Neural Networks 22 (2009) 1313–1319

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

Time Domain Parameters as a feature for EEG-based Brain–Computer Interfaces[☆]

Carmen Vidaurre [c,a,*], Nicole Krämer [c], Benjamin Blankertz [c,a], Alois Schlögl [b]

[a] IDA at Fraunhofer FIRST, Kekulestr. 7, 12489 Berlin, Germany
[b] Institute for Human–Computer Interfaces, University of Technology Graz, Krengasse 37/I, 8010 Graz, Austria
[c] Machine Learning Group, Berlin Institute of Technology, Franklinstr. 28/29, 10587 Berlin, Germany

ARTICLE INFO

Article history:
Received 14 October 2008
Received in revised form 3 July 2009
Accepted 16 July 2009

Keywords:
Brain–Computer Interface
Hjorth parameters
Band power estimates
Time Domain Parameters
Derivative operator

ABSTRACT

Several feature types have been used with EEG-based Brain–Computer Interfaces. Among the most popular are logarithmic band power estimates with more or less subject-specific optimization of the frequency bands. In this paper we introduce a feature called Time Domain Parameter that is defined by the generalization of the Hjorth parameters. Time Domain Parameters are studied under two different conditions. The first setting is defined when no data from a subject is available. In this condition our results show that Time Domain Parameters outperform all band power features tested with all spatial filters applied. The second setting is the transition from calibration (no feedback) to feedback, in which the frequency content of the signals can change for some subjects. We compare Time Domain Parameters with logarithmic band power in subject-specific bands and show that these features are advantageous in this situation as well.

1. Introduction

Amplitude modulations of sensorimotor rhythms (SMRs) can be voluntarily induced by most people, e.g. by kinesthetically imagining limb movements. This ability can be taken as a basis for Brain–Computer Interfaces (BCIs) which are devices that translate the intent of a subject measured from brain signals directly into control commands, e.g. for a computer application or a neuroprosthesis. BCI systems can also be based on other components of brain signals, see Dornhege, del R. Millán, Hinterberger, McFarland, and Müller (2007), Allison, Wolpaw, and Wolpaw (2007), Birbaumer et al. (2006), Pfurtscheller, Neuper, and Birbaumer (2005), Wolpaw, Birbaumer, McFarland, Pfurtscheller, and Vaughan (2002) and Kübler, Kotchoubey, Kaiser, Wolpaw, and Birbaumer (2001) for an overview, but only SMR-based BCIs are considered in this paper.

One of the challenges in the development of BCI systems is to minimize the time required before accurate BCI performance is possible (subject training Elbert, Rockstroh, Lutzenberger, and Birbaumer (1980), Rockstroh, Birbaumer, Elbert, and Lutzenberger (1984), Birbaumer et al. (2000) or gathering calibration data in off-line runs to feed machine learning (ML) algorithms (Blankertz, Dornhege, Krauledat, Müller, & Curio, 2007; Blankertz & Losch et al., 2008)). In this paper we present an exploration of the usefulness of several features for the use in ML-based BCIs, that aim at providing accurate feedback as early as possible.

The features we propose in this manuscript are inspired by the Hjorth parameters that have been previously used in BCI experiments; for an overview please refer to Boostani and Moradi (2004); Coyle, McGinnity, and Prasad (2006); Obermaier, Guger, Neuper, and Pfurtscheller (2001) and Vourkas, Micheloyannis, and Papadourakis (2000). We have generalized the Hjorth parameters and obtained features that can be rapidly and easily computed. They rely on the estimation of band power in wide bands and use the frequency content of the signal itself, which makes them more robust against over-fitting or non-stationarities. First, we explore the situation in which immediate feedback is provided to the subject without prior calibration measurement. This approach is not new in the BCI literature and has already been proven as useful for user training (see Blankertz & Vidaurre, 2009; Vidaurre, Schlögl, Cabeza, Scherer, & Pfurtscheller, 2007). We compare Time Domain Parameters (TDP) with the traditionally used band power estimates (BPE). Second, we explore the transition between calibration and feedback, in which a change in the frequency content of the signal can be expected for some subjects. In this setting, TDP and BPE are extracted from user-specific wide bands and combined with different spatial filters. A regularized linear discriminant classifier (RLDC) is used to estimate the separability of the features.

2. Time Domain Parameters (TDP)

As TDP are inspired by the Hjorth parameters, we will present the latter. The parameters introduced by Hjorth (1970) are three

☆ This work was supported by the grants: EU project 040666 Multiadaptive BCI, DFG (MU 987/3-1) VitalBCI Project, and EU TOBI project FP7-224631.

