Probabilistic Modeling of Sensorimotor μ-Rhythms for Classification of Imaginary Hand Movements

(BCI Competition 2003 - Data Set III)

Steven Lemm, Christin Schäfer, and Gabriel Curio

***Abstract*— Brain computer interfaces (BCI) require effective on-line processing of EEG measurements, e.g., as a part of feedback systems. Here we present an algorithm for single trial on-line classification of imaginary left and right hand movements, based on time-frequency information derived from filtering EEG wideband raw data with causal Morlet wavelets which are adapted to individual EEG spectra. Since imaginary hand movements lead to perturbations of the ongoing pericentral μ-rhythm, we estimate probabilistic models for amplitude modulation in lower (10 Hz) and upper (20 Hz) frequency bands over the sensorimotor hand cortices both contra- and ipsilaterally to the imagined movements (i.e., at EEG channels C3 and C4). We use an integrative approach to accumulate over time evidence for the subject's unknown motor intention. Disclosure of test data labels after the competition showed this approach to succeed with an error rate as low as 10.7%.**

***Index Terms*— single trial classification, BCI, imaginary hand movement, μ-rhythm, Morlet wavelet, sensorimotor hand cortex**

I. Introduction

THE human perirolandic sensorimotor cortices show rhythmic macroscopic EEG oscillations (μ-rhythm) [1], with spectral peak energies around 10 Hz (localised predominantly over the postcentral somatosensory cortex) and 20 Hz (over the precentral motor cortex). Modulations of the μ-rhythm have been reported for different physiological manipulations, e.g., by motor activity, both actual and imagined [2]–[4], as well as by somatosensory stimulation [5]. Standard trial averages of μ-rhythm power show a sequence of attenuation, termed event-related desynchronization (ERD) [3], followed by a rebound (event-related synchronization: ERS) which often overshoots the pre-event baseline level [6].

In order to distinguish between single trials (STs) of left and right hand imaginary movements, we utilize the accompanying EEG μ-rhythm perturbation. Similar approaches were pursued in [7]–[9].

This paper describes a probabilistic approach, that has been successfully applied in the 2003 international data analysis competition on BCI-tasks [10] (data set III). The particular competition data was provided by the Dept. of Med. Informatics, Inst. for Biomed. Eng., Univ. of Techn. Graz. The EEG from three channels (C3, Cz, C4) was acquired with bandfilter settings of 0.5 to 30 Hz and sampled at 128 Hz. The data consist of 140 labeled and 140 unlabeled trials of imaginary hand movements, with an equal number of left and right hand trials. Each trial has a duration of 9 s: after a 3 s preparation period a visual cue (arrow) is presented pointing either to the left or the right. This is followed by another 6 s for performing the imagination task (for details see [10]). The specific competition task is to provide an on-line discrimination between left and right movements for each of the 140 unlabeled STs. In particular, at every time instance in the interval from 3 to 9 seconds a decision and its confidence must be supplied. The objective of the competition is to detect the respective motor intention as early and as reliable as possible. Consequently, the competing algorithms have to solve the on-line discrimination task as based on information on the event onset. Thus it is not within the scope of the competition to solve the problem of detecting the event onset itself.

We approach this problem by applying an algorithm focusing on the different modulations of the μ-rhythm. Since we assume that the midline channel Cz does contain little of discriminative information, we excluded it and restricted the analysis to C3 and C4. To extract the rhythmic information we map the EEG to the time-frequency domain by means of Morlet wavelets [11]. For the extracted time courses of the rhythmic activity we estimate two probabilistic models, one for each class, on the labeled training data. The classification of an unlabeled ST is then derived using Bayes theorem.

The paper is organized as follows: section II introduces some of the underlying neurophysiological principles and elaborates on the ERD framework. Section III describes the feature extraction, starting from a short excursion on Morlet wavelets. It also introduces probabilistic models for a single point in time and explains how to combine the information from preceding time points to gather accumulating evidence for the final classification at a certain time point. In section IV the results on the competition data are given and discussed.

