Fast Feature Selection to Compare Broadband with Narrowband Phase Synchronization in Brain-computer Interfaces

E. Gysels¹, P. Renevey¹, P. Celka²
¹Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland
²Gold Coast Campus, Griffith University, School of Engineering, Gold Coast, Queensland, Australia

Summary
Objective: Brain-computer interface (BCI) research aims at developing communication devices for the motor disabled. Such devices are not driven by muscle activity, but by brain activity recorded during different mental tasks. We present here the comparison of phase synchronization and power spectral density (PSD) features, computed from broadband and narrowband filtered EEG signals and their ability to discriminate three mental tasks.
Methods: EEG signals were recorded from five subjects while performing left and right hand movement imagination and word generation. We applied a modified Fast Correlation Based Filter (FCBF) [9] for the purpose of feature selection.
Results: We found that the features were selected from electrode signals corresponding to neurophysiological evidence, i.e. electrodes lying over the motor cortex. PSD and phase locking value (PLV) features were more discriminative when computed from narrowband (8-12 Hz) and broadband (8-30 Hz) filtered signals respectively.
Conclusions: The generalization performance is as good as the one obtained with SVM-rfe, but this algorithm is faster and selects fewer features. These properties may make FCBF a valuable tool for further improvement of BCIs.

Keywords
Brain-computer interface, broadband phase synchronization, EEG, Fast Correlation Based Filter

1. Introduction
Brain-computer interfaces (BCIs) offer new communication opportunities to the motor disabled [1, 2]. Such communication does not depend on the brain’s normal output pathways of peripheral nerves and muscles, but is based on analysis of brain activity generated during different mental tasks. Different mental tasks correspond to different output commands when recognized by the system.

Actual BCI devices [3] allow for controlling cursor movement, spelling letters with a virtual keyboard, driving a wheelchair and operating a neuroprosthesis.

Surface EEG-based interfaces commonly use features extracted from spectral estimates and autoregressive coefficients. This allows for studying typical rhythms associated to different mental tasks. In [4] we proposed to use synchronization measures in BCIs in order to take into account the coupling between different brain regions during mental tasks.

In this work, we study the phase locking value (PLV) computed from broadband and narrowband filtered EEG signals by means of the Fast Correlation Based Filter (FCBF) feature selection.

This paper presents an offline analysis aiming to:
- evaluate FCBF in the framework of BCIs,
- compare synchronization and power-based features derived from broadband and narrowband filtered EEG signals,
- study from which EEG signals the selected features were derived,
- investigate the classification performance.

The data were recorded without feedback, in order to make sure the subjects could not adapt to any features or classification algorithms that would be used to provide feedback.

2. Recording Protocol
Five healthy right-handed persons (two women and three men) took part in the experiments. None of the subjects had participated in BCI experiments before. For each subject, a total of 60 minutes of EEG has been recorded during five sessions of four minutes on three consecutive days. The mental tasks performed, with opened eyes, were imagination of left hand movement (‘L’), imagination of right hand movement (‘R’) and word generation (‘W’). During a session, about every 20 seconds the operator informed the subject which task they had to perform.

EEG signals were recorded with the 32-channel Biosemi ActiveTwo system. Electrodes were placed on the scalp according to the 10-20 international electrode placement system (see Fig. 1). Two separate electrodes replace the ground electrode. They form a feedback loop, decreasing the effective impedance with a factor 100 at 50 Hz. EEG signals were digitized at a sampling rate fₛ = 512 Hz and stored for offline analysis.

3. Methods
In this section we will describe the pre-processing, the features extracted from the recorded EEG signals, the feature selection method and the classifier used to classify the selected features.
3.1 Data Preprocessing

Before computing the PLV features, in each window the mean of the signals was subtracted, and subsequently a linear phase FIR 8-30 Hz band-pass filter was applied.

The preprocessing of the PSD features included subtracting the mean, applying a Laplacian filter and a 4-40 Hz linear phase FIR band-pass filter. Finally, the signal was resampled at 1:4 of the original sampling rate to reduce the variance of the PSD coefficients.

