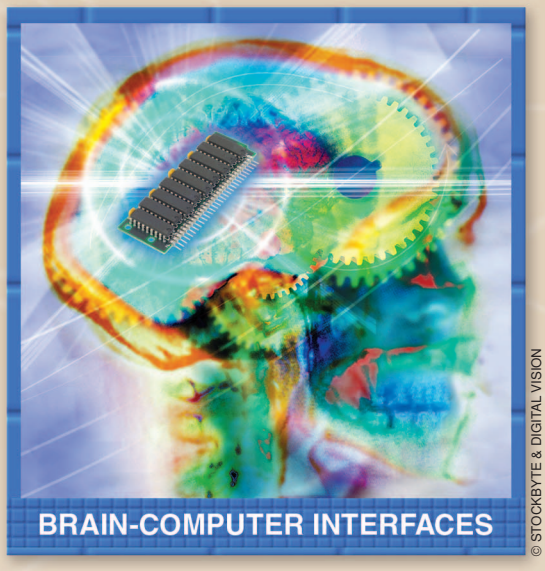
Lucas C. Parra, Christoforos Christoforou, Adam D. Gerson, Mads Dyrholm, An Luo, Mark Wagner, Marios G. Philiastides, and Paul Sajda



© STOCKBYTE & DIGITAL VISION

Spatiotemporal Linear Decoding of Brain State

Application to performance augmentation in high-throughput tasks

The conventional goal for a brain-computer interface has been to restore, for paralyzed individuals, a seamless interaction with the world. The shared vision in this research area is that one day patients will control a prosthetic device with signals originating directly from their brain. This review provides a new perspective on the brain-computer interface (BCI), by asking instead "How can BCI be used to assist neurologically healthy individuals in specifically demanding tasks?"

The limited signal-to-noise ratio (SNR) of noninvasive brain signals suggests that one must tailor the application of BCI to tasks where a small increment in information can make a large difference. High throughput tasks may provide such a scenario, as will be exemplified in this review for one such task: rapid visual target detection. BCI can assist in this task by prioritizing perceived target images.

Due to the speeded nature of this and related tasks, it is essential to use fast and effective signal processing. Effective performance is achieved by extracting spatiotemporal discriminant

Digital Object Identifier 10.1109/MSP.2007.909529

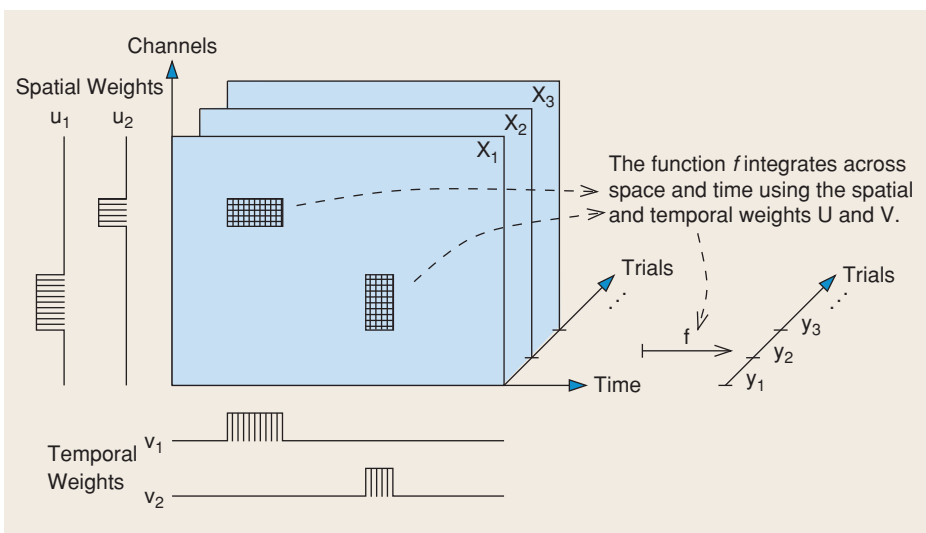
brain activity in high dimensions (typically ≥64 electrodes distributed across the skull surface, with sampling rates of 1,000 Hz). Fast processing speed is achieved by constraining the methods to linear analysis of these data. In fact, considering the low SNR of scalp electroencephalography (EEG) even linear features require careful regularization. This review summarizes linear spatiotemporal signal analysis methods that derive their power from careful consideration of spatial and temporal features of skull surface potentials.

INTRODUCTION

From an engineering point of view, the task of a BCI is to decode brain activity as reliably and as fast as possible. Specifically, the interface must identify neurophysiological activity that is associated with a subject's choices or decisions. This is essentially a signal transmission problem in which performance can be evaluated in terms of bit rate, that is, the number of binary decisions successfully communicated per unit time. It is no surprise, therefore, that a large fraction of BCI algorithms use binary classification between two alternative brain "states" as their starting point [10], [20].

Neuronal activity of the brain is reflected in changes of blood oxygenation, local field potentials, or scalp surface potentials. Blood oxygenation can be measured with functional MRI at very high spatial resolutions (≈mm). However, the sluggish response of hemodynamic regulation (>5 s) limits the effective bit rate of this modality [8], despite potentially high classification specificity. In contrast, electrical signals generated by neuronal discharge have very high temporal resolution (10 ms or less). Currently only surface electrodes (either placed on top of the scull or implanted chronically on the cortical surface) are viable on a long-term basis. For healthy subjects only scalp electrodes (scalp EEG) can be justified.

This review will provide an overview of linear classification algorithms for EEG and demonstrate applications of BCI to performance augmentation.



[FIG1] Representation of data space in which EEG data is collected. Linear decoding algorithms can be applied across different and multiple dimensions of this data-space.

