Hidden Markov model and support vector machine based decoding of finger movements using electrocorticography

Tobias Wissel1,9,10, Tim Pfeiffer1, Robert Frysch1, Robert T Knight2,3, Edward F Chang2, Hermann Hinrichs4,5,6,7, Jochem W Rieger3,8 and Georg Rose1

1 Chair for Healthcare Telematics and Medical Engineering, Otto-von-Guericke-University Magdeburg, Postfach 4120, D-39016 Magdeburg, Germany
2 Department of Neurological Surgery, University of California, San Francisco, 505 Parnassus Ave., M-779 San Francisco, CA 94143-0112, USA
3 Department of Psychology and the Helen Wills Neuroscience Institute, University of California, Berkeley, 132 Barker Hall, Berkeley, CA 94720-3190, USA
4 Clinic of Neurology, Otto-von-Guericke-University Magdeburg, Leipziger Straße 44, D-39120 Magdeburg, Germany
5 Leibniz-Institute for Neurobiology, Brenneckestraße 6, D-39118 Magdeburg, Germany
6 German Center for Neurodegenerative Diseases (DZNE), Leipziger Straße 44, D-39120 Magdeburg, Germany
7 Center of Behavioural Brain Sciences (CBBS), Universitätsplatz 2, D-39106 Magdeburg, Germany
8 Applied Neurocognitive Psychology, Faculty VI, Carl-von-Ossietzky University, D-26111 Oldenburg, Germany

E-mail: wissel@rob.uni-luebeck.de

Received 28 May 2013
Accepted for publication 28 August 2013
Published 18 September 2013
Online at stacks.iop.org/JNE/10/056020

Abstract

Objective. Support vector machines (SVM) have developed into a gold standard for accurate classification in brain–computer interfaces (BCI). The choice of the most appropriate classifier for a particular application depends on several characteristics in addition to decoding accuracy. Here we investigate the implementation of hidden Markov models (HMM) for online BCIs and discuss strategies to improve their performance.

Approach. We compare the SVM, serving as a reference, and HMMs for classifying discrete finger movements obtained from electrocorticograms of four subjects performing a finger tapping experiment. The classifier decisions are based on a subset of low-frequency time domain and high gamma oscillation features.

Main results. We show that decoding optimization between the two approaches is due to the way features are extracted and selected and less dependent on the classifier. An additional gain in HMM performance of up to 6% was obtained by introducing model constraints. Comparable accuracies of up to 90% were achieved with both SVM and HMM with the high gamma cortical response providing the most important decoding information for both techniques.

Significance. We discuss technical HMM characteristics and adaptations in
the context of the presented data as well as for general BCI applications. Our findings suggest that HMMs and their characteristics are promising for efficient online BCIs.

Online supplementary data available from stacks.iop.org/JNE/10/056020/mmedia
(Some figures may appear in colour only in the online journal)

1. Introduction

Brain–computer interface (BCI) oriented research has a principle goal of aiding disabled people suffering from severe motor impairments (Hoffmann et al 2007, Palaniappan et al 2009). The majority of research on BCI has been based on electroencephalography (EEG) data and restricted to simple experimental tasks using a small set of commands. In these studies (Cincotti et al 2003, Lee and Choi 2003, Obermaier et al 1999) information was extracted from a limited number of EEG channels over scalp sites of the right and left hemisphere.

Recently, studies have employed invasive subdural electrocorticogram (ECoG) recordings for BCI (Ganguly and Carmena 2010, Schalk 2010, Zhao et al 2010, Shenoy et al 2007). ECoG is recorded for diagnosis in clinical populations, e.g. for localization of epileptic foci. The signal quality of ECoG recorded brain activity outperforms the EEG data with respect to higher amplitudes and higher signal-to-noise ratio (SNR), higher spatial resolution, and broader bandwidth (0–500 Hz) (Crone et al 1998, Schalk 2010). Thus, ECoG-signals have the potential for improving earlier results of feature extraction and signal classification (Schalk 2010).

While approaches have been made to correlate continuous movement kinematics and brain activity exploiting regression techniques (e.g. Wiener filter, Kalman filter and others: Dethier et al 2011, Acharya et al 2010, Ball et al 2009, Kubanek et al 2009, Liang and Bougrain 2009, Pistohl et al 2008), the support vector machine (SVM) approach has been established as a gold standard for deriving class labels in cases of a discrete set of control commands (Quandt et al 2012, Liu et al 2010, Zhao et al 2010, Demirer et al 2009, Shenoy et al 2007). This is due to high and reproducible classification performance as well as robustness with a low number of training samples using SVM approaches (Guyon et al 2002).

However, depending on the individual problem, distinct properties of other machine learning methods may provide another viable BCI approach. Pascual-Marqui et al (1995) and Obermaier et al (1999) argued that brain states and state transitions can explain components of observed human brain activity. Obermaier et al applied HMMs in offline classification of non-invasive EEG recordings of brain activity to distinguish between two different imagined movements. In support of the idea of distinct mental states, they reported an improvement of BCI classification results in subjects who developed a consistent imagination strategy resulting in more focused and reproducible brain activity patterns. The authors concluded that HMMs offer a promising means to classify mental activity in a BCI framework because of the inherent HMM flexibility in structural degrees of freedom and simplicity. Some studies however reveal unstable properties and lower HMM robustness in high dimensional feature spaces. The latter point is critically related to small training sets leading to degradation in the classifier’s ability (Lotte et al 2007).

In this study we directly compare HMM and SVM classification of ECoG data recorded during a finger tapping experiment. This was done after a careful optimization of both classifiers—first, focusing on improvements of the features and secondly by introducing constraints to the HMMs reducing the amount of required training data. Additionally, the problem of high dimensional feature spaces is addressed by feature selection. It is known that feature extraction and selection have a crucial impact on classifier performance (Pechenizkiy 2005, Guyon and Elisseeff 2003). However, the exact effect on the decoding accuracy for given data characteristics and representations of information in time and space remains to be investigated. Poorly constructed feature spaces cannot be compensated for by the classifier. However, an optimal feature space does not make the choice of the classifier irrelevant. A tendency toward unstable HMM behaviour as well as the different ways HMM and SVM model feature space was addressed with the ECoG data set.

Our hypothesis was that for ECoG classification of finger representations, differences in performance between the two approaches would be due to the way features are extracted and selected and less influenced by the choice of the classifier. Here we demonstrate that constrained HMMs achieve decoding rates comparable to the SVM. We first provide an overview of how the data were acquired. We then review the analysis methods, involving feature extraction, HMMs and SVMs. Third, results on feature space setup and feature selection are evaluated. Finally, we discuss the decoding performance of the HMM classifier and directly compare it to the one-vs-one SVM.

