



# Supervision of time-frequency features selection in EEG signals by a human expert for brain–computer interfacing based on motor imagery

Alban Duprès, François Cabestaing, José Rouillard

## ► To cite this version:

Alban Duprès, François Cabestaing, José Rouillard. Supervision of time-frequency features selection in EEG signals by a human expert for brain–computer interfacing based on motor imagery. Systems, Man, and Cybernetics (SMC), Oct 2016, Budapest, Hungary. hal-01394149

**HAL Id: hal-01394149**

**<https://hal.science/hal-01394149>**

Submitted on 8 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Supervision of time-frequency features selection in EEG signals by a human expert for brain–computer interfacing based on motor imagery.

Alban Duprès, François Cabestaing, José Rouillard

Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL

Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

{firstname.lastname@univ-lille1.fr}

**Abstract**—In the context of brain–computer interfacing based on motor imagery, we propose a method which allows an expert to select manually time-frequency features. This selection is performed specifically for each subject, by analysing a set of curves that emphasize differences of brain activity recorded from electroencephalographic signals during the execution of various motor imagery tasks. We will show that expert knowledge is very valuable to supervise the selection of a sparse set of significant time-frequency features. Features selection is performed through a graphical user interface to allow an easy access to experts with no specific programming skills. In this paper, we compare our method with three fully-automatic features selection methods, using dataset 2A of BCI competition IV. Results are better for five of the nine subjects compared to the best competing method.

**Index Terms**—brain–computer interface, motor imagery, EEG signal processing, sparse feature set, feature selection, human expertise.

## I. INTRODUCTION

Since the first works of Vidal [1], researchers paid increasingly attention on brain activity used as a communication channel. Brain–Computer Interfaces (BCI) are devices that analyse a user’s brain activity in real-time in order to translate his/her intent into one or more control signals. It allows him/her to control an application, bypassing normal output pathways that are muscles and peripheral nerves. Non invasive BCI are promising for patients suffering from severe motor disabilities, allowing them to recover a partial autonomy. Such a device uses different kind of neurophysiological patterns in order to identify user’s mental states. Event related desynchronization/synchronization (ERD/ERS) elicited during motor imagery (MI) are the most commonly used patterns in BCI. Users consciously control ERD/ERS patterns at their own pace, i.e by imaging that they realize a specific motor action, independently from external events [2]. In order to discriminate different MI tasks performed by the user, appropriate features have to be defined from electroencephalographic (EEG) signals.

ERD and ERS translate changes of activity in the primary motor cortex [3]. During a MI task, desynchronization of neural activity induces a decrease of EEG power, characterizing an ERD. ERS patterns are characterized by an increase of EEG power, induced by a synchronization of neural activity at the end of MI. Spatial distribution of ERD/ERS patterns depends

on the MI task, following the cortical motor homunculus [4]. ERD and ERS are known to be recorded in frequency bands that are mu (7 - 13 Hz) and beta (13 - 25 Hz) rhythms. Nevertheless frequency bands and spatial locations that are the most relevant, depend on each subject and MI task.

The EEG signal processing usually includes a spatial filtering stage in order to increase the signal to noise ratio and therefore to facilitate the discrimination of different MI tasks. CSP (Common Spatial Patterns) is an effective linear spatial filtering method, widely used for MI-based BCI [5], involving user-specific parameters. A user-tuned CSP filter weights EEG channels in order to maximize the variance of EEG signal for one MI task while minimize the variance of EEG signal for other MI tasks or a idle mental state. The actual discriminative performance of CSP filters depends on the frequency band and on the time interval from which CSP features are extracted [6]. For instance, CSP filters computed on unfiltered signals or with inappropriate frequency bands yield poor discrimination of mental states.