AN OVERVIEW OF LINEAR ANALYSIS ALGORITHMS FOR EEG

TRADITIONAL EEG ANALYSIS

In EEG the SNR of individual signal channels is low, often at −20 dB or less. To overcome this limitation, all analysis methods perform some form of averaging, either across repeated trials, across time, or across electrodes. Traditional EEG analysis averages the signals of individual electrodes across many repeated trials. A conventional method is to directly average the measured potentials following stimulus presentation, thereby canceling uncorrelated noise that is not reproducible from one trial to the next. This averaged activity, called evoked response potentials (ERPs), captures activity that is time locked to the stimulus presentation while canceling evoked oscillatory activity that is not in phase with the stimulus. Alternatively, many studies compute the power of oscillatory activity in specific frequency bands by filtering and squaring the signal prior to averaging. In either case, the analysis is performed on individual channels and different experimental conditions are compared separately for each channel using conventional statistical tests such as Student's t-test or analysis of variance (ANOVA). These approaches are motivated by traditional concepts such as averaging, filtering, and univariate hypothesis testing. They do not, however, exploit the full spatiotemporal structure of the data.

Only in recent years have massively multivariate methods gained popularity in an effort to capture the full spatiotemporal dynamic of the EEG. BCI systems in particular have driven the development with the goal of extracting information from the EEG signal without having to average over trials.

LINEAR ERP ANALYSIS IN THE CONTEXT OF BCI

This review will focus on our analysis methods for stimulus evoked responses, i.e., EEG activity that is elicited by an external event such as the presentation of an image on a computer monitor. The cascade of neuronal processes that such a stimulus elicits might last just 1 s, yet, when recording this activity with a modern EEG system at 1 kHz with 100 electrodes, this provides a data matrix of 10^5 elements for each stimulus presentation.

Stimuli are often presented repeatedly in a sequence of trials leading to a large volume of data encompassing three dimensions: *trials*, *channels*, and *time* (see Figure 1). The goal of a BCI system is to identify the evoked activity for each event without averaging over trials. The nth trial is characterized by the data matrix X_n of dimensions *channels* × *time*, or $D \times T$. This

evoked response is to be classified as having been generated by one of two events. These may be associated with events such as target present versus no target present, perceived reaction error versus correct response, or imagined left movement versus imagined right movement. The challenge is to find a function that associates with the data matrix, $\mathbf{X}_n \in \Re^{D\times T}$, a scalar (possibly binary) value y_n that can be used to determine the class of the event

$$y_n = f(\mathbf{X}_n)\,. \tag{1}$$

This mapping represents a tremendous data reduction, e.g. from 10^5 dimensions to one. Without some form of constraint, the set of possible functions is intractable. The first constraint and simplification we propose is to consider only linear functions f. In its simplest form there are still $D \times N$ (10^5) free parameters to define this linear map. It may therefore be necessary to further constrain the number of degrees of freedom by exploiting available prior knowledge. For instance, we know that in the object recognition paradigm proposed later, images containing targets are rare. Such rare events are known to elicit a positive potential (relative to baseline activity) 300 ms after stimulus presentation. This activity, called P300, is broadly distributed over central and parietal electrodes. Given this information a reasonable linear mapping could simply be a weighted sum over appropriately chosen samples and electrodes:

$$y_n = f(\mathbf{X}_n) = \sum_{ij} W_{ji} X_{ijn} = Tr\{\mathbf{W}^T\mathbf{X}_n\}\,, \tag{2}$$

where W_{ij} are weights for the appropriate set of electrodes i and time indices j. Instead of fixing weights a priori one can also choose optimal weights based on example data. Optimal weights are those that best predict event class from value y for new (unseen) data $\mathbf{X}$ based on a set of training examples $\{\mathbf{X}_n, y_n; n = 1 \ldots N\}$. In the applications presented here a realistic number of such training examples may include as little as $N = 50$ trials. This small number of exemplars (as compared to $D \times T \approx 10^5$) certainly requires constraining the degrees of freedom of parameters $\mathbf{W}$ to avoid poor performance due to over-fitting.

SPATIAL AND TEMPORAL PROJECTIONS

The methods we have developed for this problem can be best summarized by considering the following factorization:

$$\mathbf{W} = \mathbf{U}\mathbf{V}^T = \sum_{k=1}^{K} \mathbf{u}_k \mathbf{v}_k^T\,. \tag{3}$$

We introduce this factorization because it allows us to reduce the degrees of freedom in $\mathbf{W}$ in a systematic fashion as the rank K is chosen to be less than $\min(D, T)$. In addition, this factorization can be interpreted as decomposing $\mathbf{W}$ into separate spatial and temporal projection vectors:

$$y_n = Tr\{\mathbf{U}^T\mathbf{X}_n\mathbf{V}\} = \sum_{k=1}^{K} \mathbf{u}_k^T\mathbf{X}_n\mathbf{v}_k\,. \tag{4}$$

The columns $\mathbf{u}_k$ of matrix $\mathbf{U}$ represent K linear projection in space and, similarly, each of the K columns $\mathbf{v}_k$ in matrix $\mathbf{V}$ represent linear projections in time.

With this interpretation it is now possible to express prior information on the spatial or temporal properties of the activity. For instance, one may use matrix $\mathbf{V}$ to select an appropriate time interval for the analysis by setting all coefficients outside this interval to zero. Matrix $\mathbf{V}$ can also be used to implement a filtering operation by specifying a corresponding Toeplitz structure; it could also be used to decompose the signal into its (time-) frequency components by using a Fourier or wavelet transformation matrix. In addition, the columns of matrix $\mathbf{U}$ can be used to select electrodes, or, if an anatomical model of the electrical conductivity of the head was available, one can specify in each column $\mathbf{u}_k$ a projection that recovers the kth current sources from the observed skull surface potentials (source space analysis). In this manner prior knowledge of the anatomical origins of neuronal currents can be naturally included into the analysis.

After these choices have been made based on prior information, the goal for machine learning is then to determine the remaining free parameters using a set of training examples. We have developed five methods of increasing complexity that fall within this framework.

1) Set $K = 1$ and chose the temporal component $\mathbf{v}$ to select a time window of interest (i.e., set $v_j = 1$ if j is inside the window of interest, and $v_j = 0$ otherwise). Learn the spatial weights of vector $\mathbf{u}$ from examples [5].

