42.99	43.62	-	401.00
	CSP	-	47.48	-	49.01†	60.78
	CPCA	38.52	40.81	42.78	41.98	74.05
S003	PLDA	30.50	33.88	32.94	-	57.48
	DLDA	29.90	33.38	33.97	-	7.53
	RLDA	33.47	36.28	37.57	-	386.00
	CSP	-	34.21	-	35.08	24.51
	CPCA	36.61	41.22	44.12	45.11†	50.03

Chance performance is 25%.

review is outside the scope of this paper). Since spatiotemporal ECoG data are essentially an image, it may be worthwhile to cast a classification of ECoG data in the framework of image/face recognition. However, when applied to ECoG data, several representative face recognition techniques produced very modest results [29], [48], [50]. The discrepancy in performance may be attributed to relatively high levels of noise in electrophysiological neural data.

Principal component analysis (PCA) is one of the most popular general-purpose dimensionality reduction techniques and is capable of handling the small sample size problem. It is, however, ignorant of class-specific information, and so it can be argued that PCA is suboptimal for data classification problems. Motivated by this limitation, the authors have recently developed an algorithm that identifies and discards an uninformative subspace in data by performing a classwise PCA (CPCA). The residual space is then amenable to standard classification techniques, since small sample size conditions and the curse of dimensionality are no longer obstacles. A notable distinction of the technique is that it is nonlinear, yet efficiently implemented. The application of the algorithm will be demonstrated on the four-state decoding problem introduced before, and the details are given in the Appendix. For an in-depth exposition of CPCA and its application to other types of statistical data, see [29].

TABLE III
CPCA PERFORMANCE (IN PERCENT) BREAKDOWN BY THE FOUR STATES

Subject	Fixation	Target	Memory	Reach
S008	97.33	94.21	91.34	90.59
S007	84.62	82.86	58.28	42.72
S006	75.69	65.09	46.92	49.33
S005	64.53	69.67	56.49	80.74
S004	40.31	40.21	53.60	37.00
S003	47.28	46.89	45.37	36.93
Average	68.29	66.49	58.67	56.22

IV. RESULTS

A. Performance Evaluation

The algorithm was tested against two face recognition techniques: direct LDA (DLDA) [51] and PCA in conjunction with LDA (PLDA) [52], and against two widely used BCI techniques: CSP and regularized LDA (RLDA) with the ℓ_1 -norm on the regularizer [53]. The multiclass extension of the CSP technique was achieved using the one-versus-rest approach [54]. The quadratic Bayesian classifier was used in conjunction with these feature-extraction techniques, making them comparable to the CPCA algorithm [cf., (4)]. In all methods, raw time series were used as input data. The feature subspace dimension was limited to $m \leq 4$, as no significant change in performance was found in higher dimensions. Note that as variants of LDA, the PLDA, DLDA, and RLDA methods are limited to $m \leq 3$. The number of eigenvectors chosen for CSP was varied with a maximum of 4, i.e., two eigenvectors from both ends of the spectrum. The performances of the methods, expressed as percent of correct classification, were assessed through a tenfold cross-validation (CV) [33].

For each feature subspace dimension, the performance of the best technique is typeset in bold, and the best overall performance per subject is marked by † (see Table II). In general, the (optimal) subspace dimension should not be chosen based on performance, otherwise overfitting may occur. In authors' own experience, restricting m to a low number (e.g., $2 \leq m \leq 4$) is sufficient for most practical applications. Table II also reports the computational cost of the methods in terms of CPU time per fold, i.e., time taken to train the algorithm based on 90% of data and test the remaining 10% of trials. All computations were performed on a workstation (Intel Xeon processor, CPU speed 3.60 GHz, 8 GB RAM). The breakdown by the state of the CPCA performance in 3-D feature subspace is given in Table III.

B. Analysis of Results

The nonuniform electrode placement (dictated by patients' medical condition), the use of different electrode types (grid and depth), and intrinsic variability among subjects are likely causes for the variation in performance. The CPCA algorithm is highly competitive, as it achieves the best recognition rates on four subjects, often with a significant margin. RLDA and CSP perform best on subjects S006 and S004, respectively, whereas DLDA and especially PLDA perform poorly. This is consistent

with previous studies [29], [48], [50], where these and other face recognition techniques were found ill-suited for classification of noisy biomedical data. Also note that the computational cost of RLDA is an order of magnitude higher than those of the CPCA, PLDA, and CSP techniques. Conversely, the computational cost of DLDA is the lowest. Finally, note that with few exceptions, the techniques performed better in 3-D/4-D subspaces. Especially encouraging are the performances reported on subjects S008, S007, and S005, with S008 achieving accuracy of more than 93%.