2. Material and methods

2.1. Patients, experimental setup and data acquisition

ECoG data were recorded from four patients who volunteered to participate (aged 18–35 years, right handed male). All patients received subdural electrode implants for pre-surgical planning of epilepsy treatment at UC San Francisco, CA, USA. The electrode grid was solely placed based on clinical criteria and covered cortical pre-motor, motor, somatosensory, and temporal areas. Details on the exact grid placement for each subject can be obtained from supplementary figures S1–S4 available from stacks.iop.org/JNE/10/056020/mmedia. The study was conducted in concordance with the local institutional review board approved protocol as well as the Declaration of Helsinki. All patients gave their informed consent before
the recordings started. The recordings did not interfere with the treatment and entailed minimal additional risk for the participant. At the time of recording all patients were off anti-epileptic medications. They were withdrawn when they were admitted for ECoG monitoring.

The recordings were obtained from a 64-channel grid of 8 × 8 platinum–iridium electrodes with 1 cm centre-to-centre spacing. The diameter of the electrodes was 4 mm, of which 2.3 mm were exposed (except for subject 4 who had a 16 × 8 electrode grid, 4 mm centre-to-centre spacing). The ECoG was recorded with a hardware sampling frequency of 3051.7 Hz and was then down-sampled to 1017 Hz for storage and further processing. During the recordings patients performed a serial reaction time task in which a number appeared on a screen indicating with which finger to press a key on the keyboard (1 indicated thumb, 2: index, 3: middle, 5: little finger; the ring finger was not used). The next trial was automatically initiated with a short randomized delay of (335 ± 36) ms after the previous key press (mean value for S1, session 1; others in similar range). The patients were instructed to respond as rapidly and accurately as possible, with all fingers remaining on a fixed key during the whole 6–8 min run. For each trial the requested as well as actually used finger was recorded. The rate of correct button presses was at or close to 100% for all subjects. The classification trials were labeled according to the finger that was actually moved. Each patient participated in 2–4 of these sessions (table 1). Subjects performed slight finger movements requiring minimal force. Topographic finger representations extend approximately 55 mm along the motor cortex (Penfield and Boldrey 1937). The timing of stimulus presentations and button presses was recorded in auxiliary analogue channels synchronized with the brain data.

Due to clinical constraints, the time available for ECoG data acquisition is restricted, providing a challenging dataset for any classifier. The numbers of trials for classification are listed in table 1. For pre-processing the time series was first highpass filtered using a cut-off frequency at 0.5 Hz as well as notch filtered around the power line frequency (60 Hz) using frequency domain filtering. The electric potentials on the grid were re-referenced to common average reference (CAR).

Next, the time series were visually inspected for the remaining intervals and artefacts caused by the measurement hardware, line noise, loose contacts and sections with signal variations not explained by normal physiology were excluded from the analysis. Channels with epileptic activity were removed since epileptic spiking has massive high-frequency spectral activity and can distort informative features in this range. We also removed any epochs with spread of the epileptic activity to normal brain sites to avoid high-frequency artefacts. The procedure was chosen to ensure comparability across subjects and sessions. Trial rejection was carried out without any knowledge of the corresponding class labels and in advance of any investigations on classification accuracies. Finally, the trials were aligned with respect to the detected button press being the response for each movement request.

### 2.2. Test framework

The methods described in the following assemble an overall test framework as illustrated in figure 1. The classifier, being the main focus here, is embedded as the core part among other supporting modules including pre-processing, data grouping according to the testing scheme, channel selection and feature extraction. The framework includes two separate paths for training and testing data to ensure that no prior knowledge about the test data is used in the training process.

### 2.3. Feature generation

Two different feature types have been extracted for this study. First, time courses of low-frequency components were obtained by spectral filtering in the Fourier domain. These time courses were limited to a specific interval of interest around the button press containing the activation most predictive for the four classes. The chosen frequency range of the lowpass filter covers a band of low-frequency components that contain phase-locked event-related potentials (ERPs) and prominent local motor potentials (LMP) (Kubanek et al 2009). These are usually extracted by averaging out uncorrelated noise over multiple trials (Makeig 1993). Here we focus on single-trial analysis which may not reflect LMP activity (Picton et al 1995). As a consequence, we refer to this feature type as low-frequency time domain (LFTD) features. Boundaries of time interval (length and location relative to the button press) and

### Table 1. The number of recorded and rejected trials per subject, session and class (thumb/index finger/middle finger/little finger).

| Subject # | Session # | Trials | Class breakdown | Rejected trials (breakdown) |
|-----------|-----------|--------|----------------|-----------------------------|
| S1        | 1         | 357    | 62/113/108/74  | 0 (0/0/0/0)                |
| 2         | 240       | 33/90/72/45 | 1 (0/1/0/0) |
| Total     | 587       | 95/203/180/119 | 1 (0/1/0/0) |
| S2        | 1         | 280    | 44/97/91/48   | 40 (5/13/13/9)             |
| 2         | 308       | 51/103/111/43 | 0 (0/0/0/0) |
| 3         | 271       | 38/87/92/54  | 77 (13/23/26/15)          |
| 4         | 273       | 32/88/88/65  | 41 (8/17/10/6)            |
| Total     | 1122      | 165/365/382/210 | 158 (26/53/49/30) |
| S3        | 1         | 234    | 54/70/70/40   | 114 (16/40/36/22)          |
| 2         | 249       | 47/70/83/49  | 104 (15/31/38/20)         |
| Total     | 483       | 101/140/153/89 | 218 (31/71/74/42)         |
| S4        | 1         | 359    | 57/113/116/73 | 6 (2/2/1/1)                |
| 2         | 328       | 63/104/106/55 | 33 (6/11/11/5)            |
| Total     | 687       | 120/217/222/128 | 39 (8/13/12/6)            |
lowpass filter cut-off were selected by grid search for each session and are stated in the results. The data sampling was reduced to the Nyquist frequency.

Second, event-related spectral perturbation (ERSP) was measured using a sliding Hann-window approach (window length $n_w = 260$ ms). For each window the square root of the power spectrum was computed by fast Fourier transform (FFT). The resulting coefficients were then averaged in a frequency band of interest. Taking the square root to obtain the amplitude spectrum assigned a higher weight to high-frequency components compared to the original power spectrum. The resulting feature sequence can be used as a measure for time-locked power fluctuation in the high gamma band, neglecting explicit phase information (Graimann et al 2004, Crone et al 1998, Makeig 1993).

The time interval relative to the button press, and the window overlap and exact low and high cut-off frequency of the high gamma band most discriminative for the four classes were selected by grid search. The resulting parameters for each session are stated in the results section.

A combined feature space including both high gamma (HG) and as well as LFTD features was also evaluated. The ratio $p$ of how many LFTD feature sequences are used relative to the number of HG sequences was obtained by grid search. Subset selection beyond this ratio is described in the next subsection.

