# Movement EEG Classification Using Parallel Hidden Markov Models

Martin Dobiáš, Jakub Šťastný
Department of Circuit Theory
Faculty of Electrotechnical Engineering, Czech Technical University in Prague
Technická 2, Prague 6, 166 27, Czech Republic
dobiama3@feld.cvut.cz, stastnj1@feld.cvut.cz

*Abstract*—In this contribution we examine the use and utility of parallel HMM classification in single-trial movement-EEG classification of index finger reaching and grasping movement. Parallel HMMs allow us to easily utilize the information contained in multiple channels. Using HMM classifier output in parallel from examined EEG channels we have been able to achieve as good a classification score as with single electrode results, further we do not rely on a single electrode giving persistently good results. Our parallel approach has the added benefit of not having to rely on small inter-session variability as it gives very good results with fewer classifier parameters being optimized. Without any classification optimization we can get a score improvement of 11.2% against randomly selected physiologically relevant electrode. If we use subject specific information we can further improve on the reference score by 1%, achieving a classification score of 84.2±0.7%.

*Keywords*—*EEG; BCI; parallel HMM; movement type classification.*

## I. Introduction

The classification of movement direction is of great significance in the field of Brain Computer Interface (BCI) research. There are many approaches to BCI control researched worldwide, using e.g., Visual Evoked Potentials, P300 event-related wave, or different kinds of voluntary mental activities. We focus on movement related activity, as controlling a BCI with movement related EEG feels very natural and only an imagination of the movement is sufficient to control the BCI [1]. Moreover movement EEG based systems can be designed as asynchronous ones, giving the user more freedom [1]. The ability to distinguish the direction of movement increases the number of recognizable states, thus increasing the information transfer rate. This is crucial as the existing BCIs use only a few distinct types of movement (mostly left/right hand or finger movement).

In EEG based BCI there are many ways to utilize information from multiple electrodes, e.g. the spatiotemporal development of signal power [2], channel extraction using Indepent Component Analysis (ICA) artifact suppression [3] or applying a spatial filter (either standard Laplacian or one generated based on a source separation algorithm) [4]. With Hidden Markov Models (HMMs), features from multiple channels are usually combined into single feature vector (multivariate HMM). This approach however relies on a single 1D model. With parallel HMMs information from multiple channel HMMs can be used to determine the final classifier decision.

For BCI the variability of the EEG signals is an ongoing problem. The EEG signal can vary greatly from electrode to electrode, from subject to subject and from session to session. Even though it makes sense to optimize channels and feature subsets [5], thus achieving a higher classification score. Though the optimization may last just one session it may provide a good starting point for the next session. On the other hand having too many parameters to optimize can pose problems for an online BCI application, which ideally needs to be quite robust as well as provide good results.

The parallel HMM method presented in this paper is intended to provide a way to efficiently utilize HMM-based classification without the need to accept some of the disadvantages of organizing the multichannel features into a single multivariate HMM. The method is similar to the one presented in [6], which is used here as well and further improved.

The experiments are performed on a database of EEG realizations of extension (reaching) and flexion (grasping) movements of right index finger. Classification of finger movements of the same limb is known as a complex and challenging task as the activated muscle mass is smaller than in e.g. arm movements and the different contralateral preponderance cannot be exploited as in different limb movements.
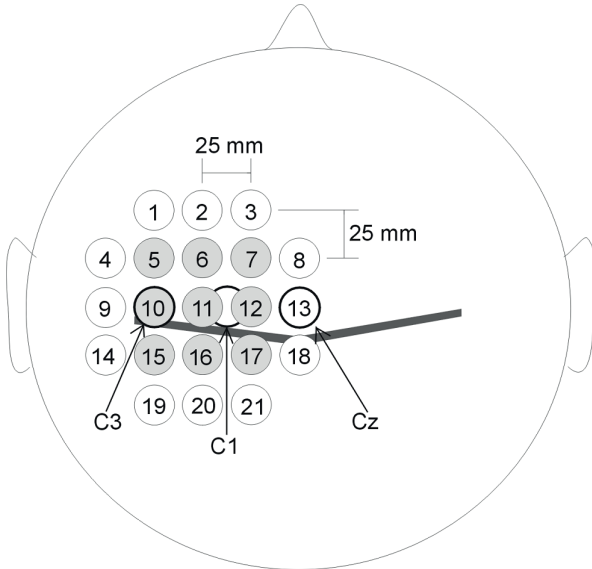
## II. Movement-Related EEG

Recorded EEG is a composition of two basic components: spontaneous activity and event-related responses. The event-related responses can be further divided into induced responses and evoked responses [1]. Induced responses are visible as changes in the power of subbands in the EEG signal and EEG has to be averaged in spectral domain to emphasize them.