From Table III, it can be concluded that in most cases, *fixation* and *target* are better predicted than the other two states, with the best rates reaching as high as 97%. In general, the decoding of the *memory* state was the most challenging, which is expected given its purely cognitive nature. A notable exception is subject S008, with a reported accuracy of more than 91%. The performances on other subjects were comparatively lower, although they were significantly better than chance (25%). The classification accuracy for the *reach* state was also comparatively low, with the exception of subjects S008 and S005, which perhaps can be attributed to the self-paced nature of this state. Namely, the state terminated whenever the subject reached the target, as detected by the touch screen interface, and only the first 1 s of data was used for training and decoding. Generally, the subjects exhibited significant trial-to-trial reach time variability. In addition, subject S003 had an uneven distribution of reach times over target positions.

In conclusion, CPCA is able to remove large portions of uninformative data from the original space, with dimensions as high as 25 200, and distill useful information down to 3-D/4-D features, thereby facilitating efficient mental state decoding. The algorithm is computationally simple and does not require expensive optimization of free parameters, such as regularizing constant or subset of electrodes and frequency bands. Unlike CSP, it is naturally defined in a multiclass setup and does not require the pairwise extension of binary classification problems. The biggest strength of the technique, however, is its ability to jointly extract information distributed in space and time. Consequently, the method may be viewed as a useful tool for analysis of spatiotemporal brain information processing.

V. DISCUSSION AND CONCLUSION

A. Discussion

We discuss the following issues based on state decoder performance. In general, from the classification rates, it follows that subjects with depth electrodes or a combination of grid and depth electrodes (S008, S007, S005) had a better performance than the subjects having only grid electrodes. Since depth electrodes provide a better source localization and have higher signal-to-noise ratio compared to the grid electrodes, this finding is not very surprising. Furthermore, grid electrodes were placed mostly in the temporal areas and over a single hemisphere, as opposed to bilateral depth electrodes that were placed in temporal, parietal, and frontal cortex, spreading over both right and left hemispheres.

1) *Application to Asynchronous BCIs*: To facilitate asynchronous BCI operation, brain data must be analyzed in real time. Practically, this means that a data window of a fixed length T is streamed to a computer for analysis, typically at a reasonably high rate (e.g., every 50 ms [12]). The choice of T is likely to be subject- and application-dependent [43] (it is nominally set to 1 s in this study). To classify a k th data window, the classification rule (4) can be adapted as

$$P(\omega_j | f_i^k) = \frac{p(f_i^k | \omega_j) P(\omega_j | f_i^{k-1})}{p(f_i^k)}, \quad i, j = 1, 2, 3, 4 \quad (1)$$

where $f_i^k = \mathbf{F}_i^T x^k$ is the projection of the data window x^k to the i th CPCA subspace (see the Appendix). Note that \mathbf{F}_i and the parameters of $p(\cdot | \omega_j)$ are to be estimated during the training process. Also note that the posterior probability $P(\omega_j | f_i^{k-1})$ is used as a prior probability at iteration k . This implies that the classification of present data depends on the classification of previous data, and this step is commonly used in recursive Bayesian filtering to smooth estimates over time [55]. It should be noted that analyzed data window may fall between two states; therefore, the performances reported in Table II and III may not be achievable during online asynchronous operation. Due to its simplicity and efficient implementation, the training of the algorithm can be repeated as needed. The retraining of the signal processing and classification algorithms may be necessary to compensate for nonstationary effects caused by learning and coadaptive processes [13], [56].

2) *Space–Time Separability Assumption*: As discussed in Section II, many BCI studies separate data processing in spatial and temporal domains. Since spatial filters, such as Laplacian, common average reference, PCA, and CSP, give rise to coefficients that remain constant in time, they are well suited to analysis of waves that are space–time separable. However, given the complex nature of brain waves governing processes (nonstationarities, inhomogeneities, nonlinearities, etc.) [57], the space–time separability assumption is unlikely to hold. To illustrate this point, optimal spatial filters for subject S008 were calculated. The filters were found through LDA, thereby maximizing the discriminant of the four states. To test whether a spatial pattern persists over time, LDA was performed at each instant using tenfold CV. Therefore, at each time point, LDA was performed ten times, and the temporal evolution of the average lead eigenvector (filter) is shown in Fig. 3. Consistent with the dynamics of ECoG signals [49], the invariance of the spatial filters began to break on time scales as short as 25 ms, and the trend was exacerbated over longer time scales. For example, large coefficient variations in the vicinity of electrode 55 at ~ 200 ms indicate that this brain region is much more important for mental state decoding than the area around electrode 23. Conversely, at ~ 800 ms, the most prominent state differences are found in the vicinity of electrode 23, while the importance of electrode 55 is negligible, and falls well within the noise fluctuations. Note that while the significant performance increase at ~ 150 ms coincides with a sharp transition in the filter shape, the performance in general is not a good measure of nonstationarity. In particular, the performances at ~ 200 and ~ 800 ms are quite