II. Neurophysiological background

In standard ERD approaches [12] the EEG signals are first filtered in a narrow frequency band, squared, lowpass filtered and averaged over trials. The ERD is then defined as the attenuation of the band power relative to the power in a preceding baseline interval. Fig. 1 visualizes the μ-rhythm ERDs in the lower frequency band at both hemispheres, induced by imaginary hand movements.

Notably, for ST classifications an ERD evaluation relative to a baseline can lead to misinterpretations. Instead we used the

C. Schäfer is with the IDA group at the Fraunhofer Institute FIRST; S. Lemm and G. Curio are with the Neurophysics Group, Dept. of Neurology, Campus Benjamin Franklin, Charité - University Medicine

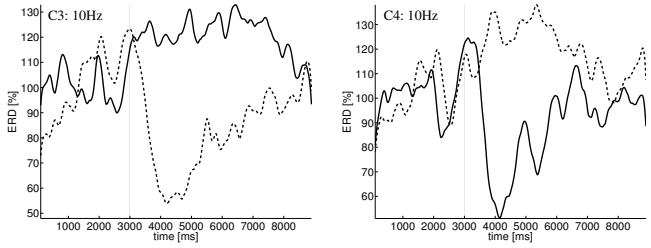

Fig. 1. The panels show the averaged ERD of the μ-rhythm at 10 Hz for the imagination of left (solid line) and right (dashed line) hand movement. The vertical line indicates the begin of the imagination period. The μ-rhythm amplitude is attenuated in relation to the preceding baseline during the motor intention. This attenuation is prominent contralateral to the intended movement, i.e., for right hand movement over the left hemisphere (C3) and over right hemisphere (C4) for the left hand.

time courses of the absolute amplitudes, independent of the level of oscillatory activity in the preceding baseline interval. To illustrate the potentially misleading influence of an ERD baseline, we sort all trials according to their amplitude level in the preparation period, then split the whole set of trials into four non-overlapping quartiles of this distribution and calculate the ERD for each subgroup separately. An ERD, as observed for the average over all trials, is actually found only for the two subgroups with the highest amplitudes during the preparation period. In striking contrast it turns even into an event-related synchronization (ERS) for the subgroup with the lowest preceding amplitude (see left panel of Fig. 2). In clear distinction to the baseline-corrected ERDs, the time courses of the absolute rhythm amplitudes behave similar for the four subgroup averages (see right panel of Fig. 2). Thus, μ-rhythm modulations should be considered independent of the oscillation level in the preceding reference interval. Given this finding we suggest that motor activity not necessarily lead to a desynchronization in general, rather a certain low, yet not the minimal level of oscillatory activity is instantiated in pericentral sensorimotor cortices. Accordingly, for the BCI

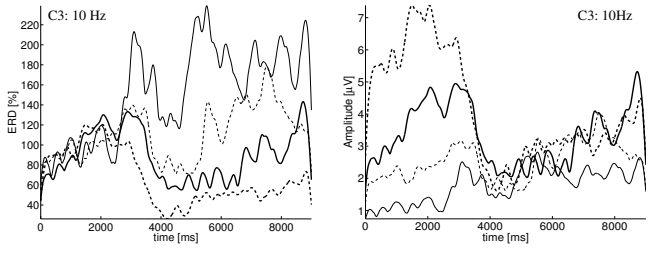

Fig. 2. The averaged time courses of the absolute amplitudes (right panel) vs. an ERD measure relative to a baseline prior to the directional command cue displayed at 3 s (left panel). Data are shown for the 10 Hz band at electrode C3 for imaginary right hand movement. Note that each of the four subgroups of different pre-stimulus activity consists of only 17 STs.

classification task we do not focus on the detection of an ERD relative to a running baseline but rather on modeling the hand-specific time course of absolute μ-rhythm amplitudes over both sensorimotor cortices for each ST.

III. On-line Classification Procedure

A. Preprocessing and feature extraction

Let $C_k^3(t)$ and $C_k^4(t)$ denote the EEG signals recorded from electrodes C3 and C4 in the k-th trial at time t . The corresponding class labels assigned to each trial of the training data are given by $y_k \in \{\mathrm{L}, \mathrm{R}\}$. For our analysis we use a time-frequency representations of the STs by filtering the EEG data with complex Morlet wavelets [11].