2.4. Feature selection

2.4.1. General remarks on subset selection. Feature sequences were computed for all 64 ECoG channels giving $2 \times 64$ sequences in total. From these sequences a subset of size $O$ has been identified, that contains the most relevant information about the finger labels. A preference towards certain channels is influenced by the grid placement as well as distinct parts of the cortex differentially activated by the task.

We tested several measures of feature selection such as Fisher’s linear discriminant analysis (Fisher 1936) and Bhattacharyya distance (Bhattacharyya 1943) and found that the Davies–Bouldin (DB) index (Wissel and Palaniappan 2011, Davies and Bouldin 1979) yielded the most robust and stable subset estimates for accurate class separation. In addition to higher robustness, the DB index also represents the fastest alternative computationally. The DB index substantially outperformed other methods for both SVM and HMM. Nevertheless it should be pointed out that observed DB subsets were in good agreement with the ranking generated by the normal vector $w$ of a linear SVM. Apart from so-called filter methods such as the DB index, an evaluation of these vector entries after training the SVM would correspond to feature selection using a wrapper method (Guyon and Elisseeff 2003). Our study did not employ any wrapper methods to avoid dependences between feature selection and classifier. This may have led to an additional bias towards SVM or HMM (Yu and Liu 2004). Based on the DB index we defined an algorithm that automatically selects an $O$-dimensional subset of feature sequences as follows.

2.4.2. DB index. The index corresponds to a cluster separation index measuring the overlap of two clusters in an arbitrarily high dimensional space, where $r \in \mathbb{R}^T$ is one of $N$ sample vectors belonging to cluster $C_i$:

$$
\mu_i = \frac{1}{N} \sum_{r \in C_i} r \quad (1)
$$

$$
d_i = \frac{1}{N} \sum_{r \in C_i} ||r - \mu_i|| \quad (2)
$$

$$
R_{ij} = \frac{d_i + d_j}{||\mu_i - \mu_j||} \quad i, j \in \mathbb{N}^+, i \neq j \quad (3)
$$

Using the cluster centroids $\mu_i \in \mathbb{R}^T$ and spread $d_i$, matrix $R = \{R_{ij}\}_{i=1...N, j=1...N}$ contains the rate between the within-class spread and the between-class spread for all class combinations. The smaller the index, the less two clusters overlap in that feature space. The index is computed for all features available.

2.4.3. Feature selection procedure. Covering different aspects of class separation, we developed a feature selection algorithm based on the DB index taking multiple classes into account. The algorithm aims at minimizing the mutual class overlap for a subset of defined size $O$ as follows. First, each feature sequence is segmented into three parts of equal length. The index matrix $R$ measuring the cluster overlap between all classes $i$ and $j$ is then computed segment-wise for each feature under consideration. Subsequently all three matrices are averaged (geometric mean for a bias towards smaller indices) for further computing. The feature yielding the smallest mean index is selected, since it provides the best separation on average from one class to at least one other class. Second, for all class combinations the features corresponding to the minimal index per combination are added. Each of
Figure 2. An HMM structure for one class consists of \( Q \) states emitting multi-dimensional observations \( o_t \) at each time step. The state as well as the output sequence is governed by a set of probabilities including transition probabilities \( a_{ij} \). To assign a sequence \( O_t \) to a particular model, the classifier decides for the highest probability \( P(O_t | \text{model}) \).

them optimally separates a certain finger combination. Next, features with the smallest geometric mean per class are added, if they are not part of the subset already. Finally, in case the predefined number of elements in the subset \( O \) admits further features, vacancies are filled according to a sorted ranking of indices averaged across class combinations per feature.

For each trial this subset results in an \( O \)-dimensional feature vector varying over time \( t \in [0, T] \). While this multivariate sequence was fed directly into the HMM classifier, the \( O \times T \) matrix \( O_t \) had to be reshaped into a \( 1 \times (T \cdot O) \) vector for the SVM. The subset size \( O \) is obtained by exhaustive search on the training data. This makes the difference between both classifiers in data modelling obvious.

While an HMM technically distinguishes between the temporal development (columns in \( O_t \)) and the feature type (rows in \( O_t \)), the SVM concatenates all channels and their time courses within the same one-dimensional feature vector. The potential flexibility of HMMs with respect to temporal phenomena arises substantially from this fact.

2.5. Classification

2.5.1. Hidden Markov models. Arising from a state machine structure, hidden Markov models (HMMs) comprise two stochastic processes. First, as illustrated in figure 2, a set of \( Q \) underlying states is used to describe a sequence \( O_t \) of \( T \) \( O \)-dimensional feature vectors. These states \( q_t \in \{s_i\}_{i=1...Q} \) are assumed to be hidden, while each is associated with an \( O \)-dimensional multivariate Gaussian \( M \)-mixture distribution to generate the observed features—the second stochastic process. Both, the state incident \( q_t \) as well as the observed output variable \( o_t \) at time \( t \) are subject to the Markov property—a conditional dependence only on the last step in time \( t - 1 \).

Among other parameters the state sequence structure and its probable variations are described by prior probabilities \( P(q_1 = s_i) \) and a transition matrix

\[
A = \{a_{ij}\}_{i,j=1...Q} = \{P(q_{t+1} = s_j | q_t = s_i)\}_{i,j=1...Q}.
\]

A more detailed account of the methodological background is given by Rabiner (1989).

For each of the four classes one HMM had to be defined, which was trained using the iterative Baum–Welch algorithm with eight expectation maximization (EM) steps (Dempster et al 1977). The classification on the test set was carried out following a maximum likelihood approach as indicated by the curly bracket in figure 2.

In order to increase decoding performance, the above standard HMM structure was constrained and adapted to the classification problem. This aims at optimizing the model hypothesis to give a better fit and reduces the number of free parameters to deal with limited training data.

2.5.2. Model constraints. First, the transition matrix \( A \) was constrained to the Bakis model commonly used for word modelling in automatic speech processing (ASR) (Schukat-Talamazzini 1995). This model admits non-zero elements only on the upper triangular part of the transition matrix—specifically on the main as well as first and second secondary diagonal. Thus this particular structure implies only ‘loop’, ‘next’ and ‘skip’ transitions and supresses state transitions on the remaining upper half. Preliminary results, using a full upper triangular matrix instead, revealed small probabilities, i.e. very unlikely state transitions, for these suppressed transitions throughout all finger models. This suggests only minor contributions to the signal characteristics described by that model. A full, ergodic transition matrix generated...
poor decoding performances in general. Recently, a similar finding was presented by Lederman and Tabrikian for EEG data (Lederman and Tabrikian 2012).

Given five states containing one Gaussian, the unleashed model would involve $4 \times 1235$ degrees of freedom, while the Bakis model uses $4 \times 1219$ for an exemplary selection of 15 channels. The significant increase of decoding performance by dropping only a few parameters seems interesting with respect to the nature of the underlying signal structure and sheds remarkable light on the importance of feature selection for HMMs. An unleashed model using all 64 channels would have $4 \times 20835$ free parameters.