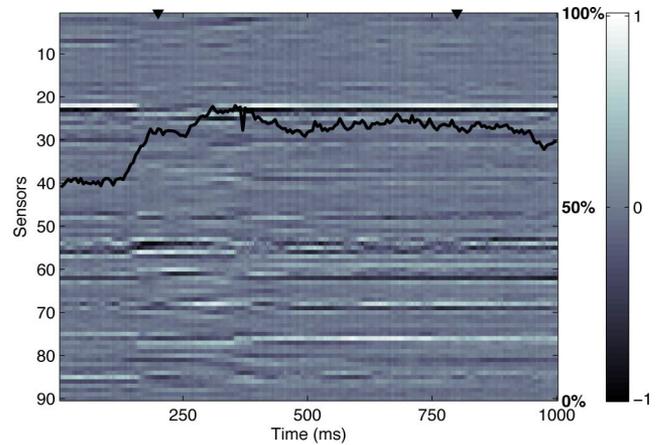


Fig. 3. (Normalized) Average coefficients of LDA spatial filters for subject S008. The solid line marks the classification accuracy (in percent), assessed through tenfold CV. Tick marks are at 200 and 800 ms.

similar ($\sim 70\%$), despite significantly different spatial filters. In conclusion, fixed spatial filters are likely to produce state classification rates that are suboptimal. The proposed method, on the other hand, avoids space–time separability assumption, and so, if state differences are encoded by two, or more, brain areas at different times, these differences will be jointly extracted into low-dimensional features.

3) *CPCA as Analysis Tool*: While features arising in CPCA are abstract, the corresponding filters (the columns of \mathbf{F}_i) have a clear physical interpretation. If the filter coefficients are organized into a spatiotemporal array (see the Appendix), they will explicitly point to brain areas, time scales, and latencies involved in encoding the mental state differences. For multidimensional features, there will be one filter per feature; however, if 1-D features are considered, filter analysis is considerably simpler. Fig. 4 shows four spatiotemporal filters computed from one of the folds of tenfold CV (see Section IV-A) for subject S008. Due to the piecewise linear nature of CPCA, each filter is locally adapted to a particular state: \mathbf{F}_1 (*fixation*), \mathbf{F}_2 (*target*), \mathbf{F}_3 (*memory*), and \mathbf{F}_4 (*reach*). This state-specific information is included in the filters through principal components $\mathbf{W}_i^{\text{CPCA}}$ (see the Appendix). On the other hand, the similarities in the filters may be attributed to matrices \mathbf{W}_B and $\mathbf{T}_i^{\text{DFE}}$. Therefore, information about all the states is present in all four filters, although it is expected that the features of each state are best extracted by their local filter, e.g., *fixation* by \mathbf{F}_1 .

By investigating the distribution of filter coefficients, an additional insight into the spatiotemporal dynamics of brain state representation is obtained. Within the first 150 ms, none of the filters point to any important areas. This is consistent with the latency of visual information processing [58], as all state transitions are triggered by visual cues (see Section III-A). All four filters feature important coefficients in the period from 150 to 300 ms, with similar spatial distributions. In particular, based on the coefficients of \mathbf{F}_2 , locally adapted to the *target* state, the following important brain areas are identified: orbitofrontal (OF) cortex, amygdala, and supplementary motor (SM) cortex.