For training the classifier, windows containing eye blinking artifacts were automatically detected with the method described in [5] and rejected. For testing, all windows were used, including the ones containing artifacts.

3.2 Feature Extraction

Features were computed from $T = 1$ s sliding windows with 7/8 s overlap. We considered two frequency bands (FBs), the $\alpha$ (8-12 Hz) and $\lambda$ (8-30 Hz) band. These frequency bands were chosen because of their involvement in the motor imagery tasks [6]. For both FBs, we extracted the following features from the preprocessed EEG signals:

The PLV [7] was computed for the 496 possible pairs of electrode signals, filtered in the FBs $\alpha$ and $\lambda$. The PLV characterizes by a value between 0 and 1 the stability of the phase difference between the phases $\varphi_x(t)$ and $\varphi_y(t)$ of the two signals $s_x(t)$ and $s_y(t)$ across subsequent time samples of one window as follows:

$$ PLV = \left| \left[ e^{j \varphi(t)} \right]_T \right|, $$

where $\Delta \varphi(t) = \varphi_x(t) - \varphi_y(t)$, $T$ is the length of the window, and $\langle \cdot \rangle_T$ is the averaging operator. We compute the instantaneous phase $\varphi(t)$ by means of the Hilbert transform. The Hilbert transform of $s(t)$ is given by

$$ \tilde{s}(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{s(t')}{t - t'} dt', $$

where p.v. denotes the Cauchy principal value, and allows us to determine the instantaneous phase $\varphi$, as follows:

$$ \tilde{s}(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{s(t')}{t - t'} dt', $$

From the PSD, we computed the band power (BP) in the $\alpha$ and $\lambda$ FB, normalized with the total power.

The normalized statistical mean frequency (SMF) was computed using the PSD estimated at the frequencies between 8 and 30 Hz:

$$ SMF = \frac{1}{(30 - 8)} \left( \int_{30Hz}^{30Hz} P(f) df - 8 \right), $$

where $P$ is the estimation of the PSD at a given frequency $f$. The SMF is independent of the FB used to compute the other features.

Thus, for each 1-second EEG window, we compute for each of the 32 electrode signals the PLV with the signals recorded from all other electrodes, filtered in FB $\alpha$ or $\lambda$ (32 groups of features); the BP in the frequency band FB $\alpha$ or $\lambda$ (one group); and the normalized SMF (one group). Each of these 34 groups of features consists of 32 values (corresponding to the 32 electrodes).

3.3 Feature Selection

We previously used genetic algorithms [8] and SVM-based recursive feature elimination, SVM-rfe, [5], for the task of feature selection. In this work, we select features with a modified version of the algorithm presented by Yu and Liu, the “Fast Correlation Based Filter” (FCBF) [9]. Not only does the algorithm determine the relevant features, it also includes a redundancy analysis. First, the relevant features are determined from the symmetrical uncertainty between each feature and the class:

$$ U(X, Y) = 2 \left( \frac{H(X) - H(X|Y)}{H(X) + H(Y)} \right), $$

where $H(X)$ is the entropy of variable $X$ and $H(X|Y)$ is the entropy of $X$ after observing values of $Y$. Second, redundant features are detected by means of approximate Markov blankets and removed. A feature $M_i$ is said to be a Markov blanket for a feature $F_j$, if $M_i$ subsumes all information that $F_j$ has about the class and about all of the other features.

Application of this method to noisy EEG data resulted in almost all features being removed. This is because the correlation between the individual features and the class is rather low. Therefore, we did not remove all features for which a considered predominant feature was an approximate Markov blanket, i.e., we did not remove all features for which the correlation with the predominant feature was greater than the correlation of the feature with the class. We removed only the 25% of features for which the difference between the correlation with the predominant feature and the correlation with the class was greatest.

This algorithm was applied in a two-step feature selection procedure. In a first step, features were selected within each of the 34 groups. In a second step, a final feature subset was selected amongst the features selected in the different groups. This was done for each of the three pairs of tasks.

3.4 Classification

The classifier consists of a combination of three support vector machines (SVMs), each distinguishing one of the three pairs of tasks, which allows for classification in three classes.

