ELSEVIER

Journal of Neuroscience Methods 167 (2008) 51–62

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Brain–computer interfaces (BCIs): Detection instead of classification

G. Schalk [a,b,*], P. Brunner [a,c], L.A. Gerhardt [b], H. Bischof [c], J.R. Wolpaw [a,d]

[a] *Brain-Computer Interface Research and Development Program, Wadsworth Center, New York State Department of Health, Albany, NY, USA*
[b] *Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA*
[c] *Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria*
[d] *School of Public Health, State University of New York, Albany, NY, USA*

Received 9 January 2007; received in revised form 14 August 2007; accepted 15 August 2007

Abstract

Many studies over the past two decades have shown that people can use brain signals to convey their intent to a computer through brain–computer interfaces (BCIs). These devices operate by recording signals from the brain and translating these signals into device commands. They can be used by people who are severely paralyzed to communicate without any use of muscle activity. One of the major impediments in translating this novel technology into clinical applications is the current requirement for preliminary analyses to identify the brain signal features best suited for communication. This paper introduces and validates signal detection, which does not require such analysis procedures, as a new concept in BCI signal processing. This detection concept is realized with Gaussian mixture models (GMMs) that are used to model resting brain activity so that any change in relevant brain signals can be detected. It is implemented in a package called SIGFRIED (SIGnal modeling For Real-time Identification and Event Detection). The results indicate that SIGFRIED produces results that are within the range of those achieved using a common analysis strategy that requires preliminary identification of signal features. They indicate that such laborious analysis procedures could be replaced by merely recording brain signals during rest. In summary, this paper demonstrates how SIGFRIED could be used to overcome one of the present impediments to translation of laboratory BCI demonstrations into clinically practical applications.

Keywords: Brain–computer interface (BCI); Electroencephalography (EEG); Electrocorticography (ECoG); Augmentative communication; Brain–machine interface; Gaussian mixture model; Background modeling; SIGFRIED

1. Introduction

1.1. Brain–computer interface (BCI) technology

Many people with severe motor disabilities need augmentative communication technology. Those who are totally paralyzed, or "locked-in," cannot use conventional augmentative technologies, all of which require some measure of muscle control. Over the past two decades, various studies have evaluated the possibility that brain signals recorded from the scalp or from within the brain could provide new augmentative technology that does not require muscle control (e.g., Birbaumer et al., 1999; Farwell and Donchin, 1988; Hochberg et al., 2006; Kennedy et al., 2000; Kübler et al., 1999, 2005; McFarland et al., 1993; Müller and Blankertz, 2006; Pfurtscheller et al., 1993; Serruya et al., 2002; Sutter, 1992; Taylor et al., 2002; Vaughan et al., 2006; Wessberg et al., 2000; Wolpaw and McFarland, 2004; Wolpaw et al., 1991, 2002).

These brain–computer interface (BCI) systems measure specific features of brain activity and translate them into device control signals. Thus, a BCI system derives and utilizes control signals to effect the user's intent, and it usually does so by allowing the user to make a selection. This selection capacity is often realized using a computer cursor (e.g., Hochberg et al., 2006; Wolpaw and McFarland, 2004), but also in other ways such as controlling an arrow on a dial (Müller and Blankertz, 2006), a moving robot (Millán et al., 2004), or controlling other external devices (Donoghue et al., 2007; Pfurtscheller et al., 2000). Key performance characteristics of BCI systems are speed (i.e., how long it takes to make a selection) and precision (i.e., how often the executed selection is the one the user intended). Current systems allow for one selection within several seconds at a relatively high accuracy (e.g., 90% accuracy in a binary task). Expressed in bit rate, which combines both speed and accuracy,

* Corresponding author.
E-mail address: schalk@wadsworth.org (G. Schalk).

0165-0270/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.jneumeth.2007.08.010

the sustained performance of typical non-invasive and invasive BCI systems is still modest (i.e., between 5 and 25 bits/min; Wolpaw et al., 2002).

These studies show that non-muscular communication and control is possible and imply that BCI technologies can serve useful purposes for those who cannot use conventional technologies.

1.2. The signal identification problem

While these laboratory studies represent impressive technical demonstrations of a novel way to communicate, significant impediments remain that make it difficult to translate this technology from the laboratory to clinical practice. One of the major issues is that creation and successful operation of a BCI device depend on identification of those brain signal features and locations that are best suited for communication. This initial and continual identification is necessary because these features and locations are usually subject-specific and may also change over time.

In addition, there is essentially no theoretical basis (and thus no mathematical model) for the choice of signals for communication. All current BCI systems are based on experimental observations that particular mental tasks (such as imagining hand movements) have particular effects on specific brain signals (such as the mu rhythm measured at a particular location). Even with this empirical evidence, the choice of signals and tasks is difficult and is likely to be suboptimal for communication such that a completely different signal and task might provide better performance. Furthermore, the choice of location, feature (e.g., frequency), and task has to be optimized for each individual and for each brain signal (e.g., the mu rhythm). The difficulty of choosing signals and tasks, and the reliance on technical experts that this currently implies, can be regarded as the signal identification problem for BCI communication.

Approaches to translate BCI signal features into device control signals typically use classification/regression procedures. For example, studies reported in the literature have used linear discriminant analysis (Babiloni et al., 2000); neural networks (Huan and Palaniappan, 2004; Pfurtscheller et al., 1997); support vector machines (Garrett et al., 2003; Gysels et al., 2005; Lal et al., 2004; Müller et al., 2003); linear regression (McFarland et al., 1993, 1997a). These procedures are summarized in a recent review article on BCI feature extraction and translation methods (McFarland et al., 2006). This article lists 12 different methods for BCI feature translation and cites 26 corresponding articles. All these articles describe different realizations of classification or regression procedures. Most important, all these approaches require that the task, location, and brain signal features be optimized for each user by collecting and analyzing a comprehensive body of preliminary data. This requirement exists because the typical classification approach assumes that the brain produces a number of discrete states that are communicated using one or more signals, and the BCI problem is framed so as to detect these states.

In summary, current BCI signal processing techniques are all similar to the feature extraction/classification approaches commonly found in communication systems, which typically demand a detailed understanding and thus a mathematical model of the transmitted signal. This has persisted despite the fact that there is little *a priori* information about the signal best suited for BCI communication.

1.3. The signal identification paradox

As described above, current BCI signal processing approaches focus on extracting one specific set of brain signal features that is modulated by one particular mental task, i.e., on the creation of a highly specific filter that provides useful results only if the user appropriately modulates that particular set of signal features. Furthermore, they require data for all relevant tasks (i.e., classes) so that they can create this filter. Thus, all current methods require an initial procedure that identifies in each user out of all possible tasks those that best modulate one or some of the many possible signal features, and the definition of a filter that extracts this particular brain signal. This characteristic, i.e., that both features and tasks need to be identified, differentiates the BCI signal identification problem from the typical classification problem in which the different data classes are defined *a priori* and only the features need to be appropriately selected and combined.

Unfortunately, there is no strong *a priori* basis for selecting tasks and features. Hence, the number of possible choices of features and tasks increases in parallel with increasing signal fidelity. In addition to potential performance benefits of better signal fidelity, this issue also implies that the necessary signal identification procedure becomes increasingly complex and impractical. Signal identification procedures become more complex with better signal recordings because the number of tasks that might modulate brain signals increases, and because, with current methods, subjects need to produce all of these mental tasks. They also become impractical, because it becomes progressively more difficult for subjects to consistently produce a particular mental task (e.g., to imagine a particular type of movement) as the specificity of the task, and thus the specificity of the brain signal features related to production of that task, increases. Furthermore, when subjects are asked to actually use the brain signal that was identified in the initial identification procedure for a different purpose (i.e., to control an output device), these brain signals (e.g., amplitudes at particular frequencies and locations) may change in response to the different task of controlling the output device rather than producing a particular mental task. Moreover, as the user adapts to the new task, brain signals change further. It can be difficult to adapt the function that translates signal features into control signals to these signal changes, in particular for complex methods (such as neural networks or support vector machines). Thus it is possible and with increasing signal specificity increasingly likely that these different signal features will not be optimally translated into device output signals even though the subject exerts good control over some aspects of detectable brain activity. In this scenario, which is compounded as the feature/task space increases with better signal recording, BCI performance may degrade.