* Corresponding author at: Machine Learning Group, Berlin Institute of Technology, Franklinstr. 28/29, 10587 Berlin, Germany.
E-mail address: vidcar@cs.tu-berlin.de (C. Vidaurre).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.07.020

features defined as follows:

$$Activity = var(x(t))$$

$$Mobility = \sqrt{\frac{Activity\left(\frac{dx(t)}{dt}\right)}{Activity\,(x(t))}}$$

$$Complexity = \frac{Mobility\left(\frac{dx(t)}{dt}\right)}{Mobility\,(x(t))}.$$

The first parameter, Activity, is the signal power (which is wide band filtered), Mobility is the mean frequency and Complexity the change in frequency. Note that the BPE is the same as the Activity, only at some specific frequency bands.

In this paper, we propose TDP to test whether different number of derivatives of the signal could improve the classification performance. Therefore, the number of derivatives calculated p is needed as parameter:

$$p_i(t) = var\left(\frac{d^i x(t)}{dt^i}\right), \quad i = 0, \ldots, p.$$

Note that TDP of order 0 is BPE of the, usually, band-pass filtered signal.

Although TDP features are defined in the time domain they can be as well interpreted as frequency domain filters. Of course, there are spectral methods like Fourier transform, wavelet analysis and autoregressive spectrum that are able to describe the whole spectral density function. These have the disadvantage that a rather large number of parameters are obtained causing difficulties in the classification step: more features require more training data and increase the danger of over-fitting.

In order to avoid the dimensionality problem, band power estimates of one or two frequency bands are used in BCI research. They often require the selection of specific frequency bands, with the risk that changes outside of these are ignored. Another option that has also been investigated in the BCI field is the use of the (adaptive) autoregressive (AAR) parameters (instead of AR spectrum). They reduce the number of features per channel to the order of the AR model. In the past, model orders in the range of 3 to 9 (see Schlögl, Flotzinger, & Pfurtscheller, 1997; Schlögl, 2000; Schlögl, Lee, Bischof, & Pfurtscheller, 2005) have been used. TDP features seem to suggest a similar small number of parameters. TDP have also another similarity with AR parameters, because the AR model is originally defined in the time domain. There is no need to know the spectral representation of the data, but the representation in the time domain is sufficient to grasp the idea of both, TDP and AR methods. A comparison between AAR and BPE parameters can be found in Vidaurre, Schlögl, Cabeza, Scherer, and Pfurtscheller (2005), where no significant differences between both methods were found.

Since BPE or TDP features are not normally distributed, we apply the logarithm to $p_p(t)$ in order to obtain features whose distribution is approximately Gaussian, since this makes linear classification more successful. The code for computing TDP as well as other methods is publicly available in the BioSig-toolbox at http://biosig.sf.net, (Schlögl & Brunner, 2008).

Finally, as subsequent derivatives of the signals are correlated, in this manuscript we used different subsets of derivative orders to calculate the final features. Studied subsets always contained order 0 and the rest of derivative orders were combined. The proper subset of derivative orders was found using cross-validation or principal component analysis (PCA). The choice between one and the other method, depended on the experimental setting. For more information see Section 4.5.

3. Data

Data were recorded in a one-day session from 80 healthy BCI-novices (39m, 41f; age 29.9±11.5 years; 4 left handed). Data of 5 subjects was disregarded due to problems during the experimental session that affected the labeling of the classes. The subjects were sitting in a comfortable chair with arms lying relaxed on armrests. Brain activity was recorded from the scalp with multi-channel EEG amplifiers using 119 Ag/AgCl electrodes in an extended 10–20 system sampled at 1000 Hz with a band-pass from 0.05 to 200 Hz. The data was filtered using a low-pass filter with a cutoff frequency of 40 Hz and down-sampled to 100 Hz.

First, the subjects performed a calibration measurement in which every 8 s one of three different visual cues (arrows pointing left, right, down) indicated to the subject which type of motor imagery to perform: left/right hand or foot. Three runs with 25 trials of each motor condition were recorded. Automatic variance based artifact rejection was made to discard noisy channels. Then, the most discriminative pair of classes were selected based in log-BPE features and the subjects performed a feedback measurement with three runs of 100 trials each, although for some subjects only two runs were recorded.

4. Spatial filtering

Different types of spatial filtering were investigated to preprocess the data before extracting the features. All of them are commonly used in EEG-BCI systems.

4.1. Bipolar montage

Bipolar channels are widely used by several BCI-groups, specially in a set of reduced channels (Krausz, Scherer, Korisek, & Pfurtscheller, 2003; Scherer, Lee, Schlögl, Bischof, & Pfurtscheller, 2008; Vidaurre, Schlögl, Cabeza, Scherer, & Pfurtscheller, 2006). Bipolar channels are computed subtracting the signals from two neighboring electrodes. We extracted 3 bipolar channels over C3, Cz and C4, using FC3-CP3, FCz-CPz and FC4-CP4, which is the most typical arrangement.

4.2. Laplacian montage

Small laplacian derivations (McFarland, McCane, David, & Wolpaw, 1997) are easy to calculate and extensively used in EEG recordings. In this study each laplacian derivation was calculated as follows: 4 surrounding channels, equally weighted, were subtracted to the central one. Using laplacian channels we computed 3 different spatial filters.