Typically wavelets are defined by a mother wavelets. In the case of Morlet wavelets it takes the form of a modulated Gauss-impulse with an characteristic eigenfrequency ω_0. To localize the wavelets in the time and frequency domain, the mother wavelet has to be scaled and temporally shifted, i.e.:

$$\Psi_{\tau,s}(t) \quad = \quad \frac{1}{\sqrt{s}}\pi^{-\frac{1}{4}} e^{\left(i\omega_0 \frac{t-\tau}{s}\right)} e^{-\frac{1}{2}\left(\frac{t-\tau}{s}\right)^2}. \qquad (1)$$

Note that Morlet wavelets are Gaussian filters in the frequency domain. The appropriate scaling factor s depends on the main receptive frequency f of the wavelet and can be calculated with respect to ω_0 as

$$s(f) = \frac{\omega_0 + \sqrt{2+\omega_0^2}}{4\pi f}. \qquad (2)$$

The effective widths of the wavelet in the time and frequency domain are often expressed as the duration of attenuation by a factor e (e-folding) and are denoted by t_{eff} and f_{eff}. With respect to the scaling factor s, these quantities are

$$t_{\mathit{eff}} = \sqrt{2}s, \qquad f_{\mathit{eff}} = \sqrt{2}\,(2\pi s)^{-1}. \qquad (3)$$

It is worth noting that an increase of the eigenfrequency ω_0 of the mother wavelet sharpens the frequency resolution at the expense of a lower temporal resolution.

The wavelet transform of a real signal $C(t)$ at time τ and frequency f is its convolution with the scaled and shifted wavelet. From this complex signal we calculate the instantaneous amplitude as:

$$a(\tau, f) = \frac{1}{\sqrt{s}} \|C(t) * \Psi_{\tau,s(f)}(t)\|. \qquad (4)$$

In order to extract the rhythmic activity in the two frequency bands, we adapt the parameters $(f^\alpha, \omega_0^\alpha)$ and $(f^\beta, \omega_0^\beta)$ of the wavelets to individual EEG spectra. All model parameters – including those used for the wavelets – are estimated using a leave-one-out (LOO) cross-validation scheme on the final model performance [13]. The LOO procedure is applied by leaving out entire STs rather than single time points. The obtained wavelet parameters for the lower and upper μ-band are $(f^\alpha, \omega_0^\alpha) = (10\ \mathrm{Hz}, 10)$ and $(f^\beta, \omega_0^\beta) = (22\ \mathrm{Hz}, 6)$. The corresponding effective frequency ranges are shown in Fig. 3, the e-folding times correspond to $t_{\mathit{eff}}^\alpha = 226\,\mathrm{ms}$ and $t_{\mathit{eff}}^\beta = 62\,\mathrm{ms}$.

It follows from Eq. (1) that Morlet wavelets are unbounded, time-symmetric filters centered on the time-point under study, thus featuring a causal half (looking backward on the time axis) and an acausal half (looking forward). Since real on-line processing permits only causal filtering, we first limit the extension of the wavelets in the time domain, by restricting them to four times the e-folding time. In addition, we shift

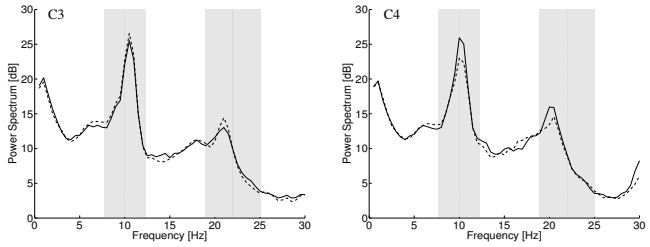

Fig. 3. The averaged frequency spectra for left (dashed) and right (solid) movements. The vertical lines and the shaded regions denote the main receptive frequencies and effective width of the wavelets in the frequency domain.

the wavelets backward on the time axis by the maximum of the remaining acausalities, i.e., by $452\,\mathrm{ms}$. The time courses of the rhythmic activity at the two electrodes over the hand motor cortices are then calculated by the convolution of each trial with these causal wavelets. The resulting amplitudes for the two frequency bands at the two electrodes are then stacked together into the feature vector

$$\mathbf{a}_k(t) = \left(a_k^3\left(t, f_\alpha\right), a_k^4\left(t, f_\alpha\right), a_k^3\left(t, f_\beta\right), a_k^4\left(t, f_\beta\right)\right)^T. \quad (5)$$

The obtained feature vectors of the labeled trials are further used for the estimation of the probabilistic models.