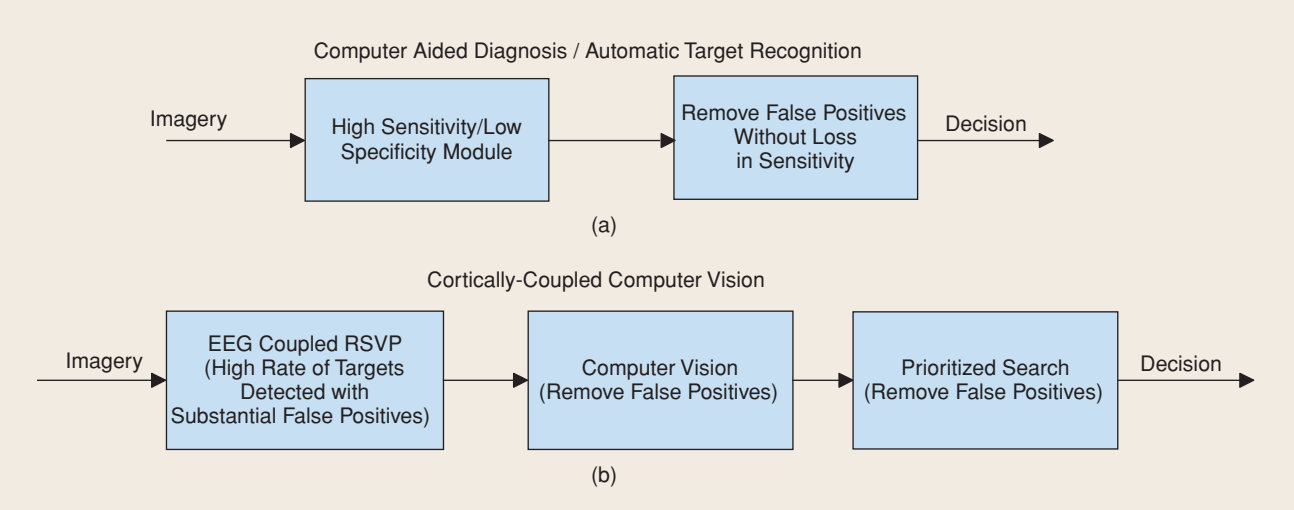
2) Select some $K > 1$ and chose the components vectors $\mathbf{v}_k$ to select multiple time windows of interest (as in 1). Learn each spatial weight vector $\mathbf{u}_k$ from examples separately and then combine these with weights $\mathbf{v}_k$ int o a single matrix $\mathbf{W}$ in a separate learning step (see Figure 4) [5].

3) Set $K = D$ while constraining $\mathbf{U}$ to be a diagonal matrix and select, separately for each channel, the time window $\mathbf{v}_k$ ($i = k$) which is most discriminative. Then train the diagonal terms of $\mathbf{U}$ resulting in a latency dependent spatial filter $\mathbf{W}$ [11]. Alternatively, in the first step, use feature selection to find the right set of time windows $\mathbf{v}_k$ simultaneously for all channels [12].

4) Set $K = 1$ and learn the spatial and temporal components $\mathbf{u}$ and $\mathbf{v}$ simultaneously. Constrain the solution to be smooth in both space and time [3].

5) Select some $K > 1$ and learn all columns of the spatial and temporal projection matrix $\mathbf{U}$ and $\mathbf{V}$ simultaneously. Constrain the solution to be smooth and resolve the inherent ambiguity of factorization (3) assuming that components k are independent across trials [4].

A number of (bi-)linear methods developed by other laboratories also fall within this general framework, e.g., [9], [19], [21], and [22]. The basic principles underling our methods will now be described, leaving the details of the specific algorithms to the original papers.



[FIG2] (a) Traditional computer aided diagnosis can be thought of as a two stage process where a number of targets are initially detected with a considerable amount of false positives. A subsequent step is then used to remove false positives without sacrificing the sensitivity of the system. (b) Cortically coupled computer vision on the other hand uses EEG signatures to initially identify images of potential interest while traditional computer vision techniques coupled with a prioritized image search (where EEG-selected images are at the top of the search list) are used to remove false positives.

LINEAR DISCRIMINATION ALGORITHMS

The learning problem as formulated above amounts to a conventional two-class linear discrimination problem with additional constraints on the structure of the coefficients $\mathbf{W}$. In the past we have used two standard approaches to linear discrimination: Fisher linear discriminants (FLDs) for its computational efficiency, and logistic regression (LR) for its robustness to outliers. In addition to their algorithmic differences, these methods differ in the optimality criterion they use to judge the quality of a classifier on a training set. LR maximizes the likelihood of correct classification assuming a general model distribution for y_n; FLD maximizes the difference between the mean value of y_n for the two classes while minimizing the variance about those means. Because of the (quadratic) variance estimate, FLD is more susceptible to outliers. LR, on the other hand, requires all data to be available at the time of learning which precludes a straightforward online implementation. The probabilistic formalism of LR is however particularly convenient when imposing additional statistical properties on the coefficient $\mathbf{W}$ such as smoothness or sparseness.

REGULARIZATION

In addition to the constraints on the structure of $\mathbf{W}$, we found it necessary to invoke additional regularization criteria to ensure good generalization performance for unseen data. This has been a common theme for BCI methods that use high-dimensional linear classification [19], [21], [22]. The maximum-likelihood formalism used in LR is particularly convenient as one can formulate standard prior probabilities for $\mathbf{U}$ and $\mathbf{V}$ to generate maximum a posteriori (MAP) estimates. We have used L2 and L1 regularization terms. L2 regularization results from assuming Gaussian process (GP) priors on spatial and temporal weight. With the choice of the covariance parameters of the GP, one can control the smoothness of the coefficient, essentially exploiting the prior knowledge that neighboring electrodes measure similar EEG activity (except for noise) and that the relevant time courses change on a slow time scale (as compared to the sampling rate). L1 normalization results from assuming Laplacian priors and encourages coefficients to be nonzero only if significant discriminant information is provided by the particular electrode and time sample. Thus, coefficients that are primarily driven by noise in the training data are suppressed.

The list of methods provided earlier indicates that we often use a two step approach: in the first step different sets of coefficients are trained on the data independently and separately. Then, in a second step, all coefficients are combined to a single overall linear classifier. This approach often shows better generalization performance as compared to training all coefficients as part of a joint optimization. The underlying assumption in this approach is that the discriminant information provided by the different sets of variables (different channels or different time windows) are independent. To see this consider for instance a FLD solution. The optimal FLD is determined by the covariance matrix of the training data. Training coefficients separately is equivalent to assuming a block-diagonal structure in this covariance matrix with zero off-diagonal blocks. By training coefficients separately, one is therefore effectively regularizing by assuming a priori the independence of the data across time-windows, components, or electrodes as the case may be.