To solve this problem, several methods have been described in the literature. Some approaches aim to improve the spatial filtering stage, using a wide frequency band, i.e. encompassing mu and beta rhythms and therefore valid for any user. For instance, [7] has compared performances of different regularized CSP (RCSP), showing that the best RCSP methods outperform classification accuracy of a basic CSP. Other researchers paid more attention on the selection of user-specific frequency bands in which ERD/ERS detection is more effective. For example, [8] proposed a filter bank common spatial pattern, which bandpass-filter EEG signals into multiple frequency bands. Then CSP features are extracted for each band and a feature selection algorithm keeps the most relevant frequency/CSP features for a given user. Later, [9] proposed a method of selecting subject specific frequency bands, based on the analysis of a channel-frequency map, for multiclass MI classification. Unlike [8], this approach selects frequency bands whose size is specific for each subject and MI. Recently, [10] proposed a method called Common Spatial-Spectral Boosting Pattern (CSSBP), which combines CSP filtering stage with a boosting strategy, to perform selection of key channels and frequency bands.

However, all the described techniques perform an au-

tonomous features selection stage, by defining empirically several meta-parameters, such as the number of features. But the *sparsity* of the feature space is not defined by taking into account some *a priori* neurophysiological knowledge during this feature selection stage. Studies have shown that the sparsity of the feature space decreases the sensitivity of the BCI to covariate shifts in EEG signals, allowing for good classification results. For example, Raza *et al.* aimed to increase the sparsity using techniques such as forward-elimination or backward-elimination of features in the space [11]. Nevertheless, they do not use a detailed neurophysiological-based analysis of ERD/ERS patterns in order to select the most relevant frequency bands, spatial locations, and time intervals in which EEG signals should be analysed.

In this paper we propose a method based on a human expert analysis to select a small number of time-frequency features during a first stage. Obviously, the raw EEG signal is not analysed by the expert, which would involve a very high level of expertise, but by plotting a set of curves using samples of specific time-frequency processed EEG signals. Then CSP filters are computed for each selected time-frequency features. This approach yields a sparse features space, which is defined by a human expert and not by blindly setting few meta-parameters in a fully automatic selection technique. Moreover, neurophysiological informations gathered during a MI task thanks to our approach are useful to understand performances of each subject.

## II. METHOD

Since our method is supervised, different processing stages are performed for the training mode as well as for the online processing mode. The training mode aims to select time-frequency intervals by analysing a set of labelled EEG trials. So the BCI protocol must be synchronous to build a set of training data. Users are informed about the period and the kind of MI to perform. During the online mode the paradigm is not necessary synchronous, users can freely perform MI tasks as their own pace.

The processing pipeline for the training mode, illustrated in figure 1, involves four successive stages: spatial filtering, band-pass frequency filtering, power estimation, and aggregation over trials. As we want to implement the same signal processing pipeline for all users in the training mode, we use the surface Laplacian for the initial spatial filtering stage. It allows spatial artefacts removal and source identification, without involving user-specific parameters [12]. Then, in order to highlight user specificities in the frequency domain, we use as other authors a bank of band-pass filters encompassing mu and beta rhythms [8], [11]. In order to show ERD/ERS time course, we compute the log-variance – equivalent to the power – of filtered signals in a sliding window of fixed duration. Finally, to better highlight ERD/ERS patterns and therefore help data analysis by the human expert, we compute the average and standard-deviation of the signal power, over all trials of each MI task.

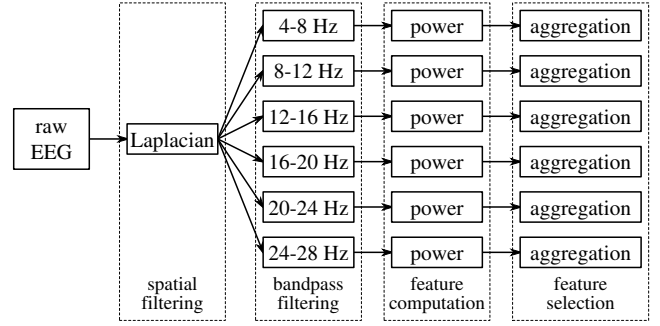


Fig. 1: Processing pipeline for the training mode.

Our method is useful to select time-frequency intervals by analysing neurophysiological patterns related to each MI task. The expert reviews, for each MI task, a set of curves that represent time-course of the EEG signal power filtered with each frequency band and known to be correlated to a specific motion. In order to highlight differences between a MI task and other MI tasks in the time-frequency domain, each curve of interest is compared to an additional baseline curve. This baseline curve represents the EEG signal power under review, averaging over all the trials corresponding to other MI tasks, i.e. following a one-versus-rest-strategy. By analysing this set of curves, the expert can select time-frequency intervals that he/she considers as the most discriminant between a particular MI and others. The first criterion that the expert pays attention to concerns the ERD and ERS patterns. He selects the time-frequency intervals where such patterns are more pronounced. Then the other criterion, enabling to refine the set of intervals, is the performance of classification.