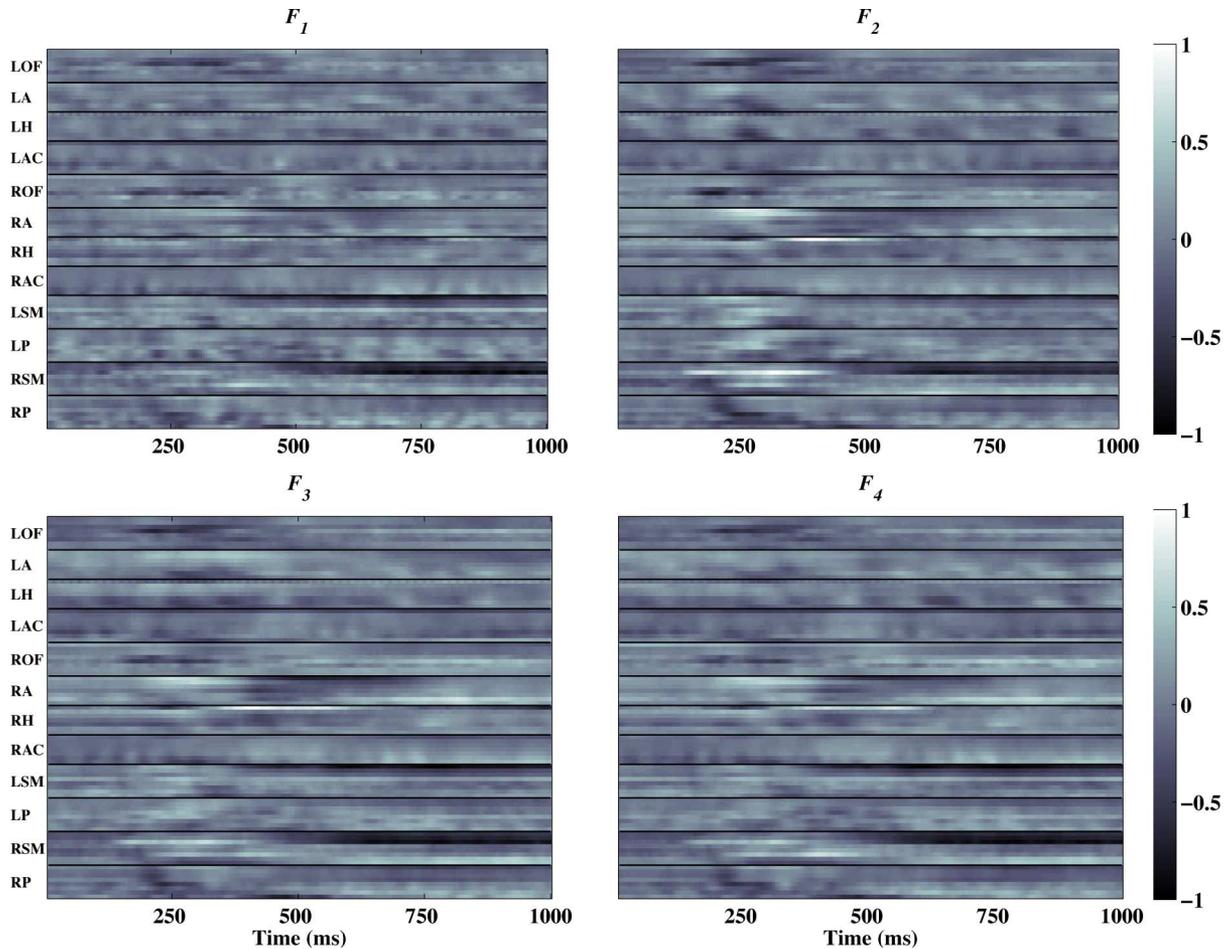


Fig. 4. Feature-extraction matrices $\mathbf{F}_i \in \mathbb{R}^{18 \times 200 \times 1}$ arranged into spatiotemporal filters. Note that the feature subspace is 1-D. Coefficients with large absolute value indicate spatiotemporal areas of importance. Electrodes were implanted over OF cortex, amygdala (A), hippocampus (H), anterior cingulate (AC) cortex, SM cortex, and parietal (P) cortex. Prefix indicates the hemisphere: left (L) and right (R).

The discriminatory information present after 500 ms is equally prominent in all four filters, especially in the SM areas.

Cortical potential studies have demonstrated that the SM areas are responsible for planning movements in humans [59]–[61]. Several studies have shown that the RP, recorded from the SM areas, appears as early as 2 s before voluntary movements [23], [60], [62]. This potential, distinguished by its slow wave-like behavior, occurs earlier in the SM areas than in any other cortical area [60]. The RP drift is also noted in persons imagining movements without actually executing any movement [63]. These findings are consistent with our result, showing that the SM areas carry significant state separability information. In addition, prominent coefficients in the SM area toward the end of the period may be movement-related. The latency of 500 ms, which is the characteristic of these coefficients, is consistent with the latency of motor response [64]. The OF cortex is reported to be involved in decision making during many decision-related tasks [65], [66], and these findings are also consistent with our study. Moreover, decision-making latencies, typically between 150 and 300 ms [64], are precisely aligned with the onset of important coefficients in the OF cortex (see Fig. 4). It is interesting to note that the contribution of the OF

cortex is not very prominent according to the filter \mathbf{F}_4 . This could be a consequence of the subject having already made a decision during the previous states: *target* and *memory*. Recent studies have shown that the amygdala region is associated with implicit learning and memory consolidation, when subjects are presented with visual stimuli [67]. This is congruous with our experimental protocol, where the subjects are given visual cues and the experiments are repeated over time. Accordingly, the amygdala region plays a significant role in the state discrimination process and shows up especially in \mathbf{F}_2 between 200 and 350 ms. It is interesting to note that the hippocampus and parietal regions did not seem to carry significant state information, as evidenced by the lack of important coefficients over these areas. These areas, however, typically exhibit directionally selective response and are, therefore, more suitable to decoding of the target location [48], [50], [68].