In summary, these properties of current BCI methods suggest that better recording and feature extraction will not necessarily continue to improve performance. In addition, the fact that current BCI systems depend on careful initial and continual adjustments by BCI experts, impedes widespread application of BCI technology to serve the communication and control needs of people with severe disabilities.

1.4. A novel approach to BCI signal processing

In the work described in this paper, we present a method that overcomes some of the issues described above. We accomplish this by showing how BCI control signals can be effectively extracted after only a reference dataset has been collected. These control signals are extracted by using a set of appropriate signal features to comprehensively characterize brain signals during rest and then to determine, at each point in time, by how much the current set of signal features differs from those in the reference dataset. Assuming that the user can modulate one or more of the features, he/she can use this difference for BCI control. This implies that our approach translates brain signals into graded one-dimensional output control signals (where the graded capacity can then be used by a specific BCI application to select different functions) and does not directly classify *brain states* into different discrete *classes*. In other words, the approach here is to detect any change that occurs in an appropriate set of brain signal features rather than to detect a specific set of changes in specific features. Because this approach relies on statistics derived from only one class (e.g., rest), it does not guarantee to provide superior performance to methods that utilize information from all possible classes (e.g., LDA, SVMs, neural networks, common spatial patterns, etc.). However, the theoretical benefit from using all possible classes is in practice countered by the often substantial intra- and inter-subject differences. Thus, the benefit of the approach presented here is simply that signal collection from all classes is laborious and impractical (and increasingly so for recordings with a larger feature/task space) and that it is likely to be less sensitive to intra- and inter-subject variations in brain activity patterns.

The paper describes a comprehensive evaluation of this novel approach which is implemented here in a software package called SIGFRIED (SIGnal modeling For Real-time Identification and Event Detection).[1] We evaluate its efficacy in the BCI context by determining whether it can be used to effectively discriminate the data collected during two different tasks. To do this, we sought to answer four questions: (1) which parameters minimize BCI error rates calculated offline; (2) whether results achieved using SIGFRIED are comparable to results achieved using a conventional method (i.e., linear regression) that relies on prior analysis of data from at least two classes and thus on a signal identification procedure; (3) whether signals produced by SIGFRIED have characteristics that make them amenable to real-time feedback; (4) whether a SIGFRIED model generated for data collected during rest (as opposed to during one of the classes in a real-time BCI experiment, which requires a trained subject) can be used subsequently to produce discriminating output for data collected during various tasks.

[1] This package is available for use by others as part of the general-purpose BCI platform BCI2000 (Schalk et al., 2004; http://www.bci2000.org).

2. Methods

In this paper, we used data that were collected during online BCI experiments and during a typical signal identification procedure, and analyzed them offline as described later in this section.

2.1. BCI data collection

The BCI data were collected from three male adults (subjects A–C, ages 29–40). Two had no disability. One had a spinal injury at the level of T7 and was confined to a wheelchair. All gave informed consent for the study, which had been reviewed and approved by the New York State Department of Health Institutional Review Board. After an initial evaluation defined the frequencies and scalp locations of each subject's mu and beta rhythm activity, the subject learned EEG-based cursor control over several months. Thus, the subjects had extensive experience with BCI operation prior to data collection for this study.

Each subject sat in a reclining chair facing a video screen and was asked to remain motionless during performance. Scalp electrodes recorded 64 channels of EEG (Sharbrough et al., 1991), each referred to an electrode on the right ear (amplification 20,000; band-pass 0.1–60 Hz). The BCI2000 software system (Schalk et al., 2004) was used to acquire signals from 64 channels (digitized at 160 Hz) and to implement a one-dimensional four-target BCI experiment as described below. Data from these experiments were originally used in the 2001 and 2003 NIPS data competition (Blanchard and Blankertz, 2004; Blankertz, 2003; Blankertz et al., 2004) and were the basis for the offline analyses in the present paper.

The subjects used mu or beta rhythm amplitude (i.e., subject-specific frequencies in the 8–12 or 18–24 Hz band, respectively) to control vertical cursor movement toward the vertical position of a target located at one of four evenly spaced positions at the right edge of the video screen. Data were collected from each subject for 8–10 sessions. Each session consisted of six runs of 180–182 s each, separated by 1 min breaks, and each run consisted of 32–35 individual trials. The total number of trials was 1681–1920. Each trial began with a 1.1 s period during which the screen was blank. Then, the target appeared at one of four possible positions (i.e., top, up, down, bottom) on the right edge. After the target was on the screen for 1.1 s, a cursor appeared at the middle of the left edge of the screen and started traveling across the screen from left to right at a constant speed (i.e., movement period). Its vertical position was controlled by the subject's EEG as described below. The subject's goal was to move the cursor to the height of the target. The movement period lasted between 1.9 and 2.3 s for the different subjects. When the cursor reached the right edge, the selected target flashed for 1.1 s. Then, the screen went blank for 1.1 s. This event signaled the end of the trial.

Cursor movement was controlled as follows. Ten times per second, the last 200 ms of digitized EEG from one to two channels over sensorimotor cortex was re-referenced using a large Laplacian spatial filter (McFarland et al., 1997b) and then submitted to frequency analysis by an autoregressive algorithm (McFarland et al., 1997a) to estimate the amplitude (i.e., the square root of power) in a mu and/or beta rhythm frequency band. The amplitudes for the one to two channels at frequencies centered at 12 or 13.5 Hz (3 Hz bandwidth) were combined to give a control signal that was used as the independent variable in a linear equation that controlled vertical cursor movement. Electrode position and center frequency remained constant for a particular subject (see Appendix A for further details.)

In summary, these online BCI experiments produced four classes of data (i.e., one for each target) in which three subjects used mu or beta rhythms at particular scalp locations to move a computer cursor in one dimension. Unless otherwise noted, the analyses described in the following sections are confined to data from two of the four classes in these experiments (i.e., those corresponding to top and bottom targets).

2.2. Additional data collection

In additional data collection, we also recorded data from 64 channels from four additional subjects (D–G) who were naive to BCI control. First, we collected 6 min of data from each subject during rest, i.e., while the subject was relaxed and not actively engaged in motor performance. We subsequently recorded 12 min of data while the subjects performed different motor tasks (i.e., moving both hands, both feet, or the left or right fist). Performance was visually cued and randomly alternated among the four tasks (each lasting 4.05 s), where each task period was followed by rest (4.05 s). The 88 rest periods between motor performance were not analyzed. There were 22 trials for each of the tasks. These motor tasks, and their imagined counterparts, are characteristic of those usually employed in BCI signal identification procedures.