Every second, strict majority voting was applied to the eight classification results to determine the class (‘L’, ‘R’ or ‘W’). If no strict majority was attained, the window was categorized as ‘unknown’.

For doing the feature selection in this study, about half of the data was used to actually select the features and train the classifier, resulting in $CR_{\text{training}}$ and $ER_{\text{training}}$. The generalization performance of the resulting classifier was then tested on the remaining, unseen, data ($CR_{\text{test}}$ and $ER_{\text{test}}$). We use CR to denote the correct classification rate and ER to denote the error rate. The remaining percentage is the rate of unknown responses.

The data were studied per day and per subject.
We performed a paired Student’s t-test to test for significant differences between the two frequency bands for the different groups of features. We used a 5% level of significance to reject the null hypothesis of equal performances for both FBs. The results obtained with the selected features of each group for the three days and the five subjects were concatenated to perform the significance test. The significance test was repeated for the three pairs of tasks.

We found that if there was a significant difference in performance between the α- and the λ-frequency bands, the λ-band was the better one for the PLV features and the α-band was the better one in case of PSD features.

Generally (averaged over all subjects all days), C4 was the most often selected electrode for the distinction of the task pairs ‘L’-‘R’ and ‘L’-‘W’ and FC1 for ‘R’-‘W’. Contrary to [5], for ‘L’-‘R’ and ‘L’-‘W’ most of the electrode pairs whose PLV was selected were in the right hemisphere. For ‘R’-‘W’ pairs were mostly selected in the left hemisphere. Considering PLV and PSD features separately learned that in case of PLV the most often selected electrodes are C4, CP2/C4 and FC1 for the task pairs ‘L’-‘R’, ‘L’-‘W’ and ‘R’-‘W’ respectively. For PSD, most often selected electrodes were C4, C4 and F7/FC1/C3 for the three respective task pairs.

The occurrence of the 32 electrode signals in the final feature subset for the three task pairs is reflected in Figure 2, for PSD-based features (BP and SMF) and PLV features respectively. Brighter regions correspond to more often selected electrodes.

Table 1 shows the three-class classification accuracies for the five subjects. As compared to results in [5], the classification accuracy on training data is much lower for FCBF than for SVM-rfe. The average classification accuracy on unseen test data was 49.0% in case of SVM-rfe feature selection [5], obtained with in the average 117 features. FCBF yields an average classification accuracy of 48.6% on test data, but obtains it with in the average 19 features only. FCBF is less prone to overfit the data.

In addition, FCBF is very fast and the selected features correspond better to neurophysiological evidence.

4. Results

We performed a paired Student’s t-test to test for significant differences between the two frequency bands for the different groups of features. We used a 5% level of significance to reject the null hypothesis of equal performances for both FBs. The results obtained with the selected features of each group for the three days and the five subjects were concatenated to perform the significance test. The significance test was repeated for the three pairs of tasks.

We found that if there was a significant difference in performance between the α- and the λ-frequency bands, the λ-band was the better one for the PLV features and the α-band was the better one in case of PSD features.

Generally (averaged over all subjects all days), C4 was the most often selected electrode for the distinction of the task pairs ‘L’-‘R’ and ‘L’-‘W’ and FC1 for ‘R’-‘W’. Contrary to [5], for ‘L’-‘R’ and ‘L’-‘W’ most of the electrode pairs whose PLV was selected were in the right hemisphere. For ‘R’-‘W’ pairs were mostly selected in the left hemisphere. Considering PLV and PSD features separately learned that in case of PLV the most often selected electrodes are C4, CP2/C4 and FC1 for the task pairs ‘L’-‘R’, ‘L’-‘W’ and ‘R’-‘W’ respectively. For PSD, most often selected electrodes were C4, C4 and F7/FC1/C3 for the three respective task pairs.

The occurrence of the 32 electrode signals in the final feature subset for the three task pairs is reflected in Figure 2, for PSD-based features (BP and SMF) and PLV features respectively. Brighter regions correspond to more often selected electrodes.