B. Conclusion

A novel feature-extraction/feature-classification algorithm [29] is applied to mental state estimation based on high-dimensional, statistically sparse ECoG signals recorded from a

large number of electrodes. The method takes a holistic approach to data analysis, wherein discriminatory information, distributed nonuniformly over time and space, is extracted jointly by low-dimensional features. Parallels between the proposed method and common BCI signal analysis tools are drawn, and important differences are highlighted. By analysis of spatiotemporal filters that are synthesized in the process, useful information can be uncovered about the spatiotemporal dynamics of the state encoding process. The method is applicable beyond BCI- and classification-related studies, as long as data can be conditioned upon two or more conditions such as stimulus versus baseline, or stimulus A versus stimulus B, etc. The success of the technique in classification of diverse statistical data [29] suggests that CPCA may be a suitable tool for other types of spatiotemporal brain data, such as functional magnetic resonance images and magnetoencephalography.

APPENDIX

Definition of Variables: Let $\mathbf{X} \in \mathbb{R}^{N_e \times 200}$ be a 1-s-long segment of ECoG signal, where N_e is the number of electrodes (see Table I). Let $x \in \mathbb{R}^n$ be a vectorized version of \mathbf{X} , where $n = N_e \times 200$. The vectorization scheme can be chosen arbitrary, but must be applied consistently. Note that the n coordinates of x contain both spatial and temporal information, which can be recovered by applying the inverse of the vectorization scheme. Let Ω be a state (label) random variable defined as $\Omega = \omega_i$ if x belongs to state, i ($i = 1, 2, 3, 4$). Let \mathcal{X}_i be a collection of data x conditioned upon ω_i . Practically, \mathcal{X}_i is obtained by pooling data (over multiple target positions) from a single period, as outlined in Section III-A. The state mean $\mu_i \in \mathbb{R}^n$ and covariance $\Sigma_i \in \mathbb{R}^{n \times n}$ are respectively defined as $\mu_i \triangleq 1/|\mathcal{X}_i| \sum_{x \in \mathcal{X}_i} x$ and $\Sigma_i \triangleq 1/(|\mathcal{X}_i| - 1) \sum_{x \in \mathcal{X}_i} (x - \mu_i)(x - \mu_i)^T$, where $|\cdot|$ denotes the cardinality of a set. The prior state probabilities are estimated empirically, i.e., $P(\omega_i) = |\mathcal{X}_i| / \sum_{j=1}^4 |\mathcal{X}_j|$.

The algorithm is a two-step procedure: feature-extraction matrices are first estimated from training data, followed by classification of test data in the feature space.

Feature Extraction: For a four-state decoding problem, the algorithm generates a family of feature subspaces $\{S_1, S_2, S_3, S_4\}$. Therefore, a notable distinction of the proposed technique is that it yields nonlinear (piecewise linear) features. However, since the individual subspaces are linear, they can be found through elementary matrix manipulations, implemented in MATLAB. Each subspace S_i is spanned by the columns of a feature-extraction matrix $\mathbf{F}_i \in \mathbb{R}^{n \times m_i}$, where m_i is the dimension of the feature space (chosen by the user). The feature matrices are constructed through the following factorization:

$$\mathbf{F}_i = [\mathbf{W}_i^{\text{CPCA}} | \mathbf{W}_B] \mathbf{T}_i^{\text{DFE}}, \quad i = 1, 2, 3, 4 \quad (2)$$

where $\mathbf{W}_i^{\text{CPCA}} \in \mathbb{R}^{n \times m_i}$ is a matrix whose columns are the m_i principal eigenvectors of Σ_i . The number m_i is determined by keeping the principal components with eigenvalues greater than the mean eigenvalue of Σ_i . In general, m_i depends on data distribution under ω_i and is likely to differ across states. Sensitivity analysis, which is performed by comparing the re-