2.3. Feature extraction

A number of studies (e.g., McFarland et al., 1997b) have shown that spatial filtering improves signal quality in EEG-based BCI systems. We applied one realization of such a filter, the common average reference (CAR), to the signal $s_h(k)$ from electrode h (where h referred to location C_3 (subject A) or CP_3 (subjects B and C)) at each time point k:

$$s'(k) = s_h(k) - \frac{1}{H}\sum_{i=1}^{H} s_i(k) \tag{1}$$

For the BCI data collected during the 1.9–2.3 s movement period (i.e., while the subjects actively modulated mu/beta rhythm amplitude), we converted the CAR-filtered time-domain signal $s'(k)$ into frequency domain amplitudes $a(n)$. We did this by submitting windows of 400 ms (no overlap) to the maximum entropy spectral estimation technique (Marple, 1987; Priestley, 1981), which is based on an autoregressive model. This model of order p ($s'(k) + c_1 s'(k-1) + \cdots + c_p s'(k-p) = \epsilon_p$) is defined by its coefficients c_1 to c_p which are estimated using the Burg method (Burg, 1967, 1968). These coefficients are converted into the amplitude spectrum $A(f)$ using:

$$A(f) = \frac{\sqrt{\epsilon_p}}{\left|1 - \sum_{k=1}^{p} c_k \, \mathrm{e}^{-\mathrm{j}2\pi k(f/f_s)}\right|} \tag{2}$$

Finally, we first discretized (step size 0.2 Hz) and then averaged these amplitudes $A(f)$ in 10 frequency bins from 10 to 30 Hz (bandwidth was 2 Hz unless otherwise noted). This resulted in 10 frequency domain features $a_f(n)$ that we combined, for each sample n, into a feature vector $\vec{a}(n) = [a_1(n), \ldots, a_{10}(n)]$. This procedure resulted in four to five feature samples per trial for each of the 1681–1920 trials (i.e., a total of 6720–9600 feature samples per subject).

2.4. Signal detection

To translate these feature samples into control signals, the characteristics of feature samples during one class are first modeled. For each new feature sample, the difference to the established model is derived. This difference yields a control signal that may drive an output device. This procedure, which combines existing modeling techniques with automatic model selection, is described in this section and in Sections 2.5 and 2.6. It is implemented in the SIGFRIED package that is made available with BCI2000.

A number of mathematical techniques could be used to describe aspects of signal distributions that could be used for signal detection. These include one or more Gaussian distributions (i.e., Gaussian mixture models (GMMs); Stauffer and Grimson, 1999), k-Nearest Neighbors (kNN) (Cover and Hart, 1967; Duda et al., 2001), Parzen Windows (Parzen, 1962), Radial Basis Functions (Cacoullos, 1966; Haykin, 1998), or one-class Support Vector Machines (SVMs) (Schölkopf et al., 2001). While detection approaches based on such techniques have been used extensively in other domains such as image processing (e.g., Pless, 2003; Radke et al., 2005; Stauffer and Grimson, 1999), they have not been used in BCI processing.

Results in image processing applications (Pless, 2003) suggested a Gaussian model and indicated that multiple Gaussian distributions (i.e., GMMs) may improve performance over one Gaussian distribution. Thus, we chose GMMs (Stauffer and Grimson, 1999) to implement our detection approach for BCI processing.

A Gaussian distribution c can be described by its mean $\vec{\mu}_c$ and its covariance matrix Σ_c. Introducing the prior probability (i.e., weight) ω_c as the proportion of data points assigned to this Gaussian distribution, the probability density function can be defined (Duda et al., 2001) and weighted with ω_c:

$$p(\vec{a}(n)|c) = \frac{\omega_c}{(2\pi)^{D/2}|\Sigma_c|^{1/2}} \, \mathrm{e}^{(-(\vec{a}(n)-\vec{\mu}_c)^T \Sigma_c^{-1}(\vec{a}(n)-\vec{\mu}_c))/2} \tag{3}$$

where D is the number of dimensions in the feature vector. Σ_c was calculated as the covariance matrix of samples assigned to

cluster c and ω_c was calculated as the proportion of the number of samples N_c assigned to that cluster compared to the total number of samples N.

The number of free parameters for each of C Gaussian distributions (i.e., clusters) is the sum of parameters in the covariance matrix Σ_c, the cluster mean $\vec{\mu}_c$, and its weight ω_c. Thus, the total number of parameters N_p in the signal model can be determined as follows:

$$N_p = \left[\frac{D(D+1)}{2} + D + 1\right] C \quad (4)$$

2.5. Parameterization

Gaussian mixture models (GMMs) can be used to approximate signal distributions and can thereby form the basis for the BCI detection approach proposed in this paper by modeling the probability density function of signals recorded during one condition. Four sets of variables must be determined to define a GMM (C, the number of clusters); and for each cluster c, $\vec{\mu}_c$, Σ_c, ω_c, the feature mean, covariance matrix, and prior probability, respectively.

To determine those $\vec{\mu}_c$, Σ_c, ω_c that produce an effective approximation of the original sample distribution given the number of clusters C, various algorithms have been proposed. One of the best known of these is the Expectation Maximization (EM) algorithm (Dempster et al., 1977). Each iteration of the EM algorithm consists of two steps: an Estimation (E) step and a Maximization (M) step. The M step maximizes a likelihood function that is redefined in each iteration by the E step. The Competitive EM (CEM) algorithm (Biernacki et al., 2003; Celeux and Govaert, 1992) improves on the EM algorithm mainly with regard to speed of convergence (and thus execution speed). It does this by assigning, in a C step, each sample to the most likely cluster, whereas the EM algorithm requires accumulation of fractional statistics. Because speed of execution is an important criterion for our application, we chose the CEM algorithm over the EM algorithm.

Each of these two algorithms requires definition of the number of Gaussian clusters C, which is typically done manually. This was not acceptable because our goal was to minimize the expert oversight required. Optimization of the number of clusters is not a straightforward problem, because the approximation of the GMM model to the source distribution improves with the number of clusters. In the end, the source distribution would be most accurately described by a model with as many Gaussian mixtures as data points. This is not resource efficient and may not provide the best level of generalization. Thus, we considered penalizing the performance of the model by a factor that reflected the complexity of the model. In the literature, such terms are generally referred to as information criteria and are typically in the form of Eq. (5):

$$\text{information criterion} = \text{measure of fit} + \text{complexity penalty} \quad (5)$$

A number of these information criteria have been described in the literature (e.g., the Akaike Information Criterion (AIC): Akaike, 1973; Vapnik's Structural Risk Minimization (SRM): Vapnik and Chervonenkis, 1974; Schwarz's Bayesian Information Criterion (BIC): Schwarz, 1978; Rissanen's Minimum Description Length (MDL) and Shortest Data Description (SSD): Rissanen, 1978 and Bozdogan's Corrected Akaike Information Criterion (CAICF) and Consistent Akaike Information Criterion: Bozdogan, 1974) (see Torr, 1997 for a comprehensive review).

Because there was no theoretical basis for selecting one of these criteria over the others, we arbitrarily picked the Akaike Information Criterion.[2] This criterion (K_{AIC} in Eq. (6)) is not influenced by the number of observations N (Akaike, 1973):

$$K_{\text{AIC}} = -2L + 2N_p \quad (6)$$

where the maximized likelihood L is defined as

$$L = \sum_{n=1}^{N} \log(p(\vec{a}(n)|c_{\text{best}}(n))) \quad (7)$$

N is the total number of samples and $c_{\text{best}}(n)$ is derived as

$$c_{\text{best}}(n) = \underset{c_i \in \mathbb{N}}{\text{argmax}}\, p(\vec{a}(n)|c_i) \quad (8)$$

We implemented automatic model selection (i.e., automated determination of the number of clusters C) by optimizing K_{AIC} as a function of C. This optimization was performed by starting with a predefined number of clusters (10) and then iteratively testing whether increase or decrease of the number of clusters improved K_{AIC}. The number of clusters was increased if splitting any existing cluster into two new clusters improved K_{AIC}. The number of clusters was decreased (i.e., the cluster with the smallest contribution to the maximized likelihood L was deleted) if that improved K_{AIC}. These two steps were performed after every fifth iteration of the CEM algorithm. On average, this optimization resulted in a model with 30 clusters.