FORWARD MODEL

As it turns out, the linear method allows recovering a model, which can be used to interpret the anatomical origin of the discriminative activity. That is, a "forward model" can be defined, which models the discriminative activity in the data as a linear spatial projection of a one-dimensional component onto the

surface electrodes [15]. For the factorized linear discriminant models, a one-dimensional forward model can be defined in the same manner ($K = 1$):

$$\mathbf{a} = \frac{\mathbf{R}\mathbf{u}}{\mathbf{u}^T\mathbf{R}\mathbf{u}}, \quad \text{with} \quad \mathbf{R} = \sum_n \mathbf{X}_n \mathbf{v}\mathbf{v}^T \mathbf{X}_n^T. \tag{5}$$

This forward model reflects the topographic distribution of activity associated with spatial projection $\mathbf{u}$ as it is observed on the sensors. In addition, a temporal forward model can be defined which can be thought of as an impulse response, i.e., the EEG temporal response evoked by the events

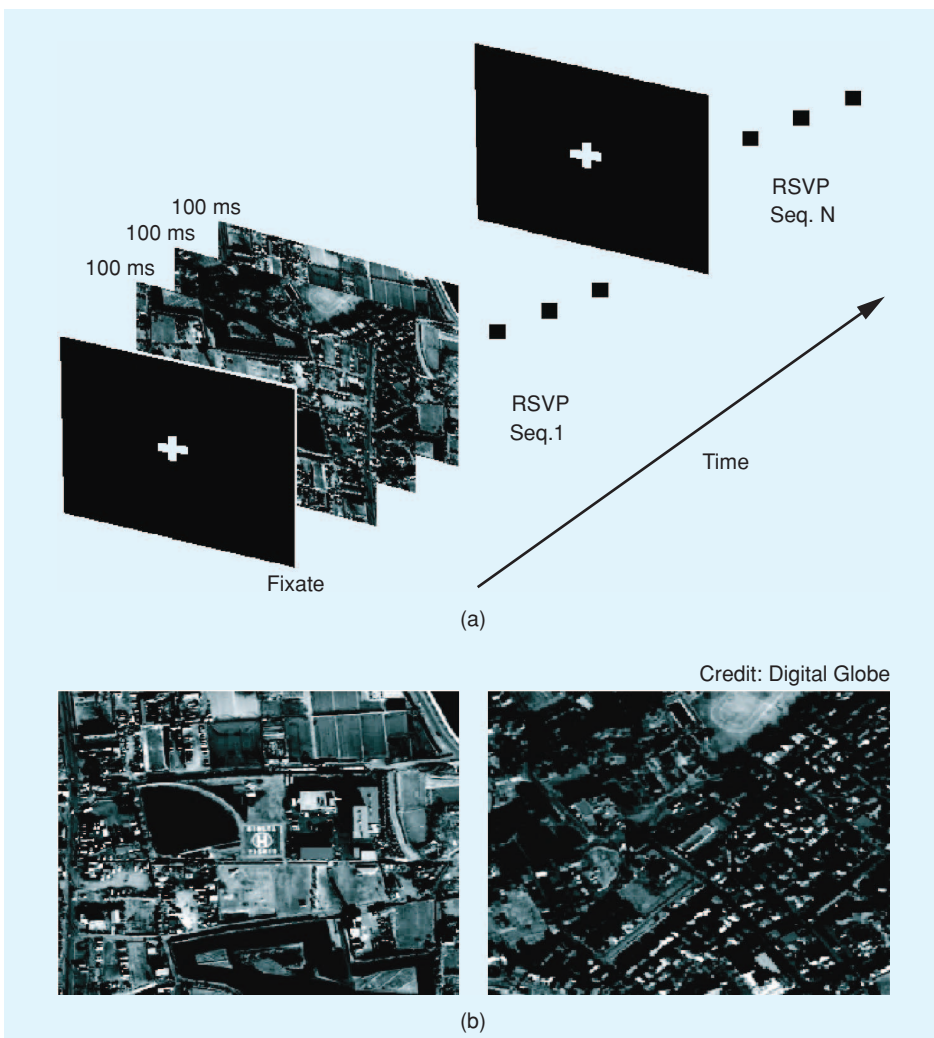
$$\mathbf{b} = \frac{\mathbf{v}^T\mathbf{R}}{\mathbf{v}^T\mathbf{R}\mathbf{v}}, \quad \text{with} \quad \mathbf{R} = \sum_n \mathbf{X}_n^T \mathbf{u}\mathbf{u}^T \mathbf{X}_n. \tag{6}$$

If multiple components are used ($K > 1$) discriminant components can be interpreted individually for each $\mathbf{u}_k$ and $\mathbf{v}_k$ after resolving a component ambiguity which is inherent in the factorization (3) [4].