4.2.1. 3 laplacian channels

This spatial filter was computed over C3, Cz and C4, for direct comparison with the bipolar channels.

4.2.2. 11 laplacian channels

Also over the motor area, the following channels were selected: C1-2-3-4, Cz, FC3-4, FCz, CP3-4, CPz.

4.3. Common Spatial Patterns, CSP

CSP is a technique to analyze multi-channel data based on recordings from two classes (tasks). It yields a data-driven supervised decomposition of the signal $\boldsymbol{x}(t)$ parametrized by a matrix $\boldsymbol{W}$ that projects the signal in the original sensor space to a surrogate sensor space $\boldsymbol{x}_{CSP}(t)$, (Blankertz, Tomioka, Lemm, Kawanabe, & Müller, 2008): $\boldsymbol{x}_{CSP}(t) = \boldsymbol{x}(t) \cdot \boldsymbol{W}$. Each column vector of a $\boldsymbol{W}$ is a spatial filter. CSP filters maximize the variance of the spatially filtered signal under one task while minimizing it for the other task. Since the variance of a band-pass filtered signal is equal to band power, CSP analysis is applied to band-pass filtered signals to obtain an effective discrimination of mental states that are characterized by ERD/ERS (even related desynchronization/synchronization) effects. Detailed information about this technique can be found in (Blankertz & Tomioka et al., 2008). For our study CSP filters were individually selected for each subject using the band-pass filtered signal. The number of filters used was automatically selected and ranged between 2 and 6 filters.

4.4. Classification

A regularized linear discriminant classifier was used because of the high-dimensionality of the features (in some of the settings) compared to the number of trials available. Recall that the Linear Discriminant Analysis (LDA) finds a one-dimensional subspace in which the classes are well separated. This is formalized by requiring that after the projection onto the subspace, the ratio of the between-class variance to the within-class variance is maximal. For the case of two classes, which we consider here, the optimal subspace is defined by

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \tag{1}$$

where $\boldsymbol{\Sigma}$ is the sample-covariance matrix, and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are the sample class means. As the covariance matrix is often typically poorly conditioned, we follow the approach by Ledoit and Wolf (2004a, 2004b) and replace $\boldsymbol{\Sigma}$ in Eq. (1) by a shrinkage estimate of the form

$$\boldsymbol{\Sigma}_\lambda = (1-\lambda)\,\boldsymbol{\Sigma} + \lambda\widetilde{\boldsymbol{\Sigma}}, \quad \lambda \in [0, 1].$$

The matrix $\widetilde{\boldsymbol{\Sigma}}$ is the sample-covariance matrix of a restricted sub-model, and the optimal shrinkage intensity λ can be estimated from the data. We use the following sub-model: all variances (i.e. all diagonal elements) are equal, and all covariances (i.e. all off-diagonal elements) are zero. (See Schäfer and Strimmer (2005) for other alternatives, and their corresponding optimal λ).

4.5. Parameter selection

The CSP filters, the order and mixture of TDP components, the frequency bands and time interval for applying the classifier when necessary, had to be estimated. In this manuscript we study two different settings and in each of them the parameters were selected in a different manner.

4.5.1. First setting, no previous subject data available

In this case only the feedback measurement was used to find parameters. As no previous data was available for a particular subject, the data of the other subjects was used to find parameters. Here, the frequency bands and time interval to calculate the features were fixed beforehand (see values in Section 5). However, a proper subset of derivative orders (see Section 2) to calculate TDP had to be fixed. This subset was selected by cross-validation in a leave one “subject” out fashion: out of 75 subjects, 74 were used to select the proper subset of derivative orders and one subject was used to obtain a test error. The errors of each subset were computed with cross-validation in the feedback measurement of each subject. This procedure was repeated 75 times until completing the full pool of subjects and the test error averaged. The stability of the selected parameter is studied in Section 5.

Three spatial filters that can be calculated in an unsupervised fashion were studied in this setting: 3 and 11 laplacian channels and 3 bipolar channels.

4.5.2. Second setting, calibration to feedback transition

For this setting the parameters were found subject-specifically using the calibration session. With log-BPE features, one or two discriminative frequency bands and optimum trial time interval were selected following the state of the art method described in Blankertz and Tomioka et al. (2008). With log-TDP features, wider frequency bands were systematically chosen. For calculating log-TDP, up to order 6 was used. The dimensionality of the features was reduced using principal component analysis (PCA) that retained a very large percent of the variance (99.5%). With the selected features, the classifier was trained in the calibration measurement and applied to the feedback measurement to obtain a test error for each subject.

The spatial filters used were 3 and 11 laplacian channels (unsupervised) and CSP (supervised).

Table 1
Subsets of derivative orders found for TDP with each spatial filter (order 0 means log-BPE) and number of iterations in which a specific subset was chosen.