2.5.3. Model structure. Experiments for all feature spaces have been conducted in which the optimal number of states $Q$ per HMM and the number of mixture distribution components $M$ per state have been examined by grid search. Since increasing the degrees of freedom will degrade the quality of model estimation, high decoding rates have mainly been obtained for a small number of states containing only a few mixture components. Based on this search a model structure of $Q = 5$ states and $M = 1$ mixture components has been chosen as the best compromise. Note, that the actual optimum might vary slightly depending on the particular subject, the feature space and even on the session, which is in line with other studies (Obermanier et al 1999, 2001a, 2001b).

2.5.4. Initialization. The initial parameters of the multivariate probability densities have been set using a $k$-means clustering algorithm (Duda et al 2001). The algorithm temporally clusters the vectors of the $O \times T$ feature matrix $O_T$. The rows of this matrix were extended by a scaled temporal counter adding a time stamp to each feature vector. This increases the probability of time-continuous state intervals in the cluster output. Then, for each of the $Q \cdot M$ clusters, the mean as well as the full covariance matrix have been calculated excluding the additional time stamp. Finally, to assign these mixture distributions to specific states, the clusters were sorted according to the mean of their corresponding time stamp values. This temporal order is suggested by the constrained forward structure of the transition matrix. The transition matrix is initialized with probability values on the diagonal that are weighted to be $c$-times as likely as probabilities off the diagonal. The constant $c$ is being calculated as the quotient of the number of samples for a feature $T$ and the number of states $Q$. In case of multiple mixture components ($M > 1$) the mixture weight matrix was initialized using the normalized number of sample vectors assigned to each cluster.

2.5.5. Implementation. The basic implementation has been derived from the open source functionalities on HMMs provided by Kevin Murphy from the University of British Columbia (Murphy 2005). This framework, its modifications as well as the entire code developed for this study were implemented using MATLAB R2012b from MathWorks.

2.5.6. Support vector machines. As a gold-standard reference we used a one-vs-one SVM for multi-class classification. In our study, training samples come in paired assignments $(x_i, y_i), i = 1, \ldots, N$, where $x_i$ is a $1 \times (T \cdot O)$ feature vector derived from the ECoG data for one trial with its class assignment $y_i = \{-1, 1\}$. Equation (4) describes the classification rule of the trained classifier for new samples $x$. A detailed derivation of this decision rule and the rationale behind SVM-classification is given e.g. by Hastie et al (2009):

$$F(x) = \text{sgn} \left( \sum_i \alpha_i y_i K(x_i, x) + b_0 \right)$$

The $\alpha_i \in [0, C]$ are weights for the training vectors and are estimated during the training process. Training vectors with weights $\alpha_i \neq 0$ contribute to the solution of the classification problem and are called support vectors. The offset of the separating hyperplane from the origin is proportional to $b_0$. The constant $C$ generally controls the trade-off between classification error on the training set and smoothness of the decision boundary. The numbers of training samples per session (table 1) were in a similar range compared to the resulting dimension of the SVM feature space $T \cdot O$ ($\sim$130–400). Influenced by this fact, the decoding accuracy of the SVM revealed almost no sensitivity with respect to regularization. Since grid search yielded a stable plateau of high performance for $C > 1$, the regularization constant was fixed to $C = 1000$. The linear kernel, which has been used throughout the study, is denoted by $K(x_i, x)$. We further used the LIBSVM (www.csie.ntu.edu.tw/~cjlin/libsvm/) package for Matlab developed by Chih-Chung Chang and Chih-Jen Lin (Chang and Lin 2011).

2.6. Considerations on testing.

In order to evaluate the findings presented in the following sections, a five-fold cross-validation (CV) scheme was used (Lemm et al 2011). The assignment of a trial to a specific fold was performed using uniform random permutations.

Each CV procedure was repeated 30 times to average out fluctuations in the resulting recognition rates arising from random partitioning of the CV sets. The number of repetitions was chosen as a compromise to obtain stable estimates of the generalization error, but also to avoid massive computational load. The training data was balanced between all classes to treat no class preferentially and to avoid the need for class prior probabilities. Feature subset selection and estimation of the parameters of each classifier was performed only on data sets allocated for training in order to guarantee a fair estimate of the generalization error. Due to prohibitive computational effort, the feature spaces have been parameterized only once for each subject and session and are hence based on all samples of each subject-specific data set. In this context preliminary experiments indicated minimal differences to the decoding performance (see table 2) when fixing the feature definitions based on the training sets only. These experiments used nested CV with an additional validation set for the feature space parameters.
Figure 3. Optimal parameters of both feature spaces for each subject and session. (a) Location and width for ROIs containing most discriminative information for LFTD features. (b) Optimal high cut-off frequency for the spectral lowpass filter to extract LFTD features. (c) Location and width for ROIs containing most discriminative information for HG features. (d) Low and high cut-off frequency for the optimal HG frequency band.

Table 2. HMM-decoding accuracies (in %) for parameter optimization on full data and on training subsets. Accuracies are given for subject 1—session 1. Parameter optimization was performed on the full dataset (standard-CV routine) as well as the training subset (nested-CV routine).

| Parameter optimization  | LFTD | HG    | LFTD + HG |
|--------------------------|------|-------|-----------|
| Full dataset             | 80.7 ± 0.3 | 97.9 ± 0.1 | 98.4 ± 0.1 |
| Training subset          | 81.1 ± 1.0  | 97.8 ± 0.3  | 98.0 ± 0.2  |

3. Results

3.1. Feature and channel selection

In order to identify optimal parameters for feature extraction a grid search for each subject and session was performed. For this optimization a preliminary feature selection was applied aiming at a fixed subset size of $O = 8$, which was found to be within the range of optimal subset sizes (see figure 6 and table 3). Results for all sessions are illustrated by the bar plots in figure 3. A sensitivity of the decoding rate has been found for temporal parameters such as location and width for the time windows (ROIs) containing the most discriminative information about LFTD or HG features (figures 3(a) and (c)).

The highest impact on accuracy was found for the HG ROI (mean optimal ranges: S1: $[-187.5, 337.5]$ ms; S3: $[-400, 475]$ ms; S4: $[-280, 420]$ ms).

Further sensitivity is given for spectral parameters namely the high cut-off frequency for the LFTD lowpass filter and the low and high cut-off for the HG bandpass (figures 3(b) and (d)). For the former, optimal frequencies ranged from 11–30 Hz and the optimal frequency bands for the latter all lay within 60–300 Hz, depending on the subject. There was hardly any impact on the decoding rate when changing the sliding window size and window overlap for the HG features. These were hence set to fixed values. According to these parameter settings, the mean numbers of samples $T$ were 26.5 (subject 1), 21.3 (subject 2), 31 (subject 3) and 22 (subject 4).