sults of keeping $m_i - 1$, m_i , and $m_i + 1$ eigenvalues of Σ_i , revealed very stable results (differences $< 0.02\%$). The columns of $\mathbf{W}_B \in \mathbb{R}^{n \times 3}$ are chosen so that they form an orthonormal basis for the range⁵ of the between-class matrix $\Sigma_B \in \mathbb{R}^{n \times n}$, which is defined as $\Sigma_B \triangleq \sum_{i=1}^4 P(\omega_i) (\mu_i - \mu)(\mu_i - \mu)^T$ [33], where $\mu \triangleq \sum_{i=1}^4 P(\omega_i) \mu_i$ is the overall mean. Including \mathbf{W}_B in the definition (2) ensures that differences in states with similar covariance matrices Σ_i , but distinct means μ_i , are captured by features [29], [48]. Finally, the matrix $\mathbf{T}_i^{\text{DFE}} \in \mathbb{R}^{(m_i+3) \times m_i}$ can be any linear discriminant feature extraction (DFE) matrix (e.g., LDA). From this standpoint, the feature extraction defined by (2) can be viewed as a two-step procedure. In the first step, the matrix $[\mathbf{W}_i^{\text{CPCA}} | \mathbf{W}_B]$ provides a significant dimensionality reduction (from n to $m_i + 3$) by discarding a large noninformative subspace. In the second step, the class separability is enhanced and the dimension is further reduced (from $m_i + 3$ to m_i) through the matrix $\mathbf{T}_i^{\text{DFE}}$. In this study, approximate information discriminant analysis (AIDA) [69] has been used to calculate $\mathbf{T}_i^{\text{DFE}}$. Similar but slightly inferior results were obtained with LDA. Finally, the feature-extraction matrices \mathbf{F}_i are orthonormalized through the Gram-Schmidt procedure. Note that the columns of \mathbf{F}_i can be converted to spatiotemporal arrays by the devectorization scheme.

Classification: Let $x \in \mathbb{R}^n$ be test data to be classified. Projecting x onto the subspace S_i yields

$$f_i = \mathbf{F}_i^T x, \quad i = 1, 2, 3, 4 \quad (3)$$

where $f_i \in \mathbb{R}^{m_i}$ are the features of x in the subspace S_i . To find which of the four subspaces is optimal for classification, statistical tests need to be devised. In particular, by a straightforward application of the Bayes rule, we have

$$P(\omega_j | f_i) = \frac{p(f_i | \omega_j) P(\omega_j)}{p(f_i)}, \quad i, j = 1, 2, 3, 4 \quad (4)$$

where $P(\omega_j | f_i)$ is the posterior probability of the state ω_j , given the feature vector f_i . Note that there are 16 such probabilities. To complete the calculations in (4), $p(f_i | \omega_j)$ is modeled as a Gaussian probability density function with the mean $\mathbf{F}_i^T \mu_j$ and covariance $\mathbf{F}_i^T \Sigma_j \mathbf{F}_i$, whereas $p(f_i) = \sum_{j=1}^4 p(f_i | \omega_j) P(\omega_j)$. To assign x to one of the states, the maximum *a posteriori* (MAP) rule is applied in each subspace

$$J(i) = \arg \max_{1 \leq j \leq 4} P(\omega_j | f_i), \quad i = 1, 2, 3, 4$$

where $J(i)$ is the index of the most likely state in S_i . To find the overall maximum, the MAP rule is applied once again

$$i^* = \arg \max_{1 \leq i \leq 4} P(\omega_{J(i)} | f_i).$$

This formally concludes the feature-extraction/feature-classification procedure; the optimal subspace is S_{i^*} and the optimal class assignment is “ x belongs to state $J(i^*)$.”

⁵The range $\mathcal{R}(\Sigma) \triangleq \{y \in \mathbb{R}^n : y = \Sigma x, \forall x \in \mathbb{R}^n\}$ of a matrix $\Sigma \in \mathbb{R}^{n \times n}$ is easily computed in MATLAB using the function `orth.m`.