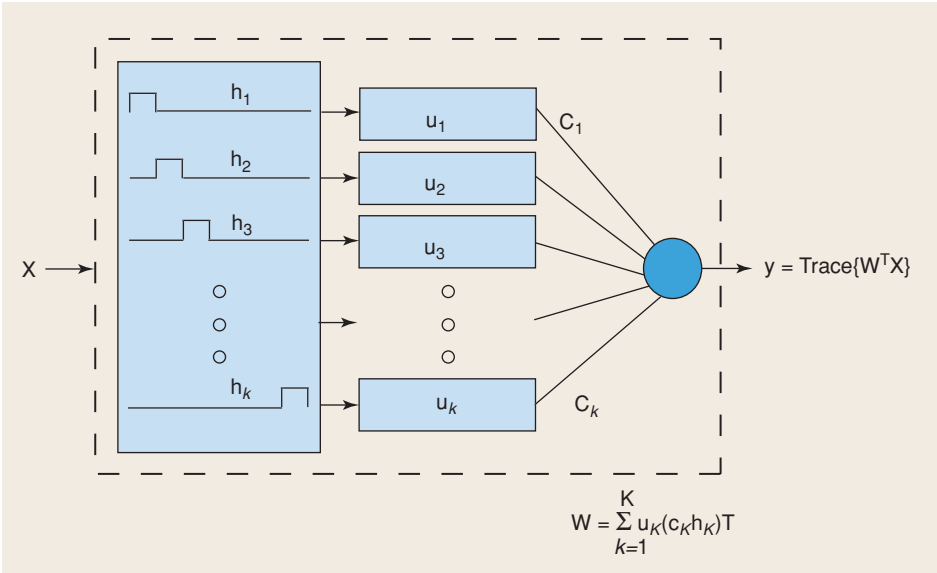
APPLICATIONS OF BCI TO HUMAN PERFORMANCE AUGMENTATION IN HIGH-THROUGHPUT TASKS

The currently limited SNR of scalp measured EEG dramatically limits the bit rates of BCI systems, with state-of-the-art systems below 50 b/min. Thus, practical applications must be designed with this limitation in mind. Those who are neurologically locked in can benefit from systems with such low bit rates, since there is no alternative means of communication, but for most other paralysis subjects there are alternative options with much higher bit rates (e.g., decoding EMG, EOG, eye tracking). However, there are scenarios in which BCI can be useful, even when a user is of normal neurological health. One example is a high-throughput search or decision task, when one is asked to rapidly make decisions or find objects of interest under severe time constraints. In such cases, one can potentially use BCI to drive throughput—i.e., increase the rate of target detection or decision making—by decoding EEG signals, in real time, that are indicative of a decision or an error. The goal would be to use this decoding to enable uninterrupted speeded responses.

Previously we have used this approach for two such tasks: speeded alternative choice [14] and rapid target detection [5]. In the case of speeded alternative force choice, subjects often notice that they make a mistake but only after the decision has occurred. BCI can assist a subject in this task by correcting, on-line, these perceived decision errors. In the case of visual target detection we find that subjects can perceive a target much quicker than their typical self-paced search speed. Performance can be dramatically improved by presenting images at a fast rate and later giving priority to the perceived target images. In both instances the BCI system leverages known electrophysiological correlates of perception, namely error related negativity (ERN), and the attention reorienting response (P300). Below we describe the BCI system we developed for rapid target detection, which we term "cortically coupled computer vision" [5].



[FIG3] (a) RSVP paradigm used in cortically coupled computer vision. A fixation cross first appears for 2 s, and it is followed by an RSVP image sequence where each image is presented for 100 ms (i.e, 10 Hz). An RSVP sequence typically lasts for 5 s (i.e., 50 consecutive images). (b) Examples of satellite image chips used for the RSVP task. A helipad target is shown on the left and a nontarget chip on the right.



[FIG4] Hierarchical discriminant component analysis. EEG signatures of target detection events are estimated using multiple linear discriminators each trained at a different time window relative to the image onset. A set of $k = 1 \ldots K$ nonoverlapping windows functions $\mathbf{h}_k$ are chosen and the discriminators u_k each associated with the corresponding time window are trained separately. At a second level, the discriminators $\mathbf{u}_k$ are fixed and coefficients c_k are learned. The resulting discriminant model takes the form of (4) where additional structure has been enforced on the temporal weights: $\mathbf{v}_k = \mathbf{h}_k c_k$.

CORTICALLY COUPLED COMPUTER VISION

Image and video repositories and databases are growing at a nearly exponential pace. An increasing problem is the efficient searching of these repositories/databases, given their size, diversity and potential sparsity of items of interest. Image search technologies have been a major focus in the the image and signal processing and machine learning communities, with methods including applications of traditional computer vision and image understanding to pattern analysis of image metadata [16], [17]. A challenge has been to develop new methods that enable the user to search in an efficient and robust way and retrieve imagery that is of particular interest to him/her.

HOW CAN BCI BE USED TO ASSIST NEUROLOGICALLY HEALTHY INDIVIDUALS IN SPECIFICALLY DEMANDING TASKS?

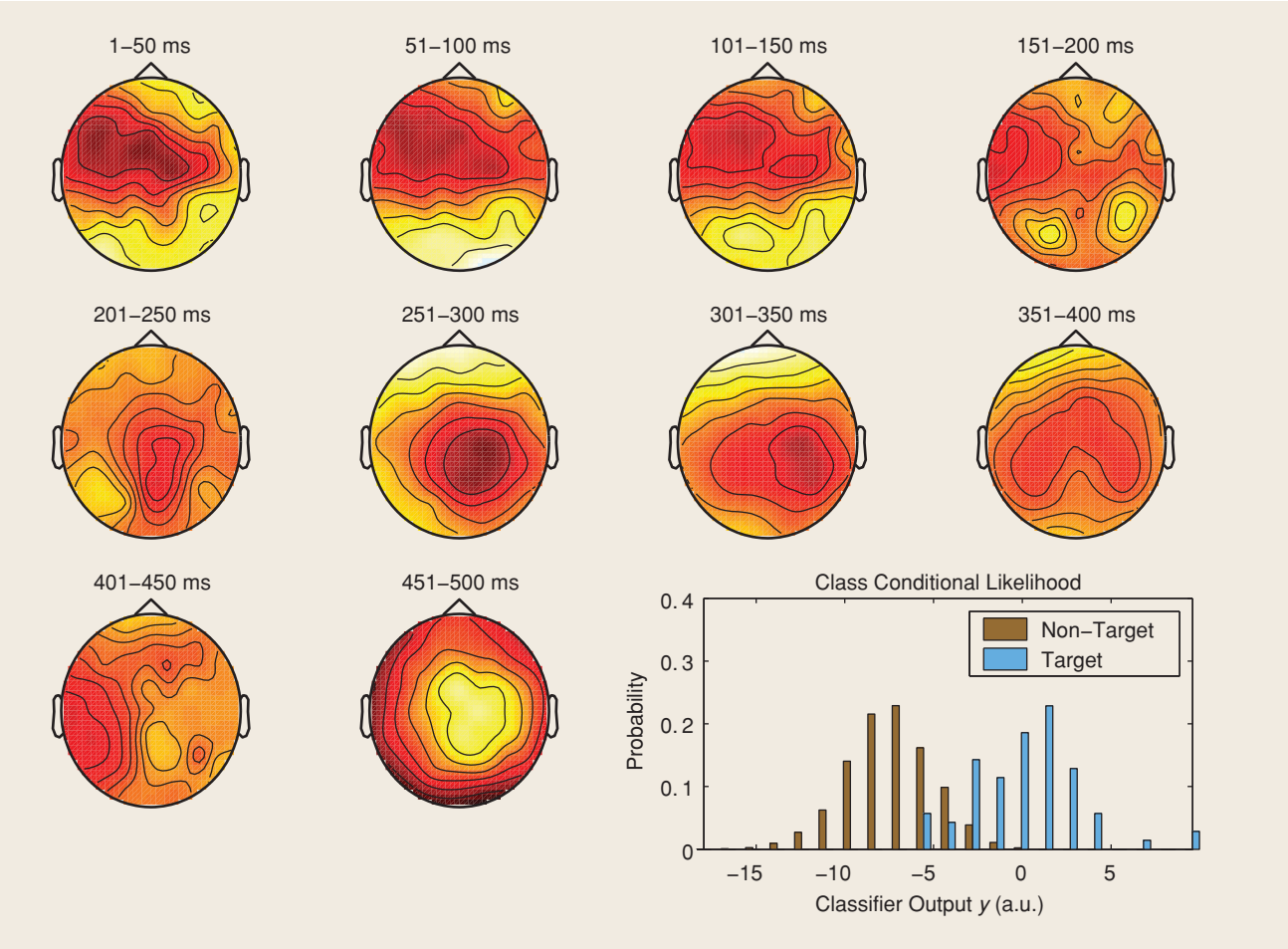
The human visual system is the most robust general-purpose visual processor that we know of; no computer vision system has been able to replicate its ability to construct complex visual scenes and recognize objects while maintaining invariances to illumination, pose, and depth. It is also well known that we are able to recognize objects in the blink of an eye [6], with various reports claiming recognition rates as fast as 100–150 ms post-stimulus [7], [18]. In addition, if one were asked to search for "interesting images" in a large database, results would depend on the individual, based on their experience, current mode of interest, etc.