An example of the grid search procedure is shown in supplementary figures S15–S17 (available at stacks.iop.org/JNE/10/056020/mmedia). The plots show varying decoding rates for the S1 session 1 within individual parameter subspaces. The optimal parameters illustrated in figure 3 represent a single point in these spaces at peaking decoding rate. Based on these parameter settings, the mean numbers of samples $T$ were 26.5 (subject 1), 21.3 (subject 2), 31 (subject 3) and 22 (subject 4).

Using the resulting parameterization, subsequent CV runs sought to identify the optimal subset by tuning $O$ based on the training set of the current fold. Figure 5 illustrates the rate at which a particular channel was selected for each subject.
Figure 4. The dashed grey time course shows a raw ECoG signal for a typical trial (S1, session 2, channel 23, middle finger). The extracted feature sequences for LFTD (cut-off at 24 Hz) and HG (frequency band [85 Hz, 270 Hz]) are overlaid as solid lines. Each HG feature has been assigned to the time point at the centre of its corresponding FFT window.

Figure 5. Channel maps for all four subjects. The channels are arranged according to the electrode grid and coloured with respect to their mean selection rate in repeated CV-runs. Channels coloured like the top of the colour bar tend to have a higher probability of being selected.

It evaluates the optimal subsets $O$ across all sessions, folds and several CV repetitions. Results are exemplarily shown for high gamma features. Full detail on subject-specific grid placement and channel maps for LFTD and HG features can be found in supplementary figures S11–S14 (available at stacks.iop.org/JNE/10/056020/mmedia).

The channel maps reveal localized regions of highly informative channels (dark red) comprising only a small subset within the full grid. While these localizations are distinct and focused for high gamma features, informative channels spread wider across the electrode grid for LFTD features. This finding is in line with an earlier study done by Kubanek et al (2009) and implies that selected subsets for LFTD features tend to be less stable across CV runs. Combined subsets were selected from both feature spaces. The selections indicated a preference for a higher proportion of LFTD features (table 3). This optimal proportion was determined by grid search.

The optimal absolute subset size $O$ was selected by successively adding channels to the subset as defined by our selection procedure and identifying the peak performance. For the high gamma case, figure 6 already shows a sharp increase in decoding performance for a subset size below five. Performance then increases less rapidly until it reaches a maximum usually in the range of 6–8 channels for the HG features (see table 3 for detailed information on the results for the other features). This behaviour again reflects the fact that a lot of information is found in a small number of HG sequences.

Increasing the subset size beyond this optimal size leads to a decrease in performance for both SVM and HMM. This coincides with an increase of the number of free parameters in the classifier model depending quadratically on the number of channels for HMMs (figure 6, bottom).

3.2. Decoding performances

Based on the configuration of the feature spaces and feature selection, a combination space containing both LFTD and HG features yielded the best decoding accuracy across all sessions and subjects. Figure 7 summarizes the highest HMM performances for each subject averaged across all sessions and compares them with the corresponding SVM accuracies. Except for subject 2, these accuracies were close to or even exceeded 90% correct classifications. However, figure 8 shows that accuracies obtained only from the HG space come very close to this combined case, implying that LFTD features contribute limited new information. Using time domain features exclusively decreased performance in most cases. An exception is given by subject 2 where LFTD features outperformed HG features. This may be explained by the wider spread of LFTD features across the cortex and the suboptimal grid placement (little coverage of motor cortex, supplementary figure S12 available at stacks.iop.org/JNE/10/056020/mmedia). The channel maps show informative channels at the very margin of the grid. This suggests that grid placement was not always optimal for capturing the most informative electrodes.

Differences in decoding accuracy compared to the SVM were found to be small (table 4) and revealed no clear superiority of any classifier (figure 9). Excluding subject 2, the mean performance difference (HMM-SVM) was −0.53% for LFTD, 0.1% for HG and −0.23% for combined features. Accuracy trends for individual subjects are similar across sessions and vary less than across subjects (figure 9). Note that, irrespective of the suboptimal grid placement, subject 2 took part in four experimental sessions. Here, subject-specific variables, e.g. grid placement, were important determinants of decoding success. Thus we interpret our results at the subject level and conclude that, in three out of four subjects, SVM and HMM decoding of finger movements led to comparable results.
Figure 6. Influence of channel selection on the HMM decoding accuracy using high gamma features only (S1, session 1). Changes in decoding accuracy when successively adding channels to the feature set applying either random or DB-index channel selection (top). Number of free HMM parameters for each number of channels (bottom).

Table 3. Number of selected channels averaged across sessions. The averaged number of selected channels and the corresponding standard derivations are given for all subjects and all feature spaces. For the combined space, square brackets denote the corresponding breakdown into its components. The average is taken across all sessions and CV repetitions.

| Subject | LFTD  | HG    | LFTD + HG |
|---------|-------|-------|-----------|
| 1       | 10.0 ± 0.0 | 6.0 ± 0.0 | 16.5 ± 0.5 (10.0 ± 0.0) + (6.5 ± 0.7) |
| 2       | 6.5 ± 1.9  | 7.3 ± 2.1 | 12.8 ± 3.4 (7.0 ± 4.5) + (5.8 ± 1.5) |
| 3       | 10.0 ± 0.0 | 8.0 ± 0.0 | 14.0 ± 1.0 (6.0 ± 1.4) + (8.0 ± 0.0) |
| 4       | 10.0 ± 0.0 | 6.0 ± 1.4 | 16.0 ± 0.7 (9.5 ± 0.7) + (6.5 ± 0.7) |

A clear superiority for SVM classification was only found for subject 2, for whom the absolute decoding performances were significantly worse than those of the other three subjects.

For the single spaces, LFTD and HG, table 4 shows the mean performance loss for the unleashed HMM. That refers to a case wherein the Bakis model constraints are relaxed to the full model. For all except two sessions of subject 2 (the full model had a higher performance of 0.2% for session 3 and 0.7% for session 4), the Bakis model outperformed the full model. The results indicate a higher benefit for the LFTD features with an up to 8.8% performance gain for subject 3, session 2. Higher accuracies were consistently achieved for feature subsets deviating from the optimal size. An example is illustrated for LFTD channel selection (subject 3, session 2) in supplementary figure S19 (available at stacks.iop.org/JNE/10/056020/mmedia). The plot compares the Bakis and full model similar to the procedure in figure 6. While table 4 lists only the improvements for the optimal number of channels, supplementary figure S19 indicates that the performance gain may be even higher when deviating from the optimal settings.

Finally, figure 10 illustrates the accuracy decompositions into true positive rates for each finger. The results reveal similar trends across subjects and feature spaces. While the thumb and little finger provided the fewest training samples, their true positive rates tend to be the highest among all fingers. An exception is given for LFTD features for subject 2. Equivalent plots for the SVM are shown in supplementary figure S110 (available at stacks.iop.org/JNE/10/056020/mmedia).