B. Probabilistic models for a single point in time

Imaginary movements modulate the μ-rhythm on the hemisphere contralateral to the respective event more than ipsilateral (see Fig. 1, [3]–[6]). Combining this with the framework of absolute ERDs, we assume the existence of two distinguishable prototypical behaviors of modulation for the *absolute* amplitude of the μ-rhythm caused by either imaginary left or, respectively, right hand movements. Based on these physiological concepts we estimate two probabilistic models, one for each class of imaginary movement. At any time instance $t \in [0-9]\,\mathrm{s}$ we assume a 4-dimensional Gaussian distribution of the feature vectors $\mathbf{a}(t)$ for each model, i.e.,

$$p(\mathbf{a}(t)|y) = \frac{|\Sigma_t^y|^{-\frac{1}{2}}}{(2\pi)^2} \cdot e^{\left(-\frac{1}{2}(\mathbf{a}(t)-\mu_t^y)^T(\Sigma_t^y)^{-1}(\mathbf{a}(t)-\mu_t^y)\right)}, \quad (6)$$

where μ_t^y and Σ_t^y are the individual means and the covariance matrices of the two classes $y \in \{\mathrm{L},\mathrm{R}\}$. The Gaussian assumption is made for computational convenience, accepting its systematic errors arising from the facts that EEG amplitudes are bounded and the empirical marginal distributions are found asymmetric. This observed skewness could be due to artifacts or, more interestingly, might be caused by STs in which the subject was not able to adequately manage the imagination task, possibly because of a temporarily declining concentration so that the ongoing oscillation stayed unperturbed at a higher average amplitude causing an overestimation of the mean. To compensate for this we estimate α-trimmed class means $\hat{\mu}_t^{\mathrm{L}}$ and $\hat{\mu}_t^{\mathrm{R}}$. Specifically, we exclude 10% of the largest observations of any class and for any dimension in order to adjust the means toward zero, whereas the size of the α-quantile to be removed is estimated by the LOO procedure as well. The covariance matrices $\hat{\Sigma}_t^{\mathrm{L}}$ and $\hat{\Sigma}_t^{\mathrm{R}}$ are estimated using the empirical means $\bar{\mathbf{a}}_t^y$

$$\hat{\Sigma}_t^y = E_{\{y=y_k\}}\left[(\mathbf{a}_k(t) - \bar{\mathbf{a}}_t^y)^T(\mathbf{a}_k(t) - \bar{\mathbf{a}}_t^y)\right]. \quad (7)$$

For $y \in \{\mathrm{L},\mathrm{R}\}$ the classification for a single point in time t is derived, using Bayes theorem, as

$$p\left(y|\mathbf{a}\left(t\right)\right) = \frac{p\left(\mathbf{a}\left(t\right)|y\right)}{p\left(\mathbf{a}\left(t\right)|\mathrm{L}\right) + p\left(\mathbf{a}\left(t\right)|\mathrm{R}\right)}. \quad (8)$$

C. Combining information across time

In order to finally derive the on-line classification at a certain time t_0, we incorporate knowledge from all preceding timepoints $t \leq t_0$, leading to an evidence accumulation over time about the binary decision process. The temporal combination is realized by taking the expectation of the class probabilities from Eq. (8) with respect to the discriminative power at each point in time, denoted by w_t that will be derived later on,

$$p\left(y|\mathbf{a}\left(0\right), \ldots, \mathbf{a}\left(t_0\right)\right) = \frac{\sum_{t\leq t_0} w_t p(y|\mathbf{a}(t))}{\sum_{t\leq t_0} w_t}. \quad (9)$$