2.6. Using signal detection for BCI signal translation

The previous two sections describe a technique that can be used to model the probability density function of brain signal features recorded during one condition (e.g., rest). This technique constructs a Gaussian mixture model with C clusters, feature means $\vec{\mu}_c$, covariance matrices Σ_c, and prior probabilities ω_c. The value of C is automatically determined. This model of resting activity can be used to convert brain signal features into output control signal features as described below.

A model of brain activity is first established.[3] Subsequently, for any new signal sample, the posterior probability $p(\vec{a}(n))$ and its negative log-likelihood $\text{LL}(\vec{a}(n))$ can be calculated. This provides an estimate of the probability that this sample was

[2] We also evaluated another criterion, the Bayesian Information Criterion. Its use gave results similar to those reported here for the AIC.

[3] For the application of SIGFRIED to BCI data (Sections 3.1–3.3), explicit rest signal recordings were not available. For these data, we modeled brain activity corresponding to the target for which the subjects produced desynchronized EEG. For the application of SIGFRIED to the motor tasks (Section 3.4), we modeled brain activity corresponding to the recorded rest period.

produced by the modeled reference data distribution:

$$p(\vec{a}(n)) = \sum_{c=1}^{C} p(\vec{a}(n)|c) \tag{9}$$

$$\mathrm{LL}(\vec{a}(n)) = -\log(p(\vec{a}(n))) \tag{10}$$

The negative log-likelihood LL can be expected to be small for samples that are likely to belong to the modeled data distribution (e.g., samples collected during rest), and large for samples that are unlikely (e.g., samples collected during imagined hand movement), and thus it produces an output signal that is under user control. In essence, this procedure translates a feature vector into a control signal that can be used for device control. In this respect, it is similar in function to traditional classification/regression methods (see Fig. 1). The critical difference from traditional approaches is that the proposed approach requires prior data samples from only one class (e.g., rest), and not from multiple classes (e.g., signals associated with a variety of different actual or imagined movements). Thus, it does not require the extensive initial signal identification procedures that currently impede clinical application of BCI technology.

In summary, we extracted feature samples (10 features per 400 ms sample corresponding to frequencies between 10 and 30 Hz) that were associated with one of two classes of data collected during an online BCI experiment. We then used parts of the data (i.e., the training set) to model one of the two classes, and calculated $\mathrm{LL}(\vec{a}(n))$ values for the two classes on the remaining data (i.e., test set). Our primary interest was to determine how well $\mathrm{LL}(\vec{a}(n))$ (i.e., the output of the proposed feature translation method) would discriminate between the two distributions, and not how well the model accounted for the original data distribution. We thus validated our method using a measure of discrimination (rather than a measure of model fit such as K_{AIC}).

2.7. Performance calculation

To measure discrimination performance in our offline analyses, we first averaged $\mathrm{LL}(\vec{a}(n))$ values within each trial (i.e., during the time the subject moved the cursor towards the current target). This produced one value of $\overline{\mathrm{LL}}_t = (1/(n_2 - n_1 + 1))\sum_{n=n_1}^{n_2} \mathrm{LL}(\vec{a}(n))$ for each trial t. These values were typically smaller for trials with the same target that was used as input to the SIGFRIED modeling procedure compared to the values for the opposite target. We then selected the threshold k that minimized the error rate E for the two classes ω_1 and ω_2 of top and bottom target, thereby determining a measure that indicated how well the two class distributions can be separated:

$$E = \int_{k}^{\infty} P(\omega_1|\overline{\mathrm{LL}}) + \int_{-\infty}^{k} P(\omega_2|\overline{\mathrm{LL}}) \tag{11}$$

In this equation, P corresponds to the discrete probabilities of classifying $\overline{\mathrm{LL}}$ into the two classes ω_1 and ω_2. Thus, each comparison of two distributions of $\overline{\mathrm{LL}}$ values resulted in one measure of the minimum error rate. This error rate reflected the fraction of incorrect target classifications (e.g., top instead of bottom target) given an optimally chosen threshold k.

We then determined this error rate (%) using 20-fold cross-validation (i.e., dividing data into 20 subsets, determining model parameters from 19 subsets, determining the minimum error rate on the remaining subset, and repeating this procedure for all 20 test subsets), which resulted, for each subject, in 20 measurements of the minimum error rate for each evaluated method. We used this set of 20 error measures, calculated between the distributions of $\overline{\mathrm{LL}}$ for top and bottom targets, as a performance metric for each of the evaluations described in the subsequent sections.

3. Results

The following sections address the four key questions in this article, i.e.: (1) which parameters minimize BCI error rates calculated offline; (2) whether results achieved using SIGFRIED are comparable to results achieved using linear regression; (3) whether signals produced by SIGFRIED have characteristics that make them amenable to real-time BCI control; (4) whether a SIGFRIED model generated for data collected during rest (as opposed to during one of the classes in a real-time BCI experiment) can be used subsequently to produce discriminating output for data collected during various tasks.

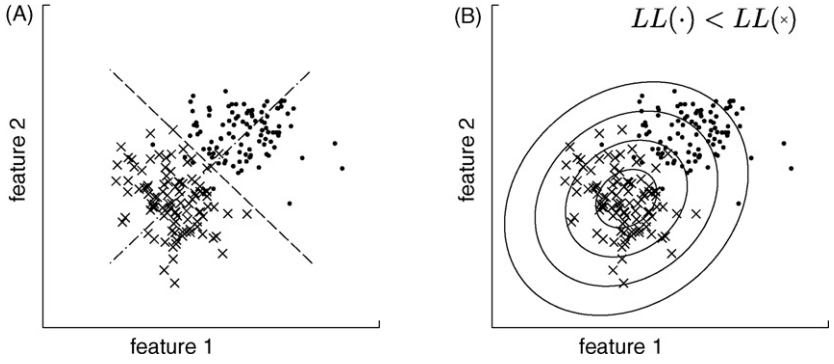

Fig. 1. Detection vs. regression. (A) Traditionally used classification (dashed line) or regression functions (dash-dotted line) can be used to discriminate between data from two classes (indicated by dots and ×). Definition of these functions requires prior analysis of data from both classes, and thus a signal identification procedure. (B) Signal detection (symbolized here by one Gaussian distribution modeling the ×) can be used to discriminate between the two classes, but requires data from only one class (e.g., signals recorded during rest).

3.1. Effect of SIGFRIED parameters on error rates

We first determined how SIGFRIED should be configured and used so that it maximizes performance in the BCI context. The following sections discuss evaluations of different feature transformation and processing methods (i.e., use of diagonal versus full covariance matrices) and of the amount of necessary training data.

3.1.1. Effect of feature transformation

Signal features can be pre-processed using a variety of transformations. In our analyses, untransformed frequency features were usually not Gaussian distributed. (When we applied the Lilliefors test (an adaptation of the Kolmogorov–Smirnov test) to test for normality, none of the 30 features (10 features per subject, 3 subjects) were normally distributed at the 0.05 level.) Some transformations of the features tended to produce distributions that were more similar to Gaussian distributions (e.g., for log-transformed features, 16 of 30 features were normally distributed). Thus, it might seem natural to choose such transformations if signals are modeled with Gaussians, as they are with our methodology. However, we felt an evaluation of this issue was warranted, first, because a good model fit on signals from one class does not guarantee that discrimination performance will also be maximized, and second, because SIGFRIED implements a GMM with automatic model selection, which further complicates interpretations.

To determine which of the selected feature transformations (i.e., no transformation, square root or log transformation) provides the best performance, we determined the classification error rates for each subject using each of the three transformations. The results of this evaluation are shown in Table 1. Irrespective of feature transformation, the achieved error rates are low. This demonstrates that SIGFRIED can discriminate data from two conditions in a BCI experiment with high accuracy. Furthermore, the error rates for the square root and log transformation are higher than those for features that are not transformed. Statistical analysis (using a paired *t*-test) indicates that at least for the comprehensive dataset (8–10 sessions per subject) used in this study, the use of features that are not transformed was significantly better than the use of features that are log transformed.