Evoked responses are slow changes in the EEG phase-locked to the movement. Induced responses are of interest in this paper so we only deal with them.

For classification we use the induced oscillatory EEG activity most prominent in $\mu$ and $\beta$ bands, the

Fig. 1. Scalp electrode layout, according to [7].

Event Related Desynchronization (ERD) and Event Related Synchronization (ERS), see [7], [4].

$\mu$ERD starts about 1 second prior to the movement onset (see Fig. 2), usually localized to the contralateral sensorimotor area (C3/CP3/C4/CP4) [8], [9]. The desynchronisation accompanies even the mere motor imagery and is present in EEG of most adults. The majority of papers dealing with left/right hand movement recognition utilize the preponderance of $\mu$ERD contralaterally to the performed movement. The frequencies of the most movement reactive component are (unlike in shoulder vs. finger movements) almost identical (11.3±1.2 Hz and 11.2±1.4 Hz) for both the extension and flexion movements [7], and there is no significant difference in the duration of the movements. On electrodes 10,11, and 15 the ERD was found stronger in flexion-extension than in extension-flexion movements.

$\beta$ERS is displayed by central $\beta$ rhythms as a rebound in the form of a phasic synchronisation [10] after the movement onset. $\beta$ERS represents a post-movement rise of power in the $\beta$ band; the phenomenon peaks about 1 s after the movement onset, see Fig. 2. $\beta$ERS is most prominent over the contralateral hemisphere [10] and is focused slightly anterior of the largest $\mu$ERD. The amplitude of $\beta$ERS is larger in extension-flexion movement, however the activity is more focused on the scalp for flexion/extension movements [11]. The mean frequency of the most reactive component is slightly higher for extension-flexion movement (19.2±2.6 Hz) than in flexion-extension movement (18.2±2.6 Hz) [11].

## III. METHODS

### A. Subjects And Experimental Procedure

The EEG database was obtained from study of Stančák et al. [7]. The database contains EEG recordings of 11 right-handed healthy subjects voluntarily performing 120 brisk right index finger extension movements followed by a return to resting position (reaching) and 120 brisk flexion movements followed by return to resting position (grasping). The movements were performed at irregular intervals of 10–12 seconds. The subjects had their eyes closed during the recording. The EEG was recorded on 21 scalp electrodes placed over the contralateral sensorimotor area (see Fig. 1), $f_s = 256$ Hz; surface EMG electrodes [7] were used to mark the onset of the movement. The EEG was filtered using an 8-neighbor surface Laplacian filter. The data were segmented into 10 seconds epochs, 5 seconds preceding and 5 seconds following the onset of the EMG. Segments containing eye or muscle artifacts were removed. The described data processing was done by the authors of [7]. The average number of artifact free EEG realizations of each movement was 66.4±16.5 per subject.

### B. Feature Extraction

Various approaches for feature extraction have been used in the literature, e.g. Davis-Bouldin index, linear spectrum [4], various types of wavelets, power spectral density, or raw EGG signals [2], [12].

Based on [4], [5] we used linear FFT coefficients with 1 second window length (with 0.2 second step) as features giving us 1 Hz frequency resolution. This setup is shown to work best on similar data and classifier setup [4]. The $k$-th feature vector consists of $p = 36$ features covering frequency lines from 5 Hz to 40 Hz $\mathcal{F}_k = (f_k[1], \ldots, f_k[p])$ where $k$ is the time index. One movement EEG realization is described by feature matrix $\mathcal{F} \in \mathcal{R}^{36,45}$.

### C. Classification

Hidden Markov Models (HMM) have several advantages for EEG classification [4], among others utilization of the context information, ease of the interpretation, and ability to model the EEG.