ACKNOWLEDGMENT

The authors would like to acknowledge the anonymous reviewers and the editorial staff of this journal.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 6, no. 113, pp. 767–791, 2002.
- [2] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, Mar. 1999.
- [3] R. Leeb, F. Lee, C. Keirnath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment," vol. 15, no. 4, pp. 473–482, Dec. 2007.
- [4] J. D. R. Millán, F. Renkens, J. Mouríño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1026–1033, Jun. 2004.
- [5] C. T. Moritz, S. I. Perlmutter, and E. E. Fetz, "Direct control of paralysed muscles by cortical neurons," *Nature*, vol. 456, no. 7222, pp. 639–642, Dec. 2008.
- [6] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. A. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, no. 6810, pp. 361–365, 2000.
- [7] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Instant neural control of a movement signal," *Nature*, vol. 416, no. 6877, pp. 141–142, 2002.
- [8] D. M. Taylor, S. I. Tillery, and A. B. Schwartz, "Direct cortical control of 3D neuroprosthetic devices," *Science*, vol. 296, no. 5574, pp. 1829–1832, 2002.
- [9] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 1, no. 2, pp. 63–71, Jun. 2004.
- [10] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, Jul. 2006.
- [11] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, 1997.
- [12] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 51, pp. 17849–17854, 2004.
- [13] A. Kübler, F. Nijboer, and N. Birbaumer, "Brain-computer interfaces for communication and motor control—perspectives and applications," in *Toward Brain Computer Interfacing*, G. Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2007, ch. 22, pp. 373–391.
- [14] A. Kübler and K.-R. Müller, "An introduction to brain-computer interfacing," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2007, ch. 1, pp. 1–25.
- [15] J. F. Borisoff, S. G. Mason, and G. E. Birch, "Brain interface design for asynchronous control," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2007, ch. 7, pp. 111–121.
- [16] G. E. Birch, P. D. Lawrence, and R. D. Hare, "Single-trial processing of event-related potentials using outlier information," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 59–73, Jan. 1993.
- [17] G. E. Birch and S. G. Mason, "Brain-computer interface research at the Neil Squire Foundation," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 193–195, Jun. 2000.
- [18] G. E. Birch, Z. Bozorgzadeh, and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 4, pp. 219–224, Dec. 2002.
- [19] S. P. Levine, J. E. Huggins, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, M. M. Rohde, E. A. Passaro, K. V. Elisevich, D. A. Ross, and B. J. Smith, "A direct brain interface based on event-related potentials," *IEEE Trans. Neural Syst. Rehabil.*, vol. 8, no. 2, pp. 180–185, Jun. 2000.
- [20] E. Yom-Tov and G. F. Inbar, "Detection of movement-related potentials from the electro-encephalogram for possible use in a brain-computer interface," *Med. Biol. Eng. Comput.*, vol. 41, no. 1, pp. 85–93, Jan. 2003.
- [21] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil.*, vol. 12, no. 2, pp. 258–265, Jun. 2004.
- [22] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface: Machine learning based detection of user specific brain states," *J. Universal Comput. Sci.*, vol. 12, no. 6, pp. 581–607, 2006.
- [23] L. Deecke, P. Scheid, and H. H. Kornhuber, "Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements," *Exp. Brain Res.*, vol. 7, no. 2, pp. 158–168, 1969.
- [24] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nat. Rev. Neurosci.*, vol. 7, no. 7, pp. 523–534, Jul. 2006.
- [25] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 954–962, Jun. 2004.
- [26] G. Schalk, J. Kubánek, K. J. Miller, N. R. Anderson, E. C. Leuthardt, J. G. Ojemann, D. Limbrick, D. Moran, L. A. Gerhardt, and J. R. Wolpaw, "Decoding two-dimensional movement trajectories using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 4, no. 3, pp. 264–275, Sep. 2007.
- [27] G. Schalk, K. J. Miller, N. R. Anderson, J. A. Wilson, M. D. Smyth, J. G. Ojemann, D. W. Moran, J. R. Wolpaw, and E. C. Leuthardt, "Two-dimensional movement control using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 5, no. 1, pp. 75–84, Mar. 2008.
- [28] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain-computer interface," *Nature*, vol. 442, pp. 195–198, 2006.
- [29] K. Das and Z. Nenadic, "An efficient discriminant-based solution for small sample size problem," *Pattern Recognit.*, vol. 42, no. 5, pp. 857–866, 2009.
- [30] B. Obermaier, G. R. Müller, and G. Pfurtscheller, "Virtual keyboard" controlled by spontaneous EEG activity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 4, pp. 422–426, Dec. 2003.
- [31] H. Scherberger, M. R. Jarvis, and R. A. Andersen, "Cortical local field potential encodes movement intentions," *Neuron*, vol. 46, pp. 347–354, 2005.
- [32] L. Shpigelman, Y. Singer, R. Paz, and E. Vaadia, "Spikernels: Predicting arm movements by embedding population spike rate patterns in inner-product spaces," *Neural Comput.*, vol. 17, no. 3, pp. 671–690, 2005.
- [33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2001.
- [34] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 732–739, Jun. 2004.
- [35] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1394–1407, Aug. 2007.
- [36] G. Saon and M. Padmanabhan, "Minimum Bayes error feature selection for continuous speech recognition," in *Proc. NIPS*, 2000, pp. 800–806.
- [37] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, 1997.
- [38] J. C. Elashoff, R. M. Elashoff, and G. E. Goldman, "On the choice of variables in classification problems with dichotomous variables," *Biometrika*, vol. 54, no. 3, pp. 668–670, Dec. 1967.
- [39] G. Toussaint, "Note on optimal selection of independent binary-valued features for pattern recognition (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 618–618, Sep. 1971.
- [40] T. M. Cover, "The best two independent measurements are not the two best," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, no. 1, pp. 116–117, Jan. 1974.
- [41] J. M.-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.
- [42] H. Ramoser, J. Müller-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.
- [43] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