Of course, having a human browse large databases is extremely cumbersome and time consuming, and it would likely yield very little utility since most images would never be seen. However, one could potentially leverage the general purpose and subjective nature of human search by creating a high-throughput version of the search task, where users browse imagery at an extremely fast rate (e.g., ten images per second) and then tag each image with a metric that represents their level of interest in what they see.

We have used our spatiotemporal linear decoding framework to develop a system for image triage, i.e., rapid search in a large database of images. We term the technology cortically coupled computer vision [5] since we potentially couple the decoding of cortical activity, measured via EEG, with more traditional computer vision and pattern recognition technology to jointly optimize, in terms of detection accuracy and time, image search (see Figure 2).

We can define the task as high throughput, namely the number of images to be searched is very large and time is limited. We have previously shown results for searching a large set of monochromatic natural images, where subjects were instructed to look for images with people in them [5]. Key is that we present the images very rapidly (10–20 images per second), in a rapid serial visual presentation (RSVP) paradigm and decode EEG online as the images are streaming by. The spatiotemporal linear filters thus decode EEG resulting from multiple overlapping (in time) stimulus events, to identify stimuli likely to be an image of interest. Below we describe a more realistic application of this approach, namely searching aerial images for objects of interest.

Trained image analysts (IAs) with experience in analyzing overhead imagery performed a search task for two different conditions, a baseline search and a search aided by cortically-coupled computer vision. In both scenarios IAs were instructed to identify helipads in monochromatic aerial images (see Figure 3). The images were commercial imagery (DigitalGlobe) and ranged in size from 28,000 × 28,000 to 30,000 × 30,000 pixels. They were instructed to mark each helipad they identified using a mouse click, which placed a red dot at the particular location in the image. During both searches, the number of helipads detected was continuously