TDP subsets of derivative orders selected during leave one subject out		
	Number of iterations	Subset orders
	58	[0 1 2 3 4 5 6]
3 bip	16	[0 2 3 4 6]
	1	[0 1 2 4 5 6]
3 lap	74	[0 1 2 3 4 6]
	1	[0 1 2 3 6]
11 lap	75	[0 1 2 3 6]

5. Results

Before presenting our results, we want to remark that the selection of tasks for the feedback measurement and the feedback measurement itself was done based uniquely in subject-specific log-BPE. Therefore the results are harder to interpret as a bias toward BPE can be expected. The bias might be specially strong in the second setting, the transition from calibration to feedback, and with CSP as spatial filter.

5.1. First setting, no previous subject data available

In our experiments we computed the features filtering the data different frequency bands and then estimated TDP from order 0 to 6. Different derivative order subsets were used, in which specific derivative orders were concatenated to form a feature vector. The proper order subset was chosen through leave one “subject” out. The subset that performed best in 74 subjects was tested in the one left. The procedure was repeated 75 times until completing all subjects. log-TDP was extracted in the fixed wide band of 8–35 Hz and the interval from 500 ms to 3500 ms after the cue. Log-TDP was compared to log-BPE of one μ band in 8–15 Hz, log-BPE of two bands from 8–15 and 16–28 Hz, in the μ and β bands respectively, and to log-BPE between 8–35 Hz. Three different spatial filters were applied to the features. None of them needs previous subject data to be calculated.

The first row of Fig. 1 depicts the scatter plot of error rates obtained with log-TDP features vs the log-BPE error rates obtained with one μ band, two bands (μ and β), and a wide band. 3 laplacian channels over C3, Cz and C4 were used to calculate the features of this Figure. Log-TDP outperforms the other features when the values are located below the diagonal. The second row of the same figure shows the scatter plots of log-TDP using different types of spatial filters.

In all three plots of the top row in Fig. 1 we can observe important improvement in performance for some subjects, specially in comparison to one narrow and one wide band (left and right plots).

Table 1 summarizes the results of subsets of derivative orders selected with each spatial filter using “leave one subject out” cross-validation. In each step, a subset of derivative orders was selected using 74 subjects and tested in the remaining data-set. The results of subset selected for 3 and 11 laplacian channels are extremely stable, as the same subset was selected in all iterations (except in one case with 3 laplacian channels). In the case of 3 bipolar channels, still 58 out of 75 iterations returned the same subset of derivative orders.