3.1.2. Effect of using full/diagonalized covariance matrices

In its standard configuration, SIGFRIED models feature variances using full covariance matrices. That is, not only are the variances of each feature estimated, but so are all covariances between all combinations of any two features. We also studied the effect of diagonalization of the covariance matrix so that only the feature variances, but not the covariances calculated between features were evaluated. We did this in two ways. First, we executed our algorithm and estimated a full covariance matrix (i.e., $10^2 = 100$ coefficients), but then set all non-diagonal coefficients to zero (i.e., *diagonal*). Second, we estimated only feature variances during the execution of the algorithm (i.e., 10 coefficients, *estimated diagonal*). We then again calculated error rates for each subject.

Table 1
Feature transformation

Subject	Normal	Sqrt		Log	
	Error (%)	Error (%)	*p*-Value	Error (%)	*p*-Value
A	3.8	4.3	$\gg 0.05$	7.7	$\ll 0.001$
B	9.7	10.5	>0.05	14.6	$\ll 0.001$
C	11.4	13.5	<0.05	17.6	$\ll 0.001$

Error rates achieved using features that were not transformed (normal), square-root transformed (sqrt), or log-transformed (log). Error rates for log transform are significantly higher than those for no transformation (see *p*-values, which were calculated using a paired *t*-test compared to normal).

Table 2
Full or diagonal covariance matrices

Subject	Full	Diagonal		Diagonal est.	
	Error (%)	Error (%)	*p*-Value	Error (%)	*p*-Value
A	3.8	2.3	<0.01	2.2	<0.01
B	9.7	9.4	$\gg 0.05$	7.5	<0.05
C	11.4	9.4	>0.05	9.6	<0.05

Error rates achieved using full covariance matrices (full), or covariance matrices in which all non-diagonal elements were set to zero after estimation of the full matrix (diagonal), or using covariance matrices for which only diagonal elements were estimated (diagonal est.). Error rates for diagonalized covariance matrices are lower than those for full covariance matrices (see *p*-values, which were calculated using a paired *t*-test compared to full). See Section 4 for further interpretation.

The results of this evaluation are shown in Table 2. They indicate the surprising finding that diagonalized covariance matrices gave lower error rates than full covariance matrices (see Section 4 for interpretation). Therefore, we used estimated diagonal covariance matrices for all subsequent evaluations.

3.1.3. Effect of amount of training data

It is important to know how much training data SIGFRIED requires to provide consistent results. It would seem likely that more features should require longer training time. To evaluate the effect of training time on performance, we evaluated the average error rates for each subject. For each evaluation, we used 1, 2, and 4 Hz frequency bins for the 10–30 Hz band (i.e., resulting in 20, 10, and 5 features, respectively) and we varied the amount of training time from 2.15 to 9.75 min (corresponding to 15–65% of the available training time in 10% increments, respectively). The results of this evaluation are shown in Fig. 2. They indicate no consistent effect of the number of features (indicated for 20, 10, and 5 features by dash-dotted, solid, and dashed lines, respectively), and only a modest effect of training time. This suggests that good performance can be achieved with small amounts of training data.

3.2. Comparison of results achieved using SIGFRIED with results achieved using linear regression

The second question was whether the results using SIGFRIED are comparable to those achieved using a conventional method that relies on initial analysis of data from all relevant classes. We chose to compare SIGFRIED to linear regression as this is a widely used method for BCI feature

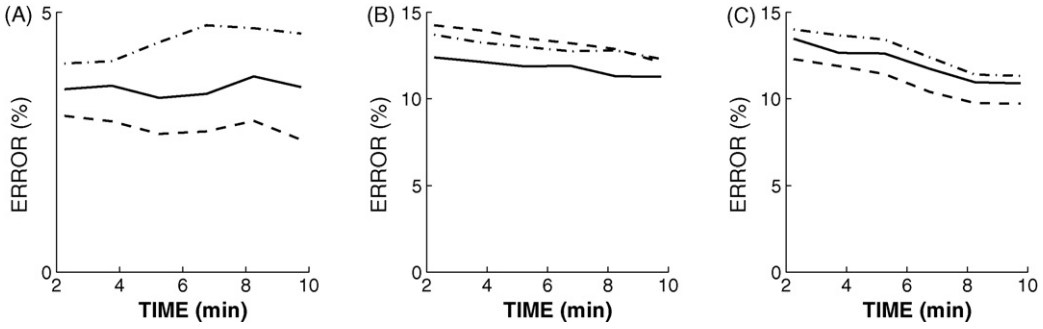

Fig. 2. Effect of training time on performance. Dash-dotted, solid, and dashed lines correspond to results achieved using 20, 10, and 5 features, respectively.

translation (e.g., McFarland and Wolpaw, 2005). To do this, we calculated average error rates using linear regression. Input to linear regression were the same features used for SIGFRIED (i.e., frequency amplitudes a_f calculated in 2 Hz bins from 10 to 30 Hz). The predicted variable was +1 or −1 for the two targets, respectively. This produced a weight vector $\vec{w}$, i.e., one weight w_i for each feature a_f. Output control signal values cs(n) were derived using the dot product of this weight vector with the feature vector at each time point n: $\mathrm{cs}(n) = \vec{w} \cdot \vec{a}(n)$. Just as with SIGFRIED above, this procedure produced output control signal values that discriminated between the two targets. To quantitatively assess the degree of discrimination, we calculated error rates the same way as described in Section 2.7 for LL($\vec{a}(n)$) values. The results of this evaluation are shown in Table 3. Error rates with SIGFRIED were lower than those for linear regression in subjects A and C and comparable in subject B.

3.3. Characteristics of BCI control signals calculated by SIGFRIED

The third question was whether control signals produced by SIGFRIED have characteristics that would make them amenable to real-time BCI experiments. In these BCI experiments, users can modulate brain signal features so as to increase or decrease the resulting SIGFRIED control signal. The amplitude of this control signal could be mapped to the velocity of a cursor, and integrated cursor velocity (i.e., cursor position) could be mapped to one of n possible choices, similar to the experiments described in this paper. In this example, the final classification (i.e., a particular choice) would be determined by a particular sequence of control signal values (e.g., by maintaining a certain value for a particular period).

Table 3
Detection vs. regression

Subject	SIGFRIED	Linear regression	
	Error (%)	Error (%)	p-Value
A	2.2	29.6	$\ll 0.001$
B	7.5	6.8	$\gg 0.05$
C	9.6	38.0	$\ll 0.001$

The results indicate that SIGFRIED produces better results than linear regression in subjects A and C, and comparable results in subject B.

Irrespective of how these control signals are utilized, such experiments demand two relevant requirements that go beyond the question of how much control the user has over these signals. First, control signals must be derived using causal procedures (i.e., procedures that use past data) and their extraction must support rapid feedback. Because SIGFRIED can be applied to the same features as conventional methods in a causal fashion, and because the generation of control signals is computationally simple, SIGFRIED meets this first requirement. The second requirement is that a subject should be able to produce the whole range of possible control signal values equally easily and that inevitable signal variations be normally distributed around the desired control signal value. For example, if the task is to produce a particular mean level of control signal values (e.g., so as to maintain cursor position at a particular level), the signal variations around this mean value should ideally be normally distributed. Because the present dataset was recorded from a four-target task in which the targets were evenly spaced long the y-axis of the screen (implying that the control signals determined online were evenly spread along the value axis), we were able to evaluate the output of SIGFRIED according to this second requirement.