The HMM classifier setup from [5] was used as reference. The used models have a left-to-right, no skips architecture with 4 emitting states, which is designed to capture the sequence of the movement-related EEG phases (resting EEG, desynchronization, post-movement synchronization, resting EEG) in the $\mu$ and $\beta$ bands [4] (see Fig. 2) generating $p$-dimensional Gaussian random processes with diagonal covariance matrices.

The design was extended by using multiple classifiers applied to several EEG channels in parallel. Only the 9 central electrodes as highlighted in Fig. 1 are used for movement type classification.

As the average number of available movement EEG realizations is relatively low, stratified 10-times repeated 5-fold Cross-Validation (CV) [5], [13] is utilized to obtain more reliable classification estimates.

In order to further increase the significance of the difference in classification score among various experiments, all presented experiments use identical initial conditions (same composition of training and testing sets) for classification.

Each of the classifiers in the parallel HMM setup is in each CV run initialized and trained separately
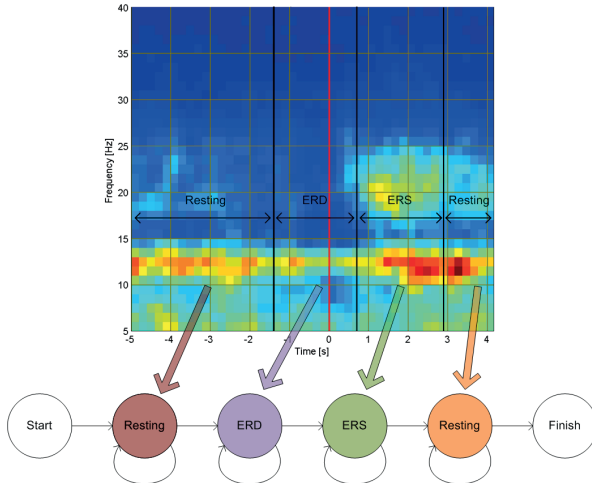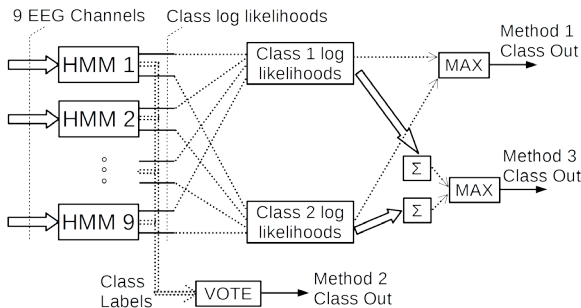
Fig. 2. HMM model of movement-related EEG.



Fig. 3. Depiction of the used classification methods.



Fig. 4. Dependence of the class score on number of channels (electrodes) used. Colored dashed lines represent the individual subjects. The solid black line is the average accuracy across the indicated number of channels.
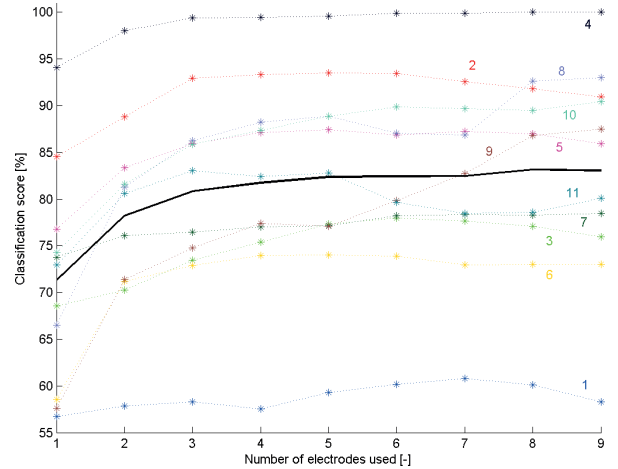


using its own training realization set. This is unlike [6], where the models are initialized using data from all channels and trained with channel specific data. Each testing realization is evaluated within all the respective EEG channel models and the output loglikelihood of each is noted. This process provides $nChannels \times nMovementTypes = 9 \times 2 = 18$ loglikelihood values on which the final class decision is based.

We applied four different decision methods (see Fig. 3):

1) *Max LogLike*: The maximum output loglikelihood decides the class of the testing realization. This method is equivalent to the one presented in [6].

2) *Max Count*: Majority voting is used to determine the output. Each channel is evaluated separately and votes equally for the final output.