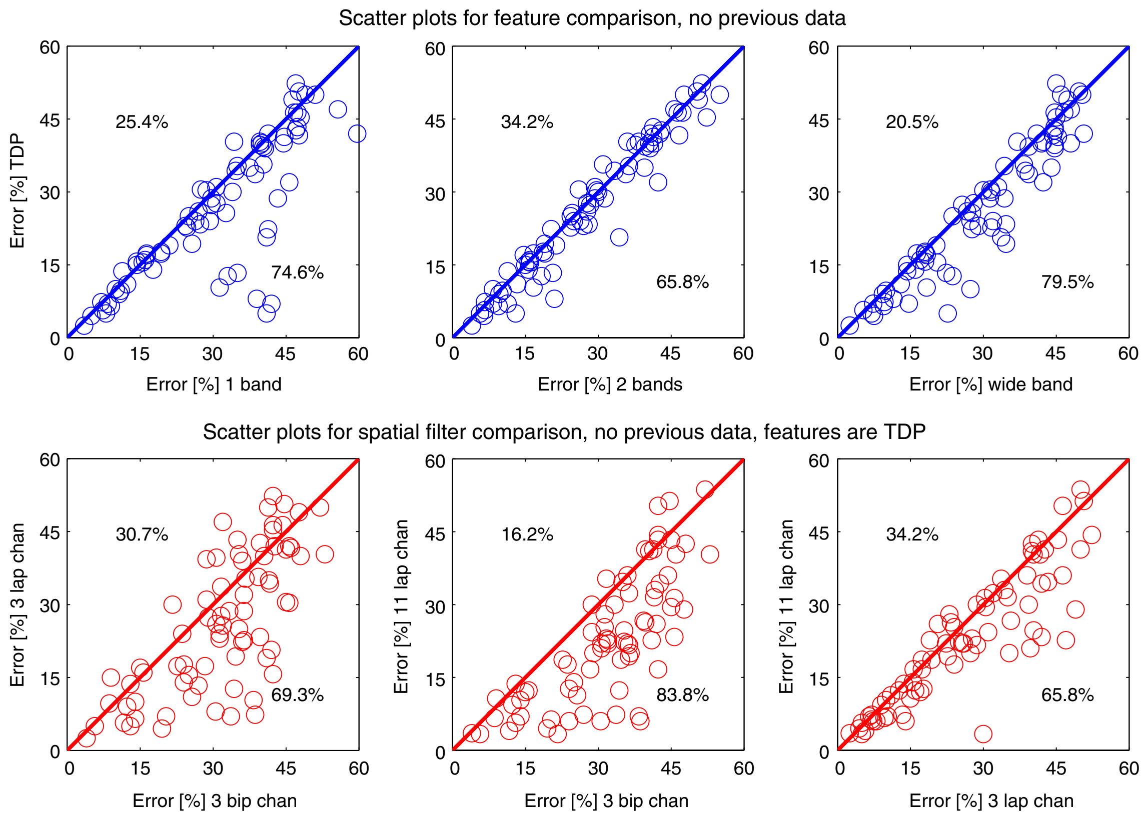

Fig. 1. Analysis of feature performance when no data from the subject is available. The first row shows scatter plots of error rates obtained with log-TDP (y-axis) vs error rates (x-axis) obtained with one μ band (left), two bands (center) and a wide band (right). The spatial filter used was 3 laplacian channels. The second row shows scatter plots of error rates obtained with log-TDP comparing different spatial filtering methods. For the values below the diagonal, the method in y-axis outperforms the method in x-axis.

Table 2
Mean error rates (%) and standard errors for log-BPE in 1 band, 2 bands, 1 wide band and log-TDP using different spatial filters. In this setting no previous data from the subject was available.

	Mean error rates (%) and standard errors of the mean			
	1 band log-BPE	2 bands log-BPE	wide band log-BPE	log-TDP
3 bip	36.21 ± 1.45	33.30 ± 1.48	35.22 ± 1.40	**31.97** ± **1.35**
3 lap	30.82 ± 1.64	27.7 ± 1.59	29.54 ± 1.56	**26.37** ± **1.64**
11 lap	27.44 ± 1.61	24.24 ± 1.57	25.65 ± 1.58	**22.81** ± **1.54**

A summary of results is presented in Table 2, with mean error rates and standard errors of the mean for each feature and spatial filter. The study of significance can be found in Section 6. In this table we see that TDP outperforms, at least in average, the rest of the presented features.

5.2. Second setting, calibration to feedback transition

This setting was selected because the spectrum of the signal can be expected to change for some of the subjects in the transition from calibration to feedback conditions. This can be due to several reasons, as e.g. differences in visual input (the information presented in the screen is different in both settings) or differences in mood or mental state due to the presence of feedback or strategy change in response to the feedback, etc. As example of this change, we chose one subject whose performance improved with TDP features vs BPE. The top row of Fig. 2 reflects the frequency spectrum in one channel during calibration and feedback conditions, both calculated in the same time interval of the trial. The bottom row shows the scalp plots with the spatial distribution of signal power between the two selected tasks. The signal greatly changed between the two conditions.

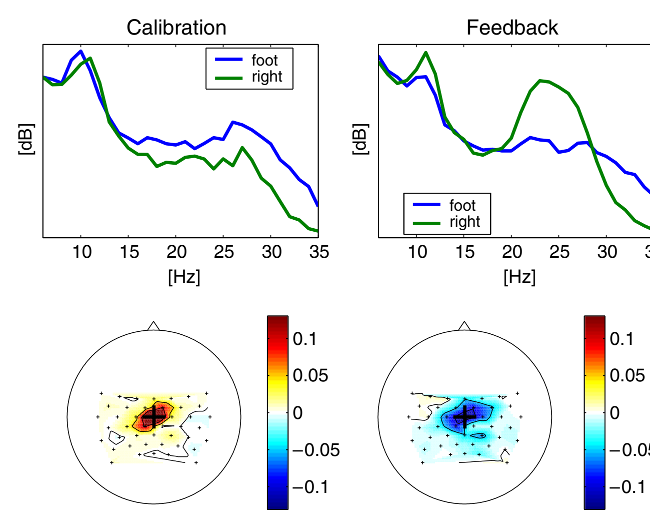

Fig. 2. The top row shows frequency spectra of one channel in calibration and in feedback conditions of a subject who benefited from the use of TDP features. The bottom row shows the scalp plots with the spatial distribution of signal power between the two selected tasks.

Given this setting, we tried log-TDP features to analyze whether they could be beneficial for this transition and compared them to the state of the art methods, that is, subject-specific log-BPE in one or two discriminative bands. The top row of Fig. 3 shows scatter plots comparing the features (log-TDP and log-BPE) with each of the spatial filters. The bottom row compares different spatial filters