We first calculated one average control signal value for each trial and compiled these values from all cross-validation folds. This produced one distribution of control signal values for each of the four targets. We then log-transformed SIGFRIED output values (i.e., deriving $\log(\mathrm{LL}(\vec{a}(n)))$),[4] and normalized mean target value to the means of top and bottom target to facilitate comparison. Fig. 3 shows the results for SIGFRIED and linear regression (top and bottom panel, respectively) for each of the three subjects A–C. These results demonstrate that SIGFRIED can be used with more than two targets provided that these targets correspond to progressive modulation of signal features. They also show that the distributions of log-transformed SIGFRIED control signal values have means that are evenly spaced along the value axis, that the standard deviations of the distributions are roughly similar for all targets, and that in these two characteristics SIGFRIED compares favorably to linear regression. These results indicate that users should be able to produce dif-

[4] We performed this log transformation because it tended to produce normal distributions. While this transformation of the one-dimensional output values will not change classification results, normally distributed control signal values are desirable for BCI purposes as described above.

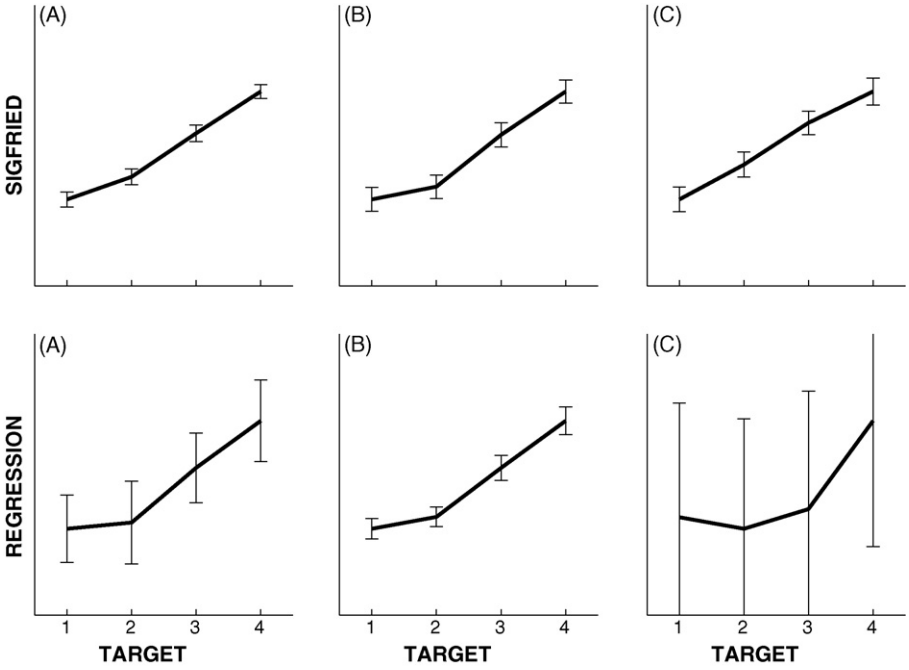

Fig. 3. Plots of averages of log-transformed control signals derived using SIGFRIED (upper panel) and control signals derived using regression (bottom panel) for each of the four target locations and for each patient. Bars represent a quarter of the standard deviation. See text for details.

ferent value ranges of SIGFRIED output control signals equally easily. This should make these signals amenable to real-time BCI operation. Together with the results in Section 3.2, these results further suggest that when used in real-time experiments, SIGFRIED should perform comparable to linear regression.

In summary, we showed that signal detection is an effective modality for BCI operation and that it can deliver results that are comparable to a widely used approach that requires signal identification.

3.4. Application of SIGFRIED to data collected during rest

In previous sections, we demonstrated that SIGFRIED can effectively discriminate between data from two classes collected during a BCI experiment. However, this requires a trained subject. In the end, the use of SIGFRIED will be most beneficial if it can be used to provide discriminating control signals without the signal identification procedures that are currently required. Consequently, the fourth question in our evaluation was whether SIGFRIED can model data collected during rest and subsequently be used to detect data collected during different kinds of tasks.

To answer this question, we analyzed the data from the four additional subjects (D–G) who were naive to BCI control. Using the same features and techniques described before, we derived a model of EEG activity for the resting dataset for each electrode location. Subsequently, we calculated, for each location and task, the error rate between data collected during rest and the task. These error rates were derived from averaged SIGFRIED output values for each 4.05 s cue period, and from 4 s epochs during the 6 min rest dataset. This procedure thus resulted in one measure of error rate for each electrode, task, and subject.

The topographies for these error rates are shown in Fig. 4. These results indicate that SIGFRIED gave low error rates when it is applied to resting data in subjects who are naive to BCI usage. Furthermore, the topographies indicate those electrode locations that change with execution of the different tasks. As expected, these locations are centered on relevant sensorimotor cortices. While there is individual variability in these topographies, their averages are comparable to those previously reported with conventional methods (McFarland et al., 2000).

4. Discussion

This paper describes a novel approach to BCI signal processing implemented in a package called SIGFRIED. It detects changes in brain signals without the need for prior determination of the specific brain signal features and tasks that produce the greatest brain signal changes. Thus, it has the potential to overcome one of the major impediments to translation of laboratory BCI demonstrations into clinical applications. The results of our study indicate that the BCI configuration problem could be reduced from a problem of identifying both the location and the frequency of the signal that the subject can modulate to the problem of merely choosing an appropriate location. Because with EEG only one or more of a few locations are usually used for BCI control (e.g., C_3, C_z, C_4), configuration for each new user should be relatively easy. Ultimately, it may be possible to combine all features from select locations into one model so that even the location would not have to be selected. In this paper we tested models that used all relevant frequency bands in one particular location. In this situation, SIGFRIED is able to support BCI control even if the frequency that is modulated by the user is unknown or changes over time.

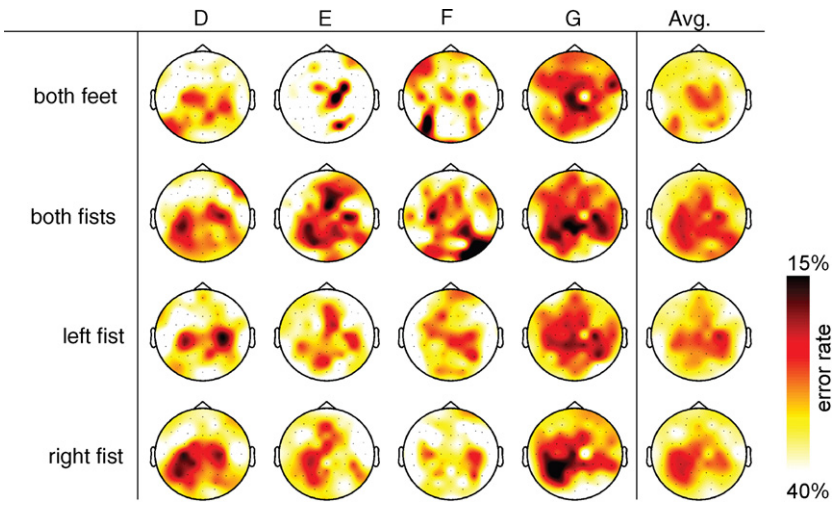

Fig. 4. Topographies of error rates for the additional subjects D–G with different motor tasks. While there is individual variability in these topographies, their average is similar to patterns given by conventional methods.

SIGFRIED is of value to BCI research mostly because of the increased practicality it provides. While we showed in this paper that SIGFRIED provides favorable performance compared to linear regression, SIGFRIED relies on the statistics of only one class. While this approach will always function properly as long as the data corresponding to the tested class are different to those for the modeled class, it may not provide optimal performance. In other words, it is likely that comprehensive optimization of signal classification using all available signal classes will outperform results achieved using only one class. For two reasons, signal detection will likely remain of value. First, there are several avenues to further optimize the technique that is presented here. For example, automatic feature selection, which has proven effective for data with more than one class, may also be possible with data from only one class. For example, one may include only those frequency features that are identified by typical mu/beta peaks. Furthermore, adaptation of the resting signal model to reflect ongoing changes in brain signals may improve performance. Second, the unique value of the detection approach is that it requires less *a priori* information. Thus, detection will always be valuable when little such information is available such as during early subject training and/or when using signals with a high feature/task space such as electrocorticographic (ECoG) signals.