For the classification of one feature sequence including all T samples, the HMM Matlab implementation took on average (3.522 ± 0.036) ms and the SVM implementation in C (0.0281 ± 0.0034) ms. The mean computation time for a feature sequence was 0.2 ms ± 2.1μs (LFTD), 1.72 ms ± 37.9 μs (HG) and 1.79 ms ± 20.0 μs (combined space). The results were obtained with an Intel Core i5-2500 K @ 3.4 GHz, 16 GB RAM, Win 7 Pro 64-Bit, Matlab 2012b for ten selected channels.
Figure 7. Decoding accuracies (in %) for HMM (green) and SVM (beige) using LFTD + HG feature space (best case). Average performances across all sessions are shown for all subjects (S1–S4) as a mean result of a $30 \times 5$ CV procedure with their corresponding error bars.

Figure 8. Decoding accuracies (in %) for the HMMs for all feature spaces and sessions Bi. Performances are averaged across 30 repetitions of the five-fold CV procedure and displayed with their corresponding error bars.

4. Discussion

An evaluation of feature space configurations revealed a sensitivity of the decoding performance to the precise parameterization of individual feature spaces. Apart from these general aspects, the grid search primarily allows for individual parameter deviations across subjects and sessions. Importantly, a sensitivity has been identified for the particular time interval of interest used for the sliding window approach to extract HG features. This indicates varying temporal dynamics across subjects and may favour the application of HMMs in online applications. This sensitivity suggests that adapting to these temporal dynamics has an influence on the degree to which information can be deduced from the data by both HMM and SVM. However, our investigations indicate that HMMs are less sensitive to changes of the ROI and give nearly stable performances when changing the window overlap. As a model property, the HMM is capable of modelling informative sequences of different lengths and rates. The latter characteristic originates from the explicit loop transition within a particular model. This means the HMM may cover uninformative samples of a temporal sequence with an idle state, before the informative part starts. Underlying temporal processes that may persist longer or shorter are modelled by more or less repeating basic model states. Lacking any time-warp properties, an SVM would expect a fixed sequence length and information rate, meaning a fixed association between a particular feature dimension $o_i$ and a specific point in time $t_i$. This highlights the strict comparability of information within one single feature dimension explicitly required by the SVM. This leads to the prediction of benefits for HMMs in online applications where the time for the button press or other
Figure 9. Differences in decoding accuracy (in %) for the HMMs with respect to the SVM reference classifier (HMM-SVM). Results are displayed for all sessions Bi and feature spaces.

Figure 10. Mean HMM decoding rates decomposed into true positive rates for each finger. Results are shown for each feature space and averaged across sessions and CV repetitions.

Table 4. Final decoding accuracies (in %) listed for HMM (left) and SVM (right). Accuracies for the Bakis model are given for each subject-session and feature space by its mean and standard error across all 30 repetitions of the CV procedure. Additionally, for each subject and feature space the mean across all sessions as well as the mean loss in performance for LFTD and HG when using an unconstrained HMM is shown.

| Subject–session | HMM          | SVM          |
|-----------------|--------------|--------------|
|                 | LFTD         | HG           | LFTD + HG    | LFTD         | HG           | LFTD + HG    |
| S1–B1 (%)       | 80.7 ± 0.3   | 97.9 ± 0.1   | 98.4 ± 0.1   | 85.8 ± 0.4   | 97.2 ± 0.2   | 98.3 ± 0.1   |
| S1–B2 (%)       | 78.8 ± 0.6   | 92.4 ± 0.4   | 95.3 ± 0.3   | 84.9 ± 0.5   | 93.4 ± 0.4   | 96.1 ± 0.3   |
| Mean (%)        | 79.8 ± 0.5   | 95.2 ± 0.3   | 96.9 ± 0.2   | 85.4 ± 0.5   | 95.3 ± 0.3   | 97.2 ± 0.2   |
| Unconstrained model (%) | −3.1 ± 0.5   | −0.9 ± 0.3   | −     | −           | −           | −           |
| S2–B1 (%)       | 44.0 ± 1.0   | 35.9 ± 0.7   | 44.4 ± 0.6   | 51.4 ± 0.8   | 38.0 ± 0.8   | 55.7 ± 0.8   |
| S2–B2 (%)       | 50.2 ± 1.0   | 48.1 ± 0.7   | 52.4 ± 0.8   | 55.5 ± 0.7   | 48.0 ± 0.6   | 60.2 ± 0.8   |
| S2–B3 (%)       | 40.2 ± 0.7   | 36.3 ± 0.8   | 40.1 ± 0.8   | 49.6 ± 0.7   | 39.5 ± 0.9   | 51.5 ± 0.8   |
| S2–B4 (%)       | 30.9 ± 0.7   | 36.4 ± 0.9   | 38.4 ± 0.9   | 39.1 ± 0.7   | 38.0 ± 1.0   | 42.7 ± 0.9   |
| Mean (%)        | 41.2 ± 0.9   | 39.2 ± 0.8   | 43.8 ± 0.8   | 48.9 ± 0.7   | 40.9 ± 0.9   | 52.5 ± 0.9   |
| Unconstrained model (%) | −2.2 ± 0.9   | −0.4 ± 0.8   | −     | −           | −           | −           |
| S3–B1 (%)       | 64.6 ± 0.7   | 85.1 ± 0.4   | 85.8 ± 0.4   | 61.2 ± 0.7   | 85.4 ± 0.5   | 88.8 ± 0.4   |
| S3–B2 (%)       | 80.1 ± 0.5   | 90.5 ± 0.3   | 93.7 ± 0.3   | 85.0 ± 0.4   | 91.4 ± 0.3   | 96.0 ± 0.3   |
| Mean (%)        | 72.4 ± 0.6   | 87.8 ± 0.4   | 89.8 ± 0.4   | 73.1 ± 0.6   | 88.4 ± 0.4   | 92.4 ± 0.3   |
| Unconstrained model (%) | −5.9 ± 0.7   | −2.5 ± 0.4   | −     | −           | −           | −           |
| S4–B1 (%)       | 71.4 ± 0.5   | 83.5 ± 0.5   | 87.2 ± 0.4   | 68.5 ± 0.6   | 82.1 ± 0.4   | 86.7 ± 0.4   |
| S4–B2 (%)       | 69.7 ± 0.7   | 83.1 ± 0.3   | 87.9 ± 0.4   | 64.3 ± 0.6   | 82.3 ± 0.3   | 84.4 ± 0.4   |
| Mean (%)        | 70.6 ± 0.7   | 83.3 ± 0.3   | 87.6 ± 0.4   | 66.4 ± 0.6   | 82.2 ± 0.3   | 85.6 ± 0.4   |
| Unconstrained model (%) | −1.1 ± 0.6   | −3.1 ± 0.5   | −     | −           | −           | −           |

11
events may not be known beforehand. The classifier may act more robustly on activity that is not perfectly time locked. Imaging paradigms for motor imagery or similar experiments may provide scenarios in which a subject carries out an action for an unspecified time interval. Given that varying temporal extent is likely, the need for training the SVM for the entire range of possible ROI locations and widths is not desirable.