To derive the discriminative power we use the Bayes error of misclassification [13] of the two estimated class distributions at time t (cf. Eq. (6)). As the Bayes error cannot be calculated directly because of the distinct covariance matrices, we approximate it from above by the Chernoff bound [13] and finally define w_t by

$$2w_t := 1 - \min_{0\leq\gamma_t\leq 1} \int p(\mathbf{a}(t)|\mathrm{L})^{\gamma_t} p(\mathbf{a}(t)|\mathrm{R})^{1-\gamma_t} d\mathbf{a} \quad (10)$$

In the case of Gaussian distributions the integral can be expressed in a closed form [13], such that the minimum solution can be easily obtained. Fig. 4 shows the estimated Chernoff bound, given the training data. Note that the most discriminative information occurs around $4.5\,\mathrm{s}$, as indicated by the minimum of the error bound that corresponds to the maximum weight in the integration process. In practice we start the integration at $3.5\mathrm{s}$, this taking into account the time required for the cognitive cue processing as well as the temporal extent of the wavelet filters.

Strictly speaking Eq. (9) gives the probabilities that the observed ST are generated by either one of the models, until time t_0. Due to the submission requirements of the competition the final decision at this point in time is

$$y_{t_0} = 1 - 2 \cdot p(L|\mathbf{a}(1), \ldots, \mathbf{a}(t_0)), \quad (11)$$

where a positive or negative sign refers to right or left movements, while the magnitude indicates the confidence in the decision on a scale between 0 and 1.

IV. Results and Discussion

As mentioned before we estimate all model hyperparameters, in particular for the wavelets and the alpha quantile by means of LOO cross-validation optimization. The selected parameters achieve a minimum binary LOO decision error of 7.9% on the training data. Using these parameters we apply the model estimated from the labeled training

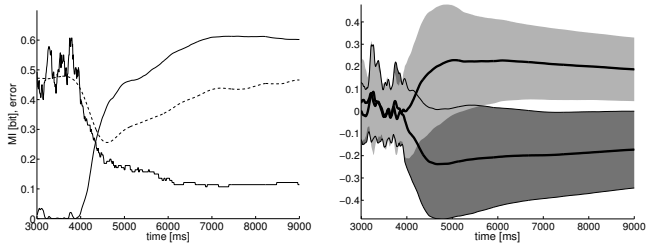

Fig. 4. Left panel shows the time course of the classification error (thin solid), the Chernoff bound on the Bayes error (dashed) and the mutual information (thick solid). Right panel displays time course of the mean and standard deviations of the decision according Eq. (11) for right (light grey) and left (dark grey) imaginary movements on the test data.

data set to the feature vectors of the unlabeled STs of the test data. To permit an objective evaluation of the model performance of each competitor, the labels for the test data were published after competition closure. The resulting time courses for both the error of the binary classification and the mutual information (MI) [14] on the previously unlabeled data are presented in Fig. 4. During the first 4 seconds the classification is rather by chance, after 4 seconds a steep ascent in the classification accuracy can be observed in both the raising MI and the decreasing classification error. Although the Bayes error bound starts to gradually increase again after 4.5 s, indicating fading separability, the full model still gains information due to the integration process so that at 6.8 s an overall minimum error of 10.7% is achieved. The MI maximum of 0.61 Bit occurs at 7.6 s indicating a peak decision confidence at this time. Demonstrating the time courses of the class means and standard deviations of the decision the right panel of Fig. 4 emphasizes the high discriminative ability of the proposed procedure: around 6 s there is no overlap between the class standard deviation tubes, reflecting the high confidence of the decisions. A comprehensive comparison of all submitted techniques to solve the specific task for data set III of the BCI-competition is provided in [10]. Basically this evaluation reveals that the proposed algorithm outperforms all competing approaches, including traditional AAR-parameter based methods.

The processing time for all 140 unlabeled STs using non-optimized MATLAB code adds up to approximately 110 s. Since each trial lasts for 9 s and is recorded at 128 Hz, this strongly emphasizes the on-line applicability of the proposed algorithm, i.e. a stream of incoming EEG data can be processed at a rate as high as about 1 kHz.