One surprising result of the present study is that full covariance matrices typically produced results inferior to those of diagonalized covariance matrices. This was true whether we estimated all parameters of the covariance matrix and subsequently set the non-diagonal entries (i.e., covariances) to zero, or whether we estimated only diagonal coefficients (i.e., variances). In other words, superior results were achieved when the model disregarded information about how particular frequency bands vary with respect to others (e.g., mu versus beta frequencies). This may indicate that important decision boundaries reflect ellipsoids whose axes are parallel to the coordinate axes, or that the covariances estimated from the training set do not generalize well to those required for the test set. It may also be related to the likely possibility that some features (i.e., frequency amplitudes) are correlated with each other. In fact, auxiliary analyses (data not shown) using simulated data that contained correlated features also produced improved performance for diagonalized covariance matrices.

The strength of the detection approach is also its main weakness. Because SIGFRIED cannot distinguish among modulation of different features within one model (e.g., one brain location), it can effectively produce only one (albeit graded) control signal for each such model. Furthermore, this control signal will only differentiate between different brain signal classes under the assumption that these classes are produced by modulation of one set of features such that they progressively differ from the resting state. Thus, one could say that SIGFRIED is a trade-off between specificity and practicality.

Because SIGFRIED depends for appropriate function strongly on data collected during rest, the quality of the recording during this resting period is important. While common sources of external noise (such as line noise, muscular artifacts, etc.) are well understood and can be minimized, it is not clear how the subject should be instructed to perform during this period so that it would produce data that are maximally different to those produced during a motor or imagery task. Furthermore, a number of external influences exist that may modulate the same signals that are modulated during movement imagery (e.g., tactile stimulation or observing movements in other people while resting; Neuper et al., 2005).

Modeling of background signals has been used extensively in other domains such as image processing (see Friedman and Russell, 1997; Harville et al., 2001; Kuo et al., 2003; Lee, 2005; Liyuan et al., 2004; Pless, 2003; Stauffer and Grimson, 1999; Toyama et al., 1999). This approach has been practically absent from biosignal processing with very few exceptions (e.g., Costa and Cabral, 2000; Harris et al., 2000; Pernkopf and Bouchaffra, 2005). Nevertheless, there are many applications within biosignal analysis that could benefit from background modeling. For example, detection of P300 evoked potentials (see Farwell and

Donchin, 1988) is typically achieved using standard classification techniques. These techniques work very well with the stable responses typically found in healthy individuals, but often fail if responses vary in time or in space as may happen with severely paralyzed users. Background modeling could make detection performance invariant to time and space. Artifact detection (see Anderer et al., 1999; Goncharova et al., 2003; Schlögl et al., 1999) is another possible and attractive application. Finally, with few exceptions (e.g., Gardner et al., 2005), detection of epileptic seizures is also often performed using hand-crafted or machine-learned classification criteria that describe the seizure signature. These approaches also have problems if this signature changes. Thus, signal detection, which is invariant to the seizure signature, may prove beneficial.

In summary, current approaches to analyzing and using brain signals typically require comprehensive analyses of data for different conditions. In this paper, we describe a procedure that complements these approaches. Our method can produce effective results using only data recorded in one condition (e.g., rest). It should thereby facilitate the development and clinical use of BCI systems.

Acknowledgements

This work was supported in the USA by the National Center for Medical Rehabilitation Research, National Institute of Child Health and Human Development, National Institutes of Health (NIH) (Grant HD30146 (JRW)) and by the National Institute of Biomedical Imaging and Bioengineering and the National Institute of Neurological Disorders and Stroke, NIH (Grants EB006356 (GS) and EB00856 (JRW)). We thank Elizabeth Winter Wolpaw for valuable comments on the manuscript.

Appendix A

The large Laplacian spatial filter was calculated by subtracting from the measured activity at the channel of interest the mean activity of four neighboring electrodes:

$$s'(k) = s_h(k) - \frac{1}{4}\sum_{i \in E_h} s_i(k) \quad (12)$$

where E_h was the set of four next-nearest neighbors surrounding electrode h.

The control signal $S(n)$ was calculated as a weighted linear combination of the estimated frequency amplitudes $a_{hj}(n)$ at a particular channel h and center frequency j:

$$S(n) = \sum_{h \in H, j \in J} w_{hj} a_{hj}(n) \quad (13)$$

The control signal $S(n)$ was translated into cursor movement $\Delta V(n)$ using a linear transformation: $\Delta V(n) = b(S(n) - a)$ where b was the gain and a was the mean control signal. These parameters were estimated from the user's previous performance.

References

Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Proceedings of the second international symposium on information theory; 1973. p. 267–81.

Anderer P, Roberts S, Schlögl A, Gruber G, Kloesch G, Herrmann W, et al. Artifact processing in computerized analysis of sleep EEG—a review. Neuropsychobiology 1999(40):150–7.

Babiloni F, Cincotti F, Lazzarini L, Millan J, Mourino J, Varsta M, et al. Linear classification of low-resolution EEG patterns produced by imagined hand movements. IEEE Trans Rehabil Eng 2000;8(2):186–8.

Biernacki C, Celeux G, Govaert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Comput Stat Data Anal 2003;41:561–75.

Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, et al. A spelling device for the paralysed. Nature 1999;398(6725):297–8.

Blanchard G, Blankertz B. BCI competition data set IIa: spatial patterns of self-controlled brain rhythm modulations. IEEE Trans Biomed Eng 2003;51(6):1062–6.

Blankertz B. BCI competition 2003 (web page); 2003. http://ida.first.fhg.de/projects/bci/competition.

Blankertz B, Müller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Trans Biomed Eng 2004;51(6):1044–51.

Bozdogan H. Model selection and Akaike's information criterion. Psychometrika 1974;52:345–70.

Burg JP. Maximum entropy spectral analysis. In: Proceedings of the 37th meeting of society of exploration geophysicists; 1967.

Burg JP. A new analysis technique for time series data. NATO Advanced Study Institute on signal processing: underwater acoustics; 1968, p. 42–8. Reprinted in Childers (1978).

Cacoullos T. Estimation of a multi-variate density. Ann Inst Math Stat (Tokyo) 1966;18(2):179–89.

Celeux G, Govaert G. A classification EM algorithm for clustering and two stochastic versions. Comput Stat Data Anal 1992;14:279–416.

Childers D, editor. Modern spectrum analysis. IEEE Press; 1978.

Costa EJ, Cabral EF. EEG-based discrimination between imagination of left and right hand movements using adaptive Gaussian representation. Med Eng Phys 2000;22(5):345–8.

Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7.

Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B 1977;39:1–38.

Donoghue JP, Nurmikko A, Black M, Hochberg LR. Assistive technology and robotic control using motor cortex ensemble-based neural interface systems in humans with tetraplegia. J Physiol 2007;579(Pt 3):603–11.

Duda R, Hart P, Stork D. Pattern classification. 2nd ed. John Wiley; 2001 [chapter 2, ISBN: 0-471-05669-3].

Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 1988;70(6):510–23.

Friedman N, Russell S. Image segmentation in video sequences: a probabilistic approach. In: IEEE proceedings of the thirteenth conference on uncertainty in artificial intelligence (UAI); 1997.

Gardner A, Krieger A, Vachtsevanos G, Litt B. One-class novelty detection for seizure analysis from intracranial EEG. J Mach Learn Res 2005;7:1025–44.

Garrett D, Peterson DA, Anderson CW, Thaut MH. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans Rehabil Eng 2003;11(2):141–4.