3) *Max Sum LogLike*: This method uses a sum of class loglikelihoods to assign a label to a test realization. The label with the greatest total loglikelihood is selected. Summing loglikelihoods is e.g. used in [12] in order to compare scalar and multivariate HMM performance.

4) *Var Num Max Ele*: This method is similar to the *Max Sum LogLike*. However not all channels are summed, only $n$ best channel loglikelihoods are summed for each trial and movement type. This method gives subject-optimized results as we need to choose $n$ so as to get the best resulting classification accuracy. For $n = 1$ this method is equivalent to *Max LogLike*, for $n = nChannels$

to *Max Sum LogLike*.

## IV. RESULTS AND DISCUSSION

We used 5-40 Hz band as reference as it encompasses the physiologically relevant phenomena and in order to stay aligned with our previous research. Moreover this band was shown to be the best general choice [5] on a similar database.

The results of our experiments are summarized in Tab. I. The reference results for each subject as well as the electrode number on which they were acquired are given. The best method for each subject is highlighted. We can see that the *Max Sum LogLike* method gives results as good as the reference ones and *Var Num Max Ele* outperforms the reference results by 1%. If we choose a method for each subject, including the reference one, we can achieve a classification score of 85.6±0.7% (boldface values in Tab. I).

Fig. 4 further extends the result by showing how the increasing number of summed channel loglikelihoods influences the classification accuracy. The classification results generally saturate at 5 summed channels out of the 9 available.

Compared with [6] it is interesting to note that the algorithm proposed there does not work well with our data as the *Max LogLike* gives the worst results of all the methods tried. This can be well seen from Fig. 4, where we can see that using only the single best value ($n = 1$) does not guarantee good results and including multiple classifier outputs in the output label decision is advantageous.

There are several important differences between our dataset and that used in [6]. The database in [6] contains left/right motor imagery and up/down mental cursor movement imagery, while we use non-lateralized movements which is a more complicated task. We have added the ability to consider multiple channel outputs, not just the most confident one. In [6] the parallel HMM approach has decreased the error

| | 5-40 Hz band | | Max LogLike | Max Count | Max Sum LogLike | Var Num Max Ele | |
|---|---|---|---|---|---|---|---|
| Sub. | Score [%] | El. | Score [%] | Score [%] | Score [%] | Score [%] | Num. El. |
| 1 | 58.2±4.4 | 16 | 56.7±3.6 | 57.1±3.6 | 58.3±4.2 | **60.8±4.4** | 7 |
| 2 | 91.0±2.0 | 10 | 84.5±1.9 | 90.0±2.0 | 90.9±2.2 | **93.5±2.3** | 5 |
| 3 | 76.0±3.9 | 6 | 68.6±4.2 | 75.2±3.4 | 75.9±2.7 | **78.0±2.9** | 6 |
| 4 | 96.6±1.2 | 10 | 94.1±2.7 | **100.0±0.0** | **100.0±0.0** | **100.0±0.0** | 8 |
| 5 | 84.1±1.5 | 17 | 76.8±2.1 | 85.6±1.2 | 85.9±1.5 | **87.4±1.3** | 5 |
| 6 | **75.1±1.1** | 17 | 58.6±4.4 | 70.2±2.1 | 73.0±2.7 | 74.0±2.4 | 5 |
| 7 | **80.7±1.7** | 17 | 73.8±2.6 | 78.2±1.7 | 78.5±1.6 | 78.5±1.6 | 9 |
| 8 | **94.4±1.2** | 6 | 66.5±3.4 | 90.8±1.8 | 93.0±2.1 | 93.0±2.1 | 9 |
| 9 | **91.4±1.1** | 5 | 57.6±7.0 | 81.6±3.0 | 87.5±1.1 | 87.5±1.1 | 9 |
| 10 | 77.9±3.5 | 6 | 74.3±2.9 | 89.2±3.8 | **90.4±2.9** | **90.4±2.9** | 9 |
| 11 | **89.9±1.8** | 5 | 72.9±3.1 | 78.8±3.4 | 80.1±2.9 | 83.0±3.3 | 3 |
| all | 83.2±0.7 | best | 71.3±1.1 | 81.5±0.8 | 83.1±0.7 | **84.2±0.7** | |
| all | 71.9±0.4 | all | 85.6±0.7%* | | | | |

*Average of the best results from both the referential and the parallel HMM results (set in boldface).

rate by 2 - 10%. Quite a significant difference in error rated was reported depending on feature sets.