Table 1 shows the three-class classification accuracies for the five subjects. As compared to results in [5], the classification accuracy on training data is much lower for FCBF than for SVM-rfe. The average classification accuracy on unseen test data was 49.0% in case of SVM-rfe feature selection [5], obtained with in the average 117 features. FCBF yields an average classification accuracy of 48.6% on test data, but obtains it with in the average 19 features only. FCBF is less prone to overfit the data.

In addition, FCBF is very fast and the selected features correspond better to neurophysiological evidence.

5. Discussion

In the framework of a brain-computer interface, we have applied a modified FCBF-
algorithm in a two-step feature selection procedure to select features and compare the classification performance achieved with PSD and PLV features computed from broadband and narrowband filtered EEG signals. To our knowledge, this method has not yet been applied to BCIs.

FCBF is a very fast feature selection method. It is a so-called filter method, i.e., it does not use a classifier to select the features. FCBF uses the symmetrical uncertainty to determine the relevant features and remove the redundant ones.

Fast feature selection methods may allow for improvement of BCIs. Between two sessions new, actual, best features could be selected and used for classification in the next session.

In addition, FCBF selects significantly less features, allowing for an additional acceleration of the classification algorithm.

A possible disadvantage of filter methods could be that the selected features are not well suited to the classifier, but in our experiments no significant generalization performance decrease has been detected, compared to the previously used SVM-rfe method [5].

Both PSD and PLV features were mostly selected from the central electrode signals. Both types of features point to the same brain region, the motor cortex. In case of PLV, the involved brain region seems to be wider, including also more frontally located areas.

The correspondence between the selected electrode signals and the brain regions predicted by neurophysiology to be involved in the considered mental tasks makes FCBF a valuable analysis tool. It could be used to investigate new protocols where we do not know a priori which brain regions are involved and which type of features carry the most discriminative information.

For the two frequency bands considered, 8-12 Hz and 8-30 Hz, PSD features yielded better results in the narrower 8-12 Hz frequency band. This frequency band corresponds to the \( \mu \)-rhythm, associated with executed or imagined motor tasks [6].

On the contrary, in case of PLV features better results were obtained using broadband, 8-30 Hz, filtered EEG signals. In this case the phase does not have a clear physical meaning. We think PLV could be more discriminative in case of broadband signals, because then synchronization is more difficult to achieve. Filtering the EEG signals in narrow frequency bands, removes the components that could desynchronize the two signals.

### Table 1

Comparison of Phase Synchronization and PSD Features

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRtrain (%)</td>
<td>66.5</td>
<td>52.6</td>
<td>70.3</td>
<td>51.1</td>
<td>57.9</td>
</tr>
<tr>
<td>CRtest (%)</td>
<td>28.0</td>
<td>41.5</td>
<td>23.6</td>
<td>41.1</td>
<td>35.5</td>
</tr>
<tr>
<td>CRmed (%)</td>
<td>56.4</td>
<td>36.8</td>
<td>61.9</td>
<td>40.1</td>
<td>47.6</td>
</tr>
<tr>
<td>CRerr (%)</td>
<td>35.9</td>
<td>57.6</td>
<td>30.7</td>
<td>51.1</td>
<td>44.4</td>
</tr>
</tbody>
</table>

6. Conclusion

In the framework of a brain-computer interface, we have applied a modified FCBF-algorithm in a two-step feature selection procedure to data recorded during the mental tasks left and right hand movement imagination and word generation. The algorithm selected features in the brain regions predicted by neurophysiological evidence to be involved in these tasks. Better classification results were obtained using PSD and PLV computed from signals filtered in the narrower frequency band 8-12 Hz and the broader 8-3 0Hz frequency band respectively. The generalization performance is as good as the one obtained with SVM-rfe, but this algorithm is faster and selects fewer features. These properties may make FCBF a valuable tool for further improvement of BCIs.

### Acknowledgments

This work was funded by the Swiss National Science Foundation through the National Centre of Competence in Research on ‘Interactive Multimodal Information Management’ (IM2). Data were provided by S. Chiappa and J. del R. Millán.

### References


### Correspondence to:

Philippe Renevey
CSEM, Swiss Center for Electronics and Microtechnology
Rue Jaquet-Droz 1
2007 Neuchâtel
Switzerland
E-mail: Philippe.Renevey@csem.ch