Goncharova I, McFarland D, Vaughan T, Wolpaw J. EMG contamination of EEG: spectral and topographical characteristics. Electroencephalogr Clin Neurophysiol 2003;114:1580–93.

Gysels E, Renevey P, Celka P. SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in brain–computer interfaces. Signal Process 2005;85(11):2178–89.

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzski G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 2000;84:401–14.

Harville M, Gordon G, Woodfill J. Foreground segmentation using adaptive mixture models in color and depth. In: IEEE proceedings, IEEE workshop on detection and recognition of events in video; 2001. p. 3–4.

Haykin S. Neural networks: a comprehensive foundation. Prentice Hall; 1998.

Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 2006;442(7099):164–71.

Huan NJ, Palaniappan R. Neural network classification of autoregressive features from electroencephalogram signals for brain–computer interface design. J Neural Eng 2004;1(3):142–50.

Kennedy PR, Bakay RA, Moore MM, Goldwaithe J. Direct control of a computer from the human central nervous system. IEEE Trans Rehabil Eng 2000;8(2):198–202.

Kübler A, Kotchoubey B, Hinterberger T, Ghanayim N, Perelmouter J, Schauer M, et al. The thought translation device: a neurophysiological approach to communication in total motor paralysis. Exp Brain Res 1999;124(2):223–32.

Kübler A, Nijboer F, Mellinger J, Vaughan TM, Pawelzik H, Schalk G, et al. Patients with ALS can use sensorimotor rhythms to operate a brain–computer interface. Neurology 2005;64(10):1775–7.

Kuo B, Yan J, Landgrebe D. Gaussian mixture classifier with regularized covariance estimator for hyperspectral data classification. In: Proceedings of the geoscience and remote sensing symposium (IGARSS '03), vol. 1; 2003. p. 276–8.

Lal TN, Schroder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, et al. Support vector channel selection in BCI. IEEE Trans Biomed Eng 2004;51(6):1003–10.

Lee D. Effective Gaussian mixture learning for video background subtraction. IEEE Trans Pattern Anal Mach Intell 2005;27(5):827–32.

Liyuan L, Huang W, Gu IY-H, Tian Q. Statistical modeling of complex backgrounds for foreground object detection. In: IEEE transactions, image processing, vol. 13; 2004. p. 1459–72.

Marple L. Digital spectral analysis with applications. Upper Saddle River, NJ: Prentice-Hall; 1987.

McFarland DJ, Wolpaw JR. Sensorimotor rhythm-based brain–computer interface (BCI): feature selection by regression improves performance. IEEE Trans Neural Syst Rehabil Eng 2005;13(3):372–9.

McFarland DJ, Neat GW, Wolpaw JR. An EEG-based method for graded cursor control. Psychobiology 1993;21:77–81.

McFarland DJ, Lefkowicz T, Wolpaw JR. Design and operation of an EEG-based brain–computer interface (BCI) with digital signal processing technology. Behav Res Methods Instrum Comput 1997a;29:337–45.

McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. Electroencephalogr Clin Neurophysiol 1997b;103(3):386–94.

McFarland DJ, Miner LA, Vaughan TM, Wolpaw JR. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr 2000;12(3):177–86.

McFarland DJ, Anderson CW, Müller K-R, Schlögl A, Krusienski DJ. BCI meeting 2005—workshop on BCI signal processing: feature extraction and translation. IEEE Trans Neural Syst Rehabil Eng; 2006;14(2)135–8.

Millán J del R, Renkens F, Mouriño J, Gerstner W. Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans Biomed Eng 2004;51(6):1026–33.

Müller K, Blankertz B. Toward noninvasive brain–computer interfaces. IEEE Signal Process Mag 2006;23(5):126–8.

Müller KR, Anderson CW, Birch GE. Linear and nonlinear methods for brain–computer interfaces. IEEE Trans Rehabil Eng 2003;11(2):165–9.

Neuper C, Scherer R, Reiner M, Pfurtscheller G. Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. Brain Res Cogn Brain Res 2005;25(3):668–77.

Parzen E. On estimation of probability density function and mode. Ann Math Stat 1962;33:1065–73.

Pernkopf F, Bouchaffra D. Genetic-based EM algorithm for learning Gaussian mixture models. IEEE Trans Pattern Anal Mach Intell 2005;27(8):1344–8.

Pfurtscheller G, Flotzinger D, Kalcher J. Brain–computer interface—a new communication device for handicapped persons. J Microcomput Appl 1993;16:293–9.

Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between imagination of right and left hand movement. Electroencephalogr Clin Neurophysiol 1997;103(6):642–51.

Pfurtscheller G, Guger C, Müller G, Krausz G, Neuper C. Brain oscillations control hand orthosis in a tetraplegic. Neurosci Lett 2000;292(3):211–4.

Pless R. Evaluation of local models of dynamic backgrounds. In: IEEE proceedings, computer vision and pattern recognition, vol. 2; 2003. p. 73–8.

Priestley M. Spectral analysis and time series. Orlando, FL: Academic Press; 1981.

Radke RJ, Andra S, Al-Kofahi O, Roysam B. Image change detection algorithms: a systematic survey. IEEE Trans Image Process 2005;14(3):294–307.

Rissanen J. Modelling by the shortest data description. Automatica 1978; 14:465–71.

Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: a general-purpose brain–computer interface (BCI) system. IEEE Trans Biomed Eng 2004;51:1034–43.

Schlögl A, Anderer P, Barbanoj M-J, Dorffner G, Gruber G, Kloesch G, et al. Artifacts in the sleep EEG—a database for the evaluation of automated processing methods. In: Proceedings of the third international congress, of the world federation of sleep research societies (WFSRS), vol. 2; 1999. p. 586.

Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Comput 2001;13(7):1443–71.

Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.

Serruya M, Hatsopoulos N, Paninski L, Fellows M, Donoghue J. Instant neural control of a movement signal. Nature 2002;416(6877):141–2.

Sharbrough F, Chatrian G, Lesser R, Luders H, Nuwer M, Picton T. American electroencephalographic society guidelines for standard electrode position nomenclature. Electroencephalogr Clin Neurophysiol 1991;8:200–2.

Stauffer C, Grimson W. Evaluation of local models of dynamic backgrounds. In: IEEE proceedings, computer vision and pattern recognition, vol. 2; 1999. p. 246–52.

Sutter EE. The brain response interface: communication through visually guided electrical brain responses. J Microcomput Appl 1992;15:31–45.

Taylor DM, Tillery SI, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. Science 2002;296:1829–32.

Torr PHS. An assessment of information criteria for motion model selection. In: Proceedings of the conference on computer vision and pattern recognition (CVPR '97); 1997. p. 47–52.

Toyama K, Krumm J, Brumitt B, Meyers B. Statistical modeling of complex backgrounds for foreground object detection. In: IEEE proceedings of the seventh IEEE international conference on computer vision, vol. 1; 1999. p. 20–7.

Vapnik VN, Chervonenkis AJ. Ordered risk minimization (I and II). Automat Remote Control 1974;34, 1226–35 and 1403–12.

Vaughan TM, McFarland DJ, Schalk G, Sarnacki WA, Krusienski DJ, Sellers EW, et al. The Wadsworth BCI research and development program: at home with BCI. IEEE Trans Neural Syst Rehabil Eng 2006;14(2):229–33.

Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, et al. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 2000;408:361–5.

Wolpaw J, McFarland D. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. Proc Natl Acad Sci 2004;101(51):17849–54.

Wolpaw JR, McFarland DJ, Neat GW, Forneris CA. An EEG-based brain–computer interface for cursor control. Electroencephalogr Clin Neurophysiol 1991;78(3):252–9.

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Electroencephalogr Clin Neurophysiol 2002;113(6):767–91.