Based on our previous research in Blind Source Separation (BSS) techniques, parallel HMMs should be advantageous when combined with BSS. It is a common practice to use e.g. the first principal component in Principal Component Analysis or the best spatially fitting Independent component in Independent Component Analysis (ICA). However the other extracted components can contain some complementary information that would otherwise not be used, especially when using ICA to select components with some specific characteristic, like the strength of ERD and ERS. Then multiple components usually comply with the characteristic. Instead of decomposing the selected components back to scalp electrodes it might be more beneficial to classify the components directly using the parallel HMMs.

Though our voting approach has managed to outperform [6], the results aren't as good as the reference single electrode results. Having equivalent vote weights can work well on our database where all the channels can be expected to contain useful information, but in general it would be prudent to assign weights to the respective channels. Those could be based e.g. on electrode position or on the similarity of the spatial filter pattern extracted using BSS to the expected pattern (based on e.g. physiological knowledge).

## V. CONCLUSIONS

We have shown that parallel HMMS are a viable alternative to the single electrode classification approach. Parallel HMMs provide the option to utilize class information from multiple channels while keeping all the advantages of Hidden Markov Model classifiers that have consistently shown to outperform other classifiers in EEG classification tasks.

Using subject-optimized number of summed output loglikelihoods we were able to improve the reference results by 1%, achieving a score of 84.2±0.7%. With subject-specific choice of classification method a further 1.4% improvement can be reached.

## REFERENCES

[1] M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain, "The brain-computer interface cycle," *Journal of Neural Engineering*, vol. 6, no. 4, p. 041001, 2009.

[2] E. Lew, R. Chavarriaga, H. Zhang, M. Seeck, and J. Del Millan, "Self-paced movement intention detection from human brain signals: invasive and non-invasive EEG," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 3280–3283, IEEE, 2012.

[3] F. Quandt, C. Reichert, H. Hinrichs, H.-J. Heinze, R. T. Knight, and J. W. Rieger, "Single trial discrimination of individual finger movements on one hand: A combined MEG and EEG study.," *NeuroImage*, vol. 59, no. 4, pp. 3316–3324, 2012.

[4] J. Šťastný and P. Sovka, "High-resolution movement EEG classification," *Computational Intelligence and Neuroscience*, 2007. Article ID 54925, 12 pages, doi:10.1155/2007/54925.

[5] M. Dobiáš and J. Šťastný, "Subband optimization for EEG-based classification of movements of the same limb," in *Applied Electronics*, Pilsen, Czech Republic, September 2014.

[6] D. Lederman and J. Tabrikian, "Classification of multichannel EEG patterns using parallel hidden Markov models," *Medical & Biological Engineering & Computing*, vol. 50, no. 4, pp. 319–328, 2012.

[7] A. Stančák, "Event-related desynchronization of the mu rhythm in extension and flexion finger movements," *Clin Neurophysiol*, no. Clin Neurophysiol at the beginning of the 21st century (supplement to Clin Neurophysiol), pp. 636–650, 2000.

[8] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography*, vol. 12, no. 3, pp. 177–186, 2000.

[9] G. Pfurtscheller, K. Zalaudek, and C. Neuper, "Event-related beta synchronization after wrist, finger and thumb movement," *Electroencephalogr and Clin Neurophysiol*, no. 109, pp. 154–160, 1998.

[10] G. Pfurtscheller, A. Stančák, and C. Neuper, "Post-movement beta synchronization: a correlate of an idling motor area?," *Electroencephalogr and Clin Neurophysiol*, no. 98, pp. 281–293, 1996.

[11] Andrej Stančák, "The electroencephalographic beta synchronization following extension and flexion finger movements in humans," *Neuroscience Letters*, no. 284, pp. 41–44, 2000.

[12] S. Zhong and J. Ghosh, "HMMs and coupled HMMs for multi-channel EEG classification," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 1254–1159, 2002.

[13] M. Dobiáš, "Optimizing crossvalidation setup of movement EEG classifier," 18th International Student Conference on Electrical Engineering POSTER 2014, May 2014.