- [44] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000.
- [45] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, pp. 539–550, 2007.
- [46] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
- [47] J. Ye, T. Xiong, Q. Li, R. Janardan, J. Bi, V. Cherkassky, and C. Kambhampettu, "Efficient model selection for regularized linear discriminant analysis," in *Proc. CIKM*, 2006, pp. 532–539.
- [48] K. Das, S. Osechinskiy, and Z. Nenadic, "A classwise PCA-based recognition of neural data for brain-computer interfaces," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 6519–6522.
- [49] D. S. Rizzuto, A. N. Mamelak, W. W. Sutherling, I. Fineman, and R. A. Andersen, "Spatial selectivity in human ventrolateral prefrontal cortex," *Nat. Neurosci.*, vol. 8, pp. 415–417, 2005.
- [50] K. Das, J. Meyer, and Z. Nenadic, "Analysis of large-scale brain data for brain-computer interfaces," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 5731–5734.
- [51] H. Yu and H. Yang, "A direct LDA algorithm for high-dimensional data—With application to face recognition," *Pattern Recognit. Lett.*, vol. 34, no. 10, pp. 2067–2070, 2001.
- [52] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [53] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 14, Cambridge, MA: MIT Press, 2002.
- [54] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
- [55] Z. Nenadic and J. W. Burdick, "A control algorithm for autonomous optimization of extracellular recordings," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 941–955, May 2006.
- [56] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 550–556, Mar. 2007.
- [57] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*. New York: Oxford Univ. Press, 1995.
- [58] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–522, 1995.
- [59] N. Amador and I. Fried, "Single-neuron activity in the human supplementary motor area underlying preparation for action," *J. Neurosurg.*, vol. 100, no. 2, pp. 250–259, 2004.
- [60] L. Deecke, W. Lang, F. Uhl, R. Beisteiner, G. Lindinger, and R. Q. Cui, "Movement-related potentials and magnetic fields: New evidence for SMA activation leading MI activation prior to voluntary movement," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 50, pp. 386–401, 1999.
- [61] H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?," *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2341–2356, Nov. 2006.
- [62] A. Ikeda, H. O. Lüders, and H. Shibasaki, "Generation of contingent negative variation in the supplementary sensorimotor area," *Adv. Neurol.*, vol. 70, pp. 153–159, 1996.
- [63] B. Libet, C. A. Gleason, E. W. Wright, and D. K. Pearl, "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act," *Brain*, vol. 106, pp. 623–642, 1983.
- [64] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortical origins of response time variability during rapid discrimination of visual objects," *NeuroImage*, vol. 28, no. 2, pp. 342–353, Nov. 2005.
- [65] J. D. Wallis, "Orbitofrontal cortex and its contribution to decision-making," *Annu. Rev. Neurosci.*, vol. 30, pp. 31–56, 2007.
- [66] G. Schoenbaum, M. R. Roesch, and T. A. Stalnaker, "Orbitofrontal cortex, decision-making and drug addiction," *Trends Neurosci.*, vol. 29, no. 2, pp. 116–124, 2006.
- [67] J. J. Paton, M. A. Belova, S. E. Morrison, and C. D. Salzman, "The primate amygdala represents the positive and negative value of visual stimuli during learning," *Nature*, vol. 439, no. 7078, pp. 865–870, Feb. 2006.
- [68] Z. Nenadic, D. S. Rizzuto, R. A. Andersen, and J. W. Burdick, "Advances in cognitive neural prosthesis: Recognition of neural data with an information-theoretic objective," in *Toward Brain Computer Interfacing*, G. Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2007, ch. 11, pp. 175–190.
- [69] K. Das and Z. Nenadic, "Approximate information discriminant analysis: A computationally simple heteroscedastic feature extraction technique," *Pattern Recognit.*, vol. 41, no. 5, pp. 1548–1557, 2008.



Koel Das received the M.S. degree in electrical engineering from Wright State University, Dayton, OH, in 2003, and the Ph.D. degree in electrical engineering and computer science from the University of California, Irvine, in 2007.

She is currently a Research Scholar at the University of California, Santa Barbara. Her current research interests include brain-machine interfaces, pattern recognition, image processing, and geometric processing.



Daniel S. Rizzuto received the Ph.D. degree in neuroscience from Brandeis University, Waltham, MA, in 2002.

He was a Postdoctoral Fellow at California Institute of Technology, Pasadena. He is currently the Neuroscience Research Manager at the Swedish Neuroscience Institute, Seattle, WA, where he is developing novel therapies for epilepsy, stroke and psychiatric indications, among others.



Zoran Nenadic (M'06) received the Diploma in control engineering from the University of Belgrade, Belgrade, Serbia, in 1995, and the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1998 and 2001, respectively.

From 2001 to 2005, he was a Postdoctoral Scholar with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena. Since 2005, he has been with the Department of Biomedical Engineering, University of California, Irvine, where he is currently an Assistant Professor. His current research interests include adaptive biomedical signal processing, control algorithms for biomedical devices, brain-computer interfaces, and modeling and analysis of biological neural networks.

Dr. Nenadic is a member of the Mathematical Association of America and the Society for Neuroscience.