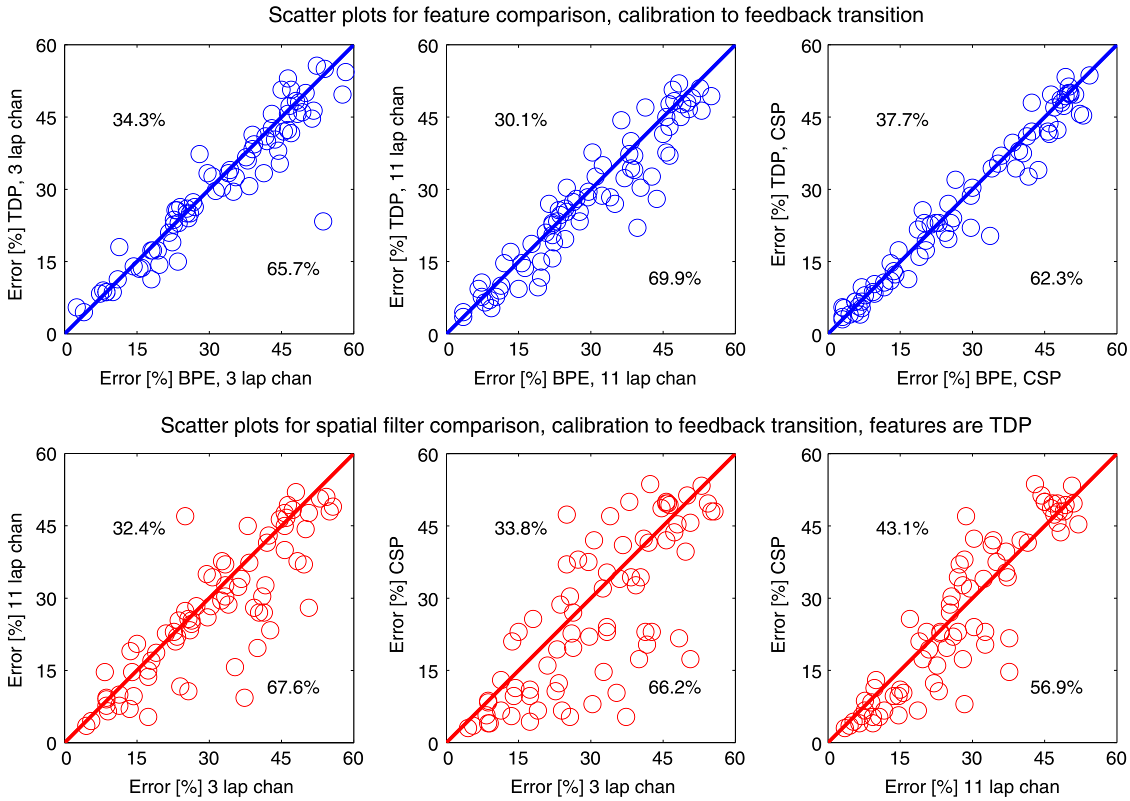

Fig. 3. Analysis of feature performance in the transition from calibration to feedback. The first row shows scatter plots of error rates obtained with log-TDP (y-axis) vs error rates (x-axis) obtained with log-BPE with 3 laplacian channels (left), 11 laplacian channels (center) and CSP (right). The second row shows scatter plots of error rates obtained with log-TDP comparing different spatial filtering methods. For the values below the diagonal, the method in y-axis outperforms the method in x-axis.

Table 3
Mean error rates (%) and standard errors for log-BPE and log-TDP using different spatial filters in the calibration to feedback transition.

	Mean error rates (%) and standard errors of the mean	
	Log-BPE	log-TDP
3 lap	32.44 ± 1.69	31.05 ± 1.62
11 lap	29.62 ± 1.65	27.81 ± 1.61
CSP	27.75 ± 1.94	26.76 ± 1.89

between each other. For the values below the diagonal, the method in y-axis outperforms the method in x-axis. The top row of Fig. 3 reveals that TDP can importantly improve the performance of BPE in some subjects, specially for 3 and 11 laplacian channels.

Table 3 summarizes the mean error rates and standard error of means of log-TDP and log-BPE and the different spatial filters applied. The study of significance can be found in Section 6.

6. Discussion

6.1. First setting, no previous subject data

Here the significance analysis of the results presented in the previous section is performed. We carried out one-sided signed-rank tests for paired samples to find out whether log-TDP performed better than log-BPE and which spatial filter was better. As three comparisons were done in both cases, the p-value was corrected using the following formula: $1-(1-\alpha)^{(1/nc)}$, where α is the original p-value, set at 5%, and nc is the number of comparisons, in this case 3. Accordingly, the new p-value = 1.70%.

All p-values found are summarized in Table 4, where we see that log-TDP significantly outperformed all BPE features. Log-TDP produces better results than log-BPE in one band (narrow or wide). The improvement in comparison to 2 bands is more modest, around 1.3%. However, as these features are fast and easy to compute, and the improvement is sustained across-subjects, we think that they are a competitive choice to the log-BPE.

With respect to the spatial filters, 3 bipolar channels is clearly not the choice. If the available number of channels allows it, mounting 3 laplacian channels is fast and the improvement in performance is big. Further improvement can be expected with 11 channels. Mounting those will depend of the time and hardware available.

We also want to note here that the error rates were computed by cross-validation in the feedback data-set of each subject. In practice, when no data of the particular the subject is available, one can construct a subject-independent classifier, using a pool of subjects and then apply some adaptation, just as in Vidaurre et al. (2007). In this way, reliable feedback can be achieved without the need of a calibration measurement.