Notably, the general Bayesian framework of the present approach is applicable to other kinds of sequential two-class on-line discrimination tasks, just by adapting the preprocessing and/or feature extraction. Moreover, different task objectives can be chosen, such as classification speed, accuracy or confidence, and, correspondigly, different modifications of the algorithm are possible. However, the optimization of one particular objective will come at the expense of the others. For speeding up the system working on the present BCI task, one can choose lower eigenfrequencies ω_0 for the mother wavelet (cf. Eq. (1)). This leads to a shorter temporal duration of the wavelets and thereby to an acceleration of the decision process. On the other hand abandoning the Gaussian assumption and fitting other distributions could increase the classification accuracy, yet determining the Bayes error of misclassification for these models will require a further appropriate elaboration.

Future work will be dedicated to the analysis of asynchronous imaginary movements, i.e., without information about onset of movement being provided. For such on-the-fly on-line classifications there are several ways how to apply the proposed algorithm. The most obvious one consists of a two-stage procedure: first a movement detector followed by the classifier proposed here, aligned to the detected onset of movement. Another possibility is to apply a sequence of any favored classifier and use the Bayesian temporal combination framework presented here to integrate its evidence over time.

Acknowledgement: The studies were supported by BMBF Grants FKZ 01IBB02A,B and DFG SFB 618-B4.

REFERENCES

[1] R. Hari and R. Salmelin, "Human cortical oscillations: a neuromagnetic view through the skull," *Trends in Neuroscience*, vol. 20, pp. 44–9, 1997.

[2] H. Jasper and W. Penfield, "Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus," *Arch. Psychiatrie Zeitschrift Neurol.*, vol. 183, pp. 163–74, 1949.

[3] G. Pfurtscheller and A. Arabibar, "Evaluation of event-related desynchronization preceding and following voluntary self-paced movement," *Electroenceph. clin. Neurophysiol.*, vol. 46, pp. 138–46, 1979.

[4] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmäki, and R. Hari, "Involvement of primary motor cortex in motor imagery: a neuromagnetic study," *Neuroimage*, vol. 6, pp. 201–8, 1997.

[5] V. Nikouline, K. Linkenkaer-Hansen, H. Wikström, M. Kesäniemi, E. Antonova, R. Ilmoniemi, and J. Huttunen, "Dynamics of mu-rhythm suppression caused by median nerve stimulation: a magnetoencephalographic study in human subjects," *Neuroscience Letters*, vol. 294, 2000.

[6] S. Salenius, A. Schnitzler, R. Salmelin, V. Jousmäki, and R. Hari, "Modulation of human cortical rolandic rhythms during natural sensorimotor tasks," *NeuroImage*, vol. 5, pp. 221–8, 1997.

[7] J. Wolpaw and D. McFarland, "Multichannel EEG-based brain-computer communication," *Electroenceph. clin. Neurophysiol.*, vol. 90, pp. 444–9, 1994.

[8] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroenceph. clin. Neurophysiol.*, vol. 103, pp. 642–51, 1997.

[9] C. Neuper, A. Schlögl, and G. Pfurtscheller, "Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery," *Journal Clin. Neurophysiol.*, vol. 16, pp. 373–82, 1999.

[10] B. Blankertz, K.-R. Müller, T. V. G. Curio, G. Schalk, J. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, and M. S. N. Birbaumer, "The BCI competition 2003," *IEEE Trans. Biomed. Eng.*, 2003.

[11] C. Torrence and G. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meterol.*, vol. 79, pp. 61–78, 1998.

[12] J. Kalcher and G. Pfurtscheller, "Discrimination between phase-locked and non-phase-locked event-related EEG activity," *Electroenceph. clin. Neurophysiol.*, vol. 94, pp. 381–4, 1995.

[13] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, New York, 2001.

[14] A. Schlögl, R. S. C. Keinrath, and G. Pfurtscheller, "Information transfer of an EEG-based brain-computer interface," in *Proc. First Int. IEEE EMBS Conference on Neural Engineering*, 2003, pp. 641–644.