For example, figure 2 shows a set of six curves, one for each frequency band, in order to exhibit differences between left hand versus right hand MI. The EEG signal recorded at location C4 over the right sensorimotor cortex is known to highlight ERD patterns during left hand MI. For this MI task, the curve of interest in each frequency band (solid red lines in figure 2) corresponds to the time-course of the EEG signal power, averaged over all corresponding trials, after spatial noise removal by a Laplacian filter. Power is computed in a time window starting one second before MI onset and ending one second and half after MI offset. Baseline curve in each frequency band (blue dotted lines in figure 2) is the time-course of the power from the same signal, averaged over all trials that do not correspond to the MI task under review. In order to bring to the expert the statistical significance of signal variations, we display two additional curves around each curve at plus and minus half standard deviation. Figure 3 shows different sets of time-frequency intervals that the expert can select, according to figure 2, as the best features to discriminate left hand versus right hand MI. Each set is represented by a time-frequency map, in which white strips delimit selected features, and its classification accuracy (in %). The set with the best trade-off between performance and sparsity should be selected for the online mode. In this case sparsity and

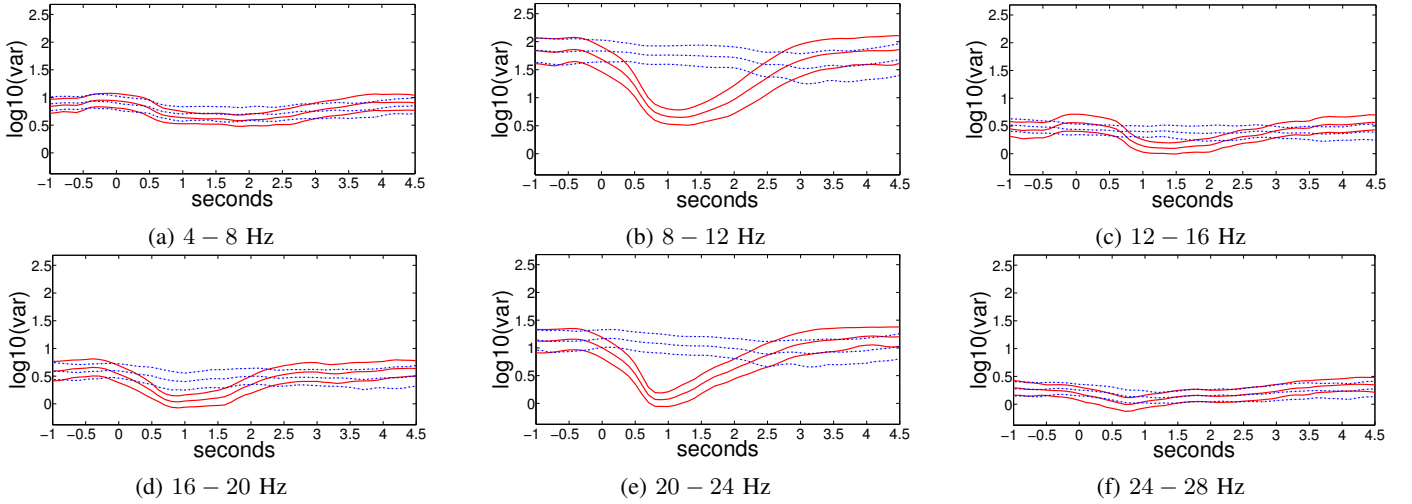


Fig. 2: Set of curves for expert review of left hand motor imagery, electrode C4

performance are convergent, allowing to select easily the second set. Nevertheless, when those parameters are divergent, sparsity is favoured in order to decrease covariate shifts in EEG signal, provided that performances are not too degraded.

were selected as the most relevant for each frequency band by the expert. Finally, features are used as input of a LDA classifier. The online processing pipeline can be performed in an asynchronous mode, using a sliding window.