6.2. Second setting, calibration to feedback transition

Again in this setting the number of comparisons performed was 3, and therefore the corrected p-value stayed at 1.70%. The statistical tests applied are the same as in Section 6.1.

All p-values found are summarized in Table 5. Log-TDP significantly outperformed all BPE features in all spatial filters. The improvement is modest but consistent, being greater for 11 laplacian channels and smaller for CSP. In all cases, but specially for laplacian channels, we can see that some subjects greatly benefit from the use of log-TDP, specially those for whose spectrum changes in the transition from calibration to feedback (as in Fig. 2). Reasons for TDP performing worse in combination with CSP could be that the mixture of channels was optimized for the filtered data, without taking into account signal derivatives or that the feedback

Table 4
p-values obtained with one-sided signed-rank test for paired samples. For the feature comparison: H0: log-TDP does not perform better than log-BPE. H1: log-TDP performs better than log-BPE. For the comparison of spatial filters, we tested whether 3 and 11 laplacian channels outperformed 3 bipolar channels and whether 11 laplacian channels outperformed 3 laplacian channels.

Feature comparison			
	p-values in (%), limit of significance 1.70%		
	1band log-BPE vs log-TDP	2bands log-BPE vs log-TDP	wband log-BPE vs log-TDP
3 bip	1.07e−06	4.80e−01	5.31e−07
3 lap	2.92e−05	1.60e−01	1.98e−06
11 lap	7.14e−04	1.40e−01	3.48e−05
Comparison of spatial filters			
	3 bip vs 3 lap	3 bip vs 11 lap	3 lap vs 11 lap
	9.02e−04	1.11e−09	7.57e−04

Table 5
p-values obtained with one-sided signed-rank test for paired samples. For the feature comparison: H0: log-TDP does not perform better than log-BPE. H1: log-TDP performs better than log-BPE. For the comparison of spatial filters, we tested whether 11 laplacian channels and CSP outperformed 3 laplacian channels and whether CSP outperformed 11 laplacian channels.

Feature comparison			
	p-values in (%), limit of significance 1.70%		
	log-BPE vs log-TDP		
3 lap	0.62		
11 lap	0.10		
CSP	1.51		
Comparison of spatial filters			
	3 lap vs 11 lap	3 lap vs CSP	11 lap vs CSP
	0.023	0.11	15.09

measurement was performed based in log-BPE features and the results are biased toward this solution.

With respect to the spatial filters, surprisingly no significant differences were found between 11 laplacian channels and CSP filters. Interestingly, CSP filters performed worse than 11 laplacian channels in almost all subjects with poor control of the system, that is, over 30% of error (see Fig. 3). One can conjecture that CSP over fits in these cases due to its supervised nature, whereas the calculation of laplacian channels is unsupervised and can be more robust to non-stationarities. A similar effect, although not that clear, can be seen in the comparison of 3 laplacian channels with CSP.

Finally, the errors are obtained applying the classifier trained in the calibration to the feedback data. This is the reason for worse average values in this setting comparing to the first (in which the errors were computed by cross-validation in the feedback data, with the parameters selected in all subjects except the one in test). Further improvement can be expected by adapting the bias, in a supervised manner like in Shenoy, Krauledat, Blankertz, Rao, and Müller (2006) or in an unsupervised fashion as in Vidaurre, Schlögl, Blankertz, Kawanabe, and Müller (2008).

7. Conclusion

In this manuscript we presented the Time Domain Parameters, a set of features inspired by as well as defined in the time domain, Hjorth parameters. They are easy and fast to calculate and only a very low number of parameters needs to be selected. We combined TDP with the most usual spatial filtering methods to have a complete performance analysis. It was found that TDP improved significantly the band power estimates in two different experimental settings that are common in BCI research. The results show that TDP is a useful feature when no previous data from the subject is available, but also can be used in the transition from calibration to feedback measurement. Regarding spatial filters, subjects' performance with 3 laplacian channels is better than with 3 bipolar channels. In case of having the possibility to perform a calibration measurement, either 11 laplacian channels or CSP can be used. 11 laplacian channels will not over-fit, but CSP reduces the feature dimensionality. The choice will depend on the application and the amount of data available to estimate covariance matrices.

Acknowledgments

This work was supported by EU project 040666 Multiadaptive BCI, DFG (MU 987/3-1) VitalBCI Project, BMBF grant FKZ 01-IS07007A (ReMind), and by the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT-216886 and EU-TOBI Project ICT-224631. This publication only reflects the authors' views. The funding agencies are not responsible for any use that may be made of the information contained herein.

References

Allison, B., Wolpaw, E., & Wolpaw, J. (2007). Brain–computer interface systems: Progress and prospects. *Expert Review of Medical Devices*, *4*(4), 463–474.

Birbaumer, N., Kübler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser,, et al. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions of Rehabitation Engineering*, *8*(2), 190–193.