The SVM characteristics also imply a need to await the whole sequence to make a decision for a class. In contrast, the Viterbi algorithm allows HMMs to state at each incoming feature vector how likely an assignment to a particular class is. In online applications this striking advantage provides more prompt classification and may reduce computational overhead, if only the baseline is employed. There is no need to await further samples for a decision. Note that this emphasizes the implicit adaptation of HMMs to real-time applications: while the SVM needs to classify the whole new subsequence again after each incoming sample, HMMs just update their current prediction according to the incoming information. They do not start from scratch. The implementation used here did not make use of this online feature, since only the classification of the entire trial was evaluated. Another reason for the SVM being faster than the HMM is given by the non-optimized implementation in Matlab code of the latter. The same applies to the feature extraction. Existing high performance libraries for these standard signal processing techniques or even a hardware field-programmable gate array implementation may significantly speed up the implementation. The average classification time for the HMMs implies a maximal possible signal rate of about 285 Hz when classifying after each sample. Thus, in terms of classification, even with the current implementation real-time is perfectly feasible for most cases. More sophisticated implementations are used in the field of speech processing where classifications have to be performed on the basis of raw data with very high sampling rates of up to 40 kHz.

Second, tuning the features to an informative subset yielded less dependence on the particular classifier for the presented offline case. Preliminary analysis using compromised parameter sets for the feature spaces as well as feature subset size led, instead, to a higher performance gap between both classifiers. Adapting to each subject individually, a strict superiority was not observed for the HMMs versus the SVM in subjects 1 and 3. For subject 4, the HMM actually outperformed the SVM in all cases. Interestingly, this coincides with the grid resolution, which was twice as high for this subject, supporting the importance of the degree and spatial precision of motor cortex coverage. For subject 2 decoding performances were significantly worse in comparison with the other subjects. This is probably due to the grid placement not being optimal for the classification (supplementary figure SI2 available at stacks.iop.org/JNE/10/056020/mmedia). Figure 5 shows no distinct informative region and channels selected at a high rate are located at the outer boundary of the grid rather than over motor cortices. This hypothesis is further supported by the decoding accuracies which are higher for LFTD than for HG features in subject 2. Since informative regions were spread more widely across the cortex for subject 2 (Kubanek et al. 2009), regions exhibiting critical high gamma motor region activity may have been located outside the actual grid.

While the results suggest that high feature quality is the main factor for high performances in this ECoG data set, two key points have to be emphasized for the application of HMMs with respect to the present finger classification problem. First, due to a high number of free model parameters, feature selection is essential for achieving high decoding accuracy. Note, that the SVM exhibits fewer free parameters to be estimated during training. These parameters mainly correspond to the weights \( \alpha \). The number of these directly depends on the number of training samples. The functional relationship between features and labels even depends on the non-zero weights of the support vectors only. Thus, SVM exhibits higher robustness to variations of the subset size. Second, imposing constraints on the HMM model structure was found to be important for achieving high decoding accuracies. As stated in the methods section, the introduction of the Bakis model does not substantially influence the number of free parameters. The subset size has a much higher impact (figure 6) on the overall number. Nevertheless, a performance gain of up to 8.8% per session or 5.9% mean per subject across all sessions was achieved for the LFTD features. This suggests that improvements may not be due to a mere reduction in the absolute number of parameters, but also due to better compliance of the model with the underlying temporal data structure, i.e. the evolution of feature characteristics over time. For the LFTD space, information is encoded in the signal phase and hence the temporal structure. The fact that the Bakis model constrains this temporal structure hence entails interesting implications. Higher overall accuracies for the HG space, along with only minor benefits from the Bakis model, indicate that the clear advantages of model constraints mainly arise if the relevant information does not appear as prominent in the data. This is further supported by a higher increase in HMM performance when not using the optimal subset size for LFTD features (see supplementary figure SI9 available at stacks.iop.org/JNE/10/056020/mmedia). This suggests that the Bakis approach tends to be more robust with respect to deviations from the optimal parameter setting.

Finally, it should be mentioned that the current test procedure does not fully allow for non-stationary behaviour. That means that the random trial selection does not take temporal evolution of signal properties into account. These might change and blur the feature space representation over time (Lemm et al. 2011). In this context and depending on the extent of non-stationarity, the presented results may slightly overestimate the real decoding rates. This applies for both SVM and HMM. Non-stationarity should be addressed in an online setup by updating the HMM on-the-fly. Training data should only be used from past recordings and the signal characteristics of future test set data remain unknown. While online model updates are generally unproblematic for the HMM training procedure, investigations on this side are still needed.
5. Conclusions

For a defined offline case, we have shown that high decoding rates can be achieved with both the HMM and SVM classifiers in most subjects. This suggests that the general choice of classifier is of less importance once HMM robustness is increased by introducing model constraints. The performance gap between both machine learning methods is sensitive to feature extraction and particularly electrode subset selection, raising interesting implications for future research. This study restricted HMMs to a rigid time window with a known event—a scenario that suits SVM requirements. This condition may be relaxed and classification improved by use of the HMM time-warp property (Sun et al 1994). These allow for the same spatio-temporal pattern of brain activity occurring at different temporal rates.

By exploiting the actual potential of such models, we expect them to be individually adapted to specific BCI applications. The idea of adaptive online learning may be implemented for non-stationary signal properties.

Future research may encompass extended and additional model constraints for special cases, as successfully used in ASR, such as fixed transition probabilities or the sharing of mixture distributions among different states. Finally, incorporation of other sources of prior knowledge and context awareness may further optimize decoding accuracy.

Acknowledgments

The authors acknowledge the work of Fanny Quandt who substantially contributed to the acquisition of the ECoG data. Supported by NINDS grant NS21135, Land-Sachsen-Anhalt grant MK48-2009/003 and ECHORD 231143. This work is partly funded by the German Ministry of Education and Research (BMBF) within the Forschungscampus STIMULATE (grant 03FO16102A).