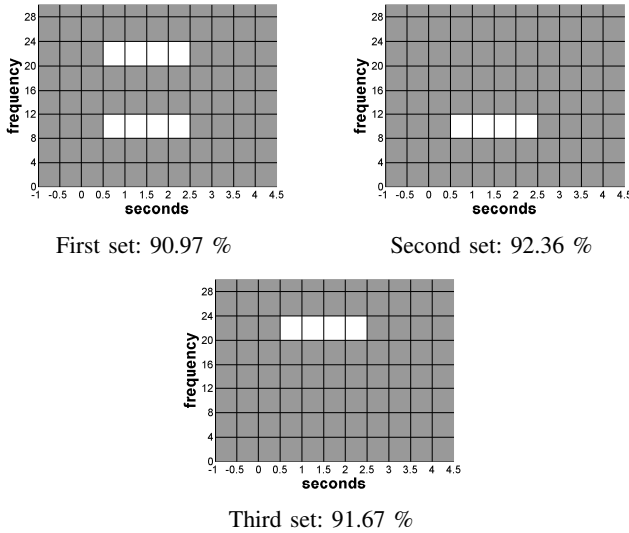


Fig. 3: Time-frequency features of subject 9 and their corresponding classification accuracy.

Then, user-tuned CSP filters are computed using EEG signals from each selected time-frequency interval, in order to optimize the spatial filtering stage compared to the non-adaptive Laplacian. As recommended by [13], we keep three pairs of CSP filters in order to get a compromise between performance and classifier overfitting.

Figure 4 depicts the processing pipeline for the online mode, including four successive stages. First, EEG signals are filtered in the spatial and frequency domains by CSP/band-pass filters pairs defined during the training mode. Then, the log-variance of filtered signals are computed over the time intervals that

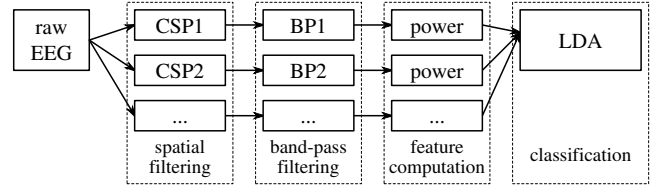


Fig. 4: Processing pipeline for the online mode.

When the BCI protocol involves more than two MI tasks, one processing pipeline is implemented for each MI. As we perform a one-versus-rest strategy for the time-frequency selection, the same approach is used for the classification stage. The confidence score, for instance the distance between the feature vector and the hyperplane, that each classifier outputs is used as input to a standard voting procedure to determine the most appropriate class.

### III. RESULTS AND COMPARISON

In order to assess our method, we used EEG signals publicly available in data set 2A of BCI competition IV, which has been widely used for comparison purposes [14]. This data set comprises raw EEG signals recorded by 22 electrodes from 9 subjects who performed left hand, right hand, feet, and tongue MI. All these MI tasks were performed during four seconds after the presentation of a cue. Each user achieved two sessions on different days in order to obtain a training data set and an evaluation data set, both containing 72 trials for each MI.

Our method is compared with those described in [8], [9], and [10], that is why we used in this study EEG signals recorded during left hand, right hand, feet, and tongue MI. For the training mode, the EEG signals are spatially filtered with surface Laplacians centred on locations C3, C4, and Cz.

EEG signals from C3 (resp. C4, and Cz) are known to exhibit an ERD pattern when the user performs right hand (resp. left hand and feet) MI. According to [15], EEG signals from C3 and C4 are known to exhibit an enhancement of the mu rhythm during tongue and feet MI. Each signal was filtered in the frequency domain by a bank of six band-pass filters (5<sup>th</sup> order Butterworth), yielding eighteen signals of interest. Finally, their log-variance was computed over a sliding window, and aggregated over trials to determine time-course of average and standard deviation.

For each MI the expert reviewed curves of each frequency band in order to keep the most discriminant time-frequency intervals between one MI and others. Using data of the training set, CSP filters are then determined for each time-frequency intervals, LDA classifiers trained, and kappa scores computed. As we have one classifier for each MI, trials are labelled by the classifier with the highest confidence score.