Birbaumer, N., Weber, C., Neuper, C., Buch, E., Haapen, K., & Cohen, L. (2006). Physiological regulation of thinking: Brain–computer interface (BCI) research. *Progress in Brain Research*, *159*, 369–391.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., & Curio, G. (2007). The non-invasive Berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, *37*(2), 539–550.

Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., & Müller, K.-R. (2008). The Berlin brain–computer interface: Accurate performance from first-session in BCI-naive subjects. *IEEE Transactions on Biomedical Engineering*, *55*(10), 2452–2462.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, *25*(1), 41–56.

Blankertz, B., & Vidaurre, C. (2009). Towards a cure for BCI illiteracy: Machine learning based co-adaptive learning. *BMC Neuroscience*, *10*(Suppl. 1), 85.

Boostani, R., & Moradi, M. H. (2004). A new approach in the bci research based on fractal dimension as feature and adaboost as classifier. *Journal of Neural Engineering*, *1*(4), 212–217.

Coyle, D., McGinnity, T.M., & Prasad, G. (2006). Creating a nonparametric brain–computer interface with neural time-series prediction preprocessing. In *Conf proc IEEE eng med biol soc. 2006*. 1, (pp. 2183–2186).

Dornhege, G., del R. Millán, J., Hinterberger, T., McFarland, D., & Müller, K.-R. (Eds.). (2007). *Toward brain–computer interfacing*. Cambridge, MA: MIT Press.

Elbert, T., Rockstroh, B., Lutzenberger, W., & Birbaumer, N. (1980). Biofeedback of slow cortical potentials. I. *Electroencephalography and Clinical Neurophysiology*, *48*, 293–301.

Hjorth, B. (1970). EEG analysis based on time-domain properties. *Electroencephalography and Clinical Neurophysiology*, *29*, 306–310.

Krausz, G., Scherer, R., Korisek, G., & Pfurtscheller, G. (2003). Critical decision-speed and information transfer in the Graz brain–computer interface. *Applied Psychophysiology and Biofeedback*, *28*(3), 233–240.

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychology Bulletin*, *127*(3), 358–375.

Ledoit, O., & Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, *30*, 110–119.
Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.
McFarland, D. J., McCane, L. M., David, S. V., & Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*, *103*, 386–394.
Obermaier, B., Guger, C., Neuper, C., & Pfurtscheller, G. (2001). Hidden markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, *22*(11), 1299–1309.
Pfurtscheller, G., Neuper, C., & Birbaumer, N. (2005). Human brain–computer interface. In A. Riehle, & E. Vaadia (Eds.), *Motor cortex in voluntary movements* (pp. 367–401). New York: CRC Press, (chapter 14).
Rockstroh, B., Birbaumer, N., Elbert, T., & Lutzenberger, W. (1984). Operant control of EEG and event-related and slow brain potentials. *Biofeedback and Self-Regulation*, *9*(2), 139–160.
Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. In *Statistical applications in genetics and molecular biology* (pp. 4–32).
Scherer, R., Lee, F., Schlögl, A., Bischof, H., & Pfurtscheller, G. (2008). Towards self-paced brain–computer communication: Navigation through virtual worlds. *IEEE Transactions on Biomedical Engineering*, *55*, 675–682.
Schlögl, A., Flotzinger, D., & Pfurtscheller, G. (1997). Adaptive autoregressive modeling used for single-trial EEG classification. 42, 162–167.
Schlögl, A. (2000). *The electroencephalogram and the adaptive autoregressive model: Theory and applications*. Aachen, Germany: Shaker Verlag.
Schlögl, A., Lee, F. Y., Bischof, H., & Pfurtscheller, G. (2005). Characterization of four-class motor imagery eeg data for the bci-competition 2005. *Journal of Neural Engineering*, *2*(4), L14–L22.
Schlögl, A., & Brunner, C. (2008). Biosig: A free and open source software library for bci research. *Computer*, *41*(10), 44–50.
Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P. N., & Müller, K.-R. (2006). Towards adaptive classification for BCI. *Journal of Neural Engineering*, *3*(1), R13–R23.
Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2005). Adaptive on-line classification for EEG-based brain computer interfaces with AAR parameters and band power estimates. *Biomedizinische Technik. Biomedical Engineering*, *50*(11), 350–354.
Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2006). A fully on-line adaptive BCI. *IEEE Transactions on Biomedical Engineering*, *53*(6), 1214–1219.
Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2007). Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Transactions on Biomedical Engineering*, *54*(3), 550–556.
Vidaurre, C., Schlögl, A., Blankertz, B., Kawanabe, M., & Müller, K.-R. (2008). Unsupervised adaptation of the LDA classifier for Brain–Computer Interfaces. In *Proceedings of the 4th international brain–computer interface workshop and training course 2008* (pp. 122–127). Verlag der Technischen Universität Graz.
Vourkas, M., Micheloyannis, S., & Papadourakis, G. (2000). Use of ann and hjorth parameters in mental-task discrimination. In *Advances in medical signal and information processing, 2000. First international conference on (IEE conf. publ. no. 476)* (pp. 327–332).
Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–Computer Interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791.