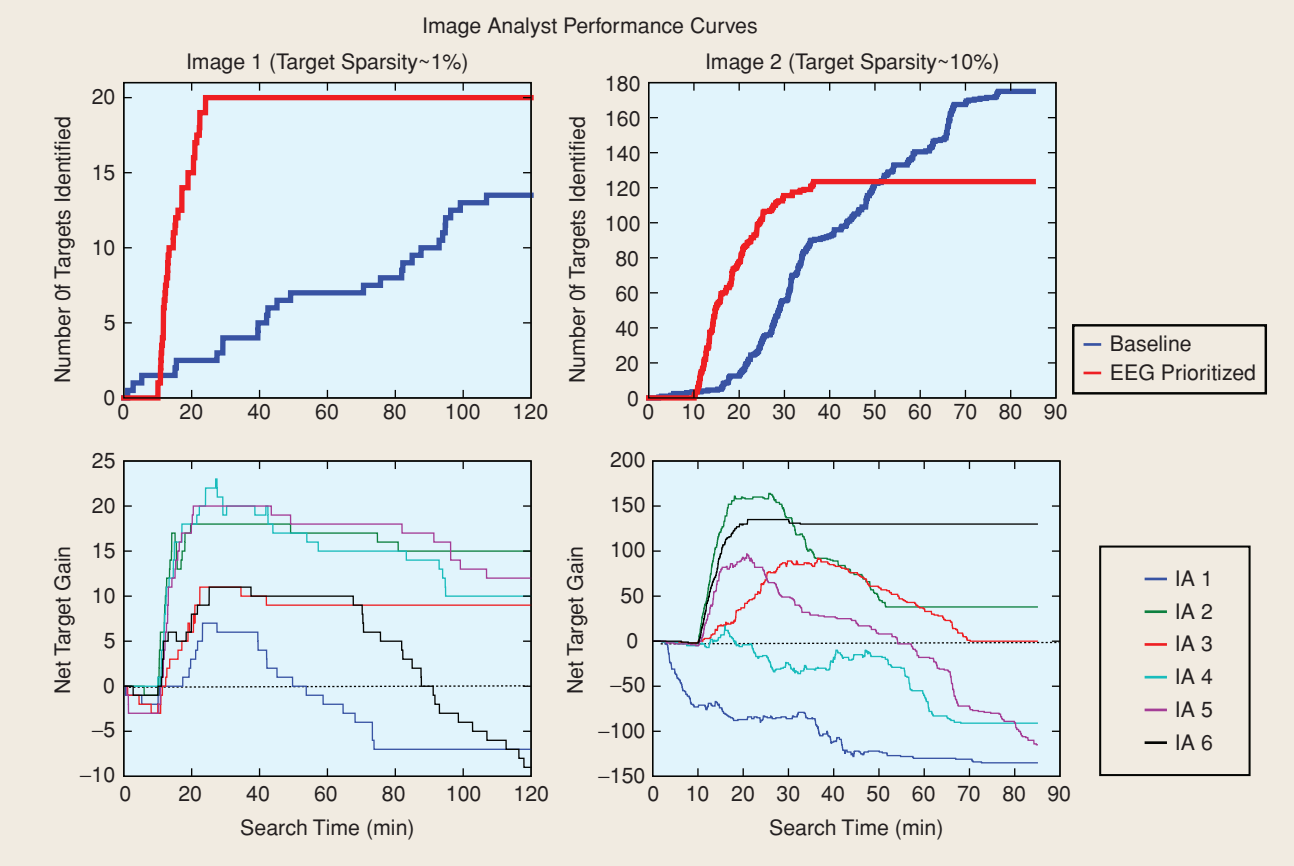


[FIG5] Scalp maps for one IA. The IA made no behavioral response (no button press) so all activity is purely decoded brain state unrelated to motor activity. Each scalp map represents the forward model $\mathbf{a}_k$ computed for the linear discriminator $\mathbf{u}_k$, each learned for a time window specified on the top of each scalp map (time relative to stimulus onset). The $K = 10$ time windows are linearly combined using the learned weightings c_k. The strongest weight here corresponds to a time window around 300 ms post-stimulus which shows a parietal scalp distribution. This is consistent with the hypothesis that the main contribution to discrimination is the single-trial correlate of the P300. The histograms of the discrimination signal y_n show that the linear classification, based on multiple windows, learns an integrated signal which can discriminate target from nontarget.

tracked as a function of search time. IAs searched until they decided to terminate the task.

In the baseline search, IAs could pan through the imagery at full resolution using the arrow keys on the keyboard, examining one 1,024×768 subimage at a time, or zoom out to 1/40 resolution using the 'z' key to view the entire image at once and their current location within it. Most IAs used a "mow-the-lawn" strategy, panning across rows or columns in a systematic way. A few used a more directed strategy, in which they utilized the low-resolution image to identify candidate regions on which to spend most time searching at full resolution.

In the cortically coupled computer vision condition, the IAs first performed an RSVP triage of the imagery, followed by a prioritized manual search of the imagery. 512 × 512 subimages were presented at 5 or 10 Hz, depending on preliminary experiments on the individual IAs, and EEG was recorded at 1,000 Hz sampling rate using a 64 electrode system (BioSemi) in a standard 10–20 montage. A linear classifier was used to decode 64 channels of EEG and assign a probability, to each image, that a target was present. Spatial weights $\mathbf{u}_k$ were trained independently for multiple windows ($K = 10$) covering 1 s of data following the onset of the image; these were then linearly combined to assign a target-probability for each single trial (image) as explained in Figure 4. (LR used all samples within each window as separate training exemplars and mean activity preceding the stimulus was subtracted for each trial. These measures make the resulting classifier robust to slow drifts and fast sample-by-sample fluctuations. The resulting classifier is shown in Figure 5.) IAs then performed the prioritized manual search at their own pace, however, the order of the chips was now determined by the EEG decoding. If decoding accuracy is high, images with targets would be examined first. IAs again marked, using the mouse, all helipads they saw, while we tracked their detection rate as a function of time. Figure 6 summarizes our results for six subjects for two different broad-area images.



[FIG6] Performance curves for six subjects (IAs 1-6) on two images (target rich on the right, and sparse targets on the left). Top panels show the number of targets found as a function of time (median across the six subjects) contrasting baseline performance (blue) versus cortically coupled computer-based vision system (red). Two graphs on the bottom show relative improvement for each subject as a function of time. Except for two out of 12 conditions, the assisted search shows significant gains within the first hour of search.

Clear is a consistent and substantial increase in the rate of target detection, which is a result of the triaging and re-prioritization. For the image with fewer targets (targets are more sparse) the relative difference between the cortically-couple computer vision and baseline search is most dramatic.

CONCLUSIONS

BCIs offer tremendous potential for improving the quality of life for those with severe neurological disabilities. At the same time, it is now possible to use noninvasive systems to improve performance for time-demanding tasks. Signal processing and machine learning are playing a fundamental role in enabling applications of BCI and in many respects, advances in signal processing and computation have helped to lead the way to real utility of noninvasive BCI.

The linear methods we present in this article are only a subset of signal processing and machine learning with potential applications. Attractive about our approach is the common framework of how we represent the data space and systematically decompose it spatially, temporally, or both simultaneously. A challenge continues to be how to best regularize these decompositions given the large parameter space and relative small number of examples. Future work will involve exploiting prior knowledge of both the signal characteristics as well as the task to improve the learning of the spatiotemporal filters and achieve high decoding accuracy.

ACKNOWLEDGMENT

This work was supported by DARPA and the National Geospatial Intelligence Agency/NGA under contract HM 1582-05-C-0043.

AUTHORS

Lucas C. Parra (parra@ccny.cuny.edu) received his Ph.D. in Physics from the Ludwig-Maximilian University in Germany in 1996. He is an associate professor of biomedical engineering at the City College of New York. Previously he was head of the adaptive image and signal processing group at Sarnoff Corporation and a member of the machine learning and the imaging departments at Siemens Corporate Research. His areas of expertise include machine learning, acoustic array processing, emission tomography, and electro-encephalography. His current research in biomedical signal processing focuses on functional brain imaging and computational models of the central nervous system.

Christoforos Christoforou (cchristoforou@gc.cuny.edu) is a Ph.D. candidate in computer science at the City University of New York. He received his B.S. and M.S. degrees in computer

science from Intercollege, Cyprus and City College, New York, respectively. He is a research assistant for the Neural Signals Laboratory of the City College and Pattern Recognition Lab of the Graduate Center. His research interests include pattern recognition methods, machine learning algorithms, bilinear structures, EEG analysis, and brain computer interfaces.