References

Acharya M, Fifer M, Benz H, Crone N and Thakor N 2010 Electroencephalographic amplitude predicts finger positions during slow grasping motions of the hand J. Neural Eng. 7 13
Ball T, Schulze-Bonhage A, Aertsen A and Mehring C 2009 Differential representation of arm movement direction in relation to cortical anatomy and function J. Neural Eng. 6 016
Bhattacharyya A 1943 On a measure of divergence between two statistical populations defined by their probability distributions Bull. Calcutta Math. Soc. 35 99–1909
Chang C-C and Lin C-J 2011 LIBSVM: a library for support vector machines ACM Tran. Intell. Syst. Technol. 2 1–27
Cincotti F et al 2003 Comparison of different feature classifiers for brain computer interfaces Proc. 1st Int. IEEE EMBS Conf. on Neural Engineering (Capri, Italy) pp 645–7
Crone N, Miglioretti D, Gordon B and Lesser R 1998 Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis II, event-related synchronization in the gamma bands Brain 121 2301–15
Davies D and Bouldin D 1979 A cluster separation measure IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 224–7
Demirer R, Ozerdem M and Bayrak C 2009 Classification of imaginary movements in ECoG with a hybrid approach based on multi-dimensional Hilbert-SVM solution J. Neurosci. Methods 178 214–21
Dempster A, Laird N and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm J. R. Stat. Soc. B 39 1–38
Dethier J, Gilja V, Nuyujukian P, Elassaad S A, Shenoy K V and Boahen K 2011 Spiking neural network decoder for brain-machine interfaces EMBS’11: Proc. IEEE 5th Int. Conf. on Neural Engineering in Medicine and Biological Society (Cancun, Mexico) pp 396–9
Duda R O, Hart P E and Stork D G 2001 Pattern Classification 2nd edn (New York: Interscience)
Fisher A R 1936 The use of multiple measurements in taxonomic problems Ann. Eugen. 7 179–88
Ganguly K and Carmena J M 2010 Neural correlates of skill acquisition with a cortical brain-machine interface J. Motor Behav. 42 355–60
Graimann B, Huggins J, Levine S and Pfurtscheller G 2004 Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis IEEE Trans. Biomed. Eng. 51 954–62
Guyon I and Elisseeff A 2003 An introduction to variable and feature selection J. Mach. Learn. Res. 3 1157–82
Guyon I, Weston J, Barnhill S and Vapnik V 2002 Gene selection for cancer classification using support vector machines J. Mach. Learn. 46 389–422
Hastie T, Tibshirani R and Friedman J 2009 The Elements of Statistical Learning 2nd edn (New York: Springer)
Hoffmann U, Vesin J and Diserens K 2007 An efficient P300-based brain-computer interface for disabled subjects J. Neurosci. Methods 167 115–25
Kubanek J, Miller K, Ojemann J, Wolpaw J and Schalk G 2009 Decoding flexion of individual fingers using electrocorticographic signals in humans J. Neural Eng. 6 14
Lederman D and Tabrikian J 2012 Classification of multichannel EEG patterns using parallel hidden Markov models Med. Biol. Eng. Comput. 50 319–28
Lee H and Choi S 2003 PCA+HMM-sVSM for EEG pattern classification 7th Proc. Int. Symp. on Signal Processing and Its Applications, vol I pp 541–4
Lemm S, Blankertz B, Dickhaus T and Mueller K-R 2011 Introduction to machine learning for brain imaging NeuroImage 56 387–99
Liang N and Bougrain L 2009 Decoding finger flexion using amplitude modulation from band-specific ECoG ESANN: Eur. Symp. on Artificial Neural Networks pp 467–72
Liu C, Zhao H B, Li C S and Wang H 2010 Classification of ECoG motor imagery tasks based on CSP and SVM BMET: 3rd Int. Conf. on Biomedical Engineering and Informatics vol 2 pp 804–7
Lotte F, Congedo M, Lecuyer A, Lamarche F and Arnaldi B 2007 A review of classification algorithms for EEG-based brain-computer interfaces J. Neural Eng. 4 R1–R13
Makeig S 1993 Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones Electroencephalogr. Clin. Neurophysiol. 86 283–93
Murphy K 2005 Hidden Markov model (HMM) toolbox for matlab www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html
Obermaier B, Guger C, Neuper C and Pfurtscheller G 2001a Hidden Markov models used for the offline classification of EEG data Pattern Recognit. Lett. 22 1299–309
Obermaier B, Guger C and Pfurtscheller G 1999 Hidden Markov models used for the offline classification of EEG data Biomed. Tech./Biomed. Eng. 44 158–62
Obermaier B, Neuper C, Guger C and Pfurtscheller G 2001b Information transfer rate in a five-classes brain-computer interface IEEE Trans. Neural Syst. Rehabil. Eng. 9 283–8
Palaniappan R, Chanan R and Syan S 2009 Current practices in electroencephalogram based brain-computer interfaces J. Neural Eng. 10 (2013) 056020
Encyclopedia of Information Science and Technology, II
(Hershey, PA: IGI Global) pp 888–901
Pasqual-Marqui R D, Michel C M and Lehmann D 1995
Segmentation of brain electrical activity into microstates: model estimation and validation
IEEE Trans. Biomed. Eng. 42 658–65
Pechenizkiy M 2005 The impact of feature extraction on the performance of a classifier: kNN, Naive Bayes and C4.5
Proc. 18th Canadian Society Conf. on Advances in Artificial Intelligence (Berlin: Springer) pp 268–79
Penfield W and Boldrey E 1937 Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation
Brain 60 389–443
Picton T W, Lins O G and Scherg M 1995 The recording and analysis of event-related potentials
Handbook of Neuropsychology ed F Boller and J Grafman (Amsterdam: Elsevier) vol 10 pp 3–73
Pistohl T, Ball T, Schulze-Bonhage A, Aertsen A and Mehring C 2008 Prediction of arm movement trajectories from ECoG-recordings in humans
J. Neurosci. Methods 167 105–14
Quandt F, Reichert C, Hinrichs H, Heinze H, Knight R and Rieger J 2012 Single trial discrimination of individual finger movements on one hand: a combined MEG and EEG study
NeuroImage 59 3316–24
Rabiner L 1989 A tutorial on hidden Markov models and selected applications in speech recognition
Proc. IEEE 77 257–86
Schalk G 2010 Can electrocorticography (ECoG) support robust and powerful brain-computer interfaces
Front. Neuroeng. 3 1
Schukat-Talamazzini E G 1995 Automatische Spracherkennung—Grundlagen, statistische Modelle und effiziente Algorithmen
Künstliche Intelligenz (Braunschweig: Vieweg)
Shenoy P, Miller K, Ojemann J and Rao R 2007 Finger movement classification for an electrocorticographic BCI
CNE’07: 3rd Int. IEEE/EMBS Conf. on Neural Engineering pp 192–5
Sun D X, Deng L and Wu C F J 1994 State-dependent time warping in the trended hidden Markov model
Signal Proc. 39 263–75
Wissel T and Palaniappan R 2011 Considerations on strategies to improve EOG signal analysis
Int. J. Artif. Life Res. 6 6–21
Yu L and Liu H 2004 Efficient feature selection via analysis of relevance and redundancy
J. Mach. Learn. Res. 5 1205–24
Zhao H B, Liu C, Wang H and Li C S 2010 Classifying ECoG signals using probabilistic neural network
ICIE’10: WASE Int. Conf. on Information Engineering vol 1 pp 77–80