Adam D. Gerson (adg71@columbia.edu) received his B.S.E. in bioengineering from the University of Pennsylvania in 1997. He received his Ph.D. in biomedical engineering from Columbia University in 2006. While earning his doctorate, he worked at the Laboratory for Intelligent Imaging and Neural Computing (LIINC). His research interests include the application of machine learning to functional neuroimaging and the development of brain computer interfaces. He is in his second year of medical school at the University of South Florida.

Mads Dyrholm (md2581@columbia.edu) received his Ph.D. degree in 2006 from the Technical University of Denmark. He worked at City College for applied work in EEG pattern recognition. His current appointment is with Columbia University where his focus is multivariate analysis of fMRI and signal processing challenges in simultaneous EEG/fMRI.

An Luo (al2082@columbia.edu) is a Ph.D. candidate in the Department of Biomedical Engineering at Columbia University. She received her M.S. degree in 2002 from Chinese Academy of Sciences in pattern recognition and intelligent systems. She received her B.E. degree in electrical engineering from Tsinghua University in 1999. Her current research interest includes developing robust and adaptive methods for neuroimaging and applying them for cognitive neuroscience research.

Mark Wagner (mjwagner@fas.harvard.edu) will receive his B.A. and M.S. both in biomedical engineering from the School of Engineering and Applied Sciences at Harvard University in 2008. His research interests include single-trial detection in noninvasive neuroimaging as well as computational modeling of human feedforward motor learning and online feedback control.

Marios G. Philiastides (marios@mpib-berlin.mpg.de) received the B.S. degree in biomedical engineering from Rutgers University in 2001, the M.S. degree in electrical engineering from Stanford University in 2003, and the Ph.D. degree in biomedical engineering from Columbia University in 2007. His research focuses on neural engineering, neuroimaging of visual and cognitive function, machine learning applied to neural data and computational neural modeling. He has expertise in single-trial analysis of the EEG and event-related functional MRI.

Paul Sajda (ps629@columbia.edu) received his B.S. in electrical engineering from MIT in 1989 and his M.S. and Ph.D. in bioengineering from the University of Pennsylvania in 1992 and 1994, respectively. In 1994 he joined the David Sarnoff Research Center where he went on to become the Head of the Adaptive Image and Signal Processing Group. He is currently an associate professor of biomedical engineering and radiology at Columbia University where he is director of the Laboratory for Intelligent Imaging and Neural Computing (LIINC). His research focuses on neural engineering, neuroimaging, computational neural modeling and machine learning applied to image understanding. He is a Senior Member of the IEEE and associate editor for *IEEE Transactions on Biomedical Engineering*.

REFERENCES

[1] R. Bro, P. Geladi, and A.K. Smilde *Multi-way Analysis*. Hoboken, NJ: Wiley, 2004.

[2] M. Dyrholm, S. Makeig, and L.K. Hansen, "Model selection for convolutive ICA with an application to spatio-temporal Analysis of EEG," *Neural Computat.*, vol. 19, no. 4, pp. 934–955, 2007.

[3] M. Dyrholm and L.C. Parra, "Smooth bilinear classification of EEG," in *Proc. 28th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, New York, Aug. 2006.

[4] M. Dyrholm, C. Christoforos, and L.C. Parra, "Bilinear discriminant component analysis," *J. Mach. Learn. Res.*, vol. 8 pp. 1097–1111, May 2007.

[5] A.D. Gerson, L.C. Parra, and P. Sajda, "Cortically-coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[6] M. Gladwell, *Blink: The Power of Thinking Without Thinking*, 1st ed. New York: Little, Brown, and Co., 2005.

[7] C. Keysers, D.K. Xiao, P. Földiák, and D.I. Perrett, "The speed of sight," *J. Cognitive Neurosci.*, vol. 13, no. 1, pp. 90–101, 2001.

[8] S.G. Kim, W. Richter, and K. Ugurbil, "Limitations for temporal resolution in functional MRI," *Magn. Reson. Med.*, vol. 37, pp. 631–636, 1997.

[9] S. Lemm, B. Blankertz, G. Curio, and K.R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.,* vol. 52, no. 9, pp. 1541–1548, 2005.

[10] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain computer interfaces," *J. Neural Eng.* vol. 4, pp. R1–R13, 2007.

[11] A. Luo and P. Sajda, "Learning discrimination trajectories in EEG sensor space: Application to inferring task difficulty," *J. Neural Eng.*, vol. 3, no. 1, pp. L1–L6, 2006.

[12] A. Luo and P. Sajda, "Using single-trial EEG to estimate the timing of target onset during rapid serial visual presentation," in *Proc. Engineering in Medicine and Biology Society (EMBC2006)*, New York, 2006, pp. 79–82.

[13] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda, "Linear spatial integration for single trial detection in encephelography," *Neuroimage*, vol. 17, pp. 223–230, 2002.

[14] L. Parra, C. Spence, A. Gerson, and P. Sajda, "Response error correction—A demonstration of improved human-machine performance using real-time EEG monitoring," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no. 2, pp. 173–177, 2003.

[15] L.C. Parra, C.D. Spence, A.D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *Neuroimage*, vol. 28, no. 2, pp. 342–353, 2005.

[16] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *Int. J. Comput. Vis.*, vol. 18, no. 3, pp. 233–254, 1996.

[17] Y. Rui and T.S. Huang, " Image retrieval: Current techniques, promising directions, and open issues," *J. Vis. Commun. Image Represent.*, vol. 10, pp. 39–62, 1999.

[18] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520–522, 1996.

[19] R. Tomioka and K. Aihara, "Classifying matrices with a spectral regularization," in *Proc. 24th Annu. Int.Conf. Machine Learning (ICML2007)*, Oregon, June 2007.

[20] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, and T.M. Vaughan, "Brain computer interface technology: A review of the first international meeting," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[21] B. Blankertz, G. Dornhege, C. Schafer, R. Krepki, J. Kohlmorgen, K.-R. Muller, V. Kunzmann, F. Losch, and F.G. Curio, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no. 2, pp. 127–131, 2003.

[22] R. Tomioka, K. Aihara, and K.-R. Müller, "Logistic regression forsingle trial EEG classification," *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, T. Hoffman, Eds. Cambridge, MA: MIT Press, pp. 1377–1384, 2007.

SP