Using the online processing pipeline described in figure 4, our method was compared to: 1) the Event-Related Brain Dynamics Analysis (ERBDA) [9]; 2) the Filter Bank Common Spatial Pattern algorithm (FBCSP) [8]; 3) the Common Spatial-Spectral Boosting Pattern (CSSBP) [10]. For those methods CSP filters are computed on user-specific frequency bands and time interval between 0.5 and 2.5 s after the cue.

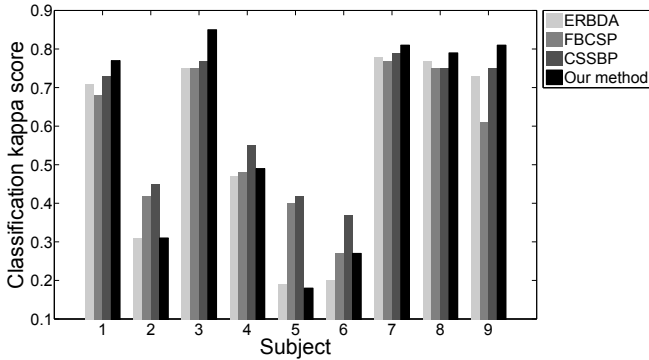


Fig. 5: Kappa scores for each subject.

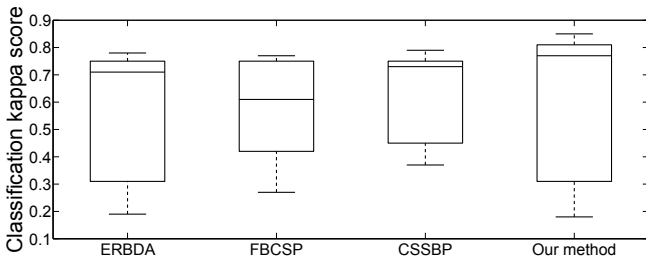


Fig. 6: Box plots for each method.

As these methods are assessed using kappa score, we use the same performance indicator for comparison purpose, as shown in figure 5. Our method (mean kappa score  $m = 0.59 \pm 0.27$  std) outperforms the best competing method for 5 of the 9 subjects. According to these results, our method is mainly profitable for subjects 3 and 9 for whom kappa score increases

respectively by 0.08 and 0.06 compared to the best competing method. On the other hand, performances of subjects 2 and 5 decrease respectively by 0.24 and 0.14 compared to the CSSBP method. In figure 6, we showed the box plots for each method. Although our method presents the best median, an important difference of performance is observed between subjects, as illustrated by the value of the first and the third quartiles.

Figure 7 shows performance of each classifier, using our method. We observe for subjects 1, 3, 7, 8, and 9 that at least three classifiers have kappa score above 0.81. This threshold is considered as an almost perfect result according to [16].

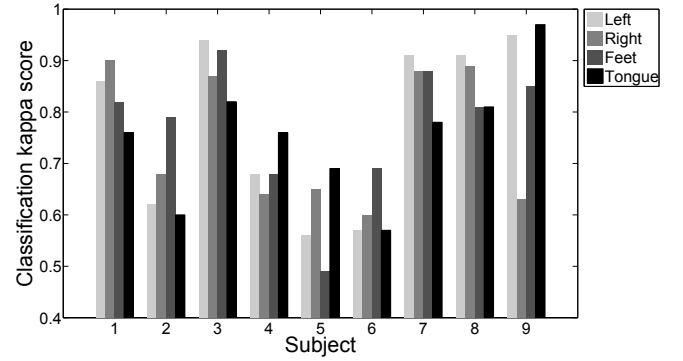


Fig. 7: Kappa scores for each MI.

The first time-frequency interval that our expert considered as the most discriminant between MI under review (red solid lines) and other MI (blue dashed lines) are shown in figure 8 for subjects 2, 3, 5, and 9. For each frequency band, the time interval selected by our expert is delineated by vertical black dashed lines. According to this figure, differences between MI are more pronounced for subjects 3 and 9 than for subjects 2 and 5.

#### IV. DISCUSSION

According to figure 5, our method outperforms others for five of the nine subjects, demonstrating its efficiency to deal with a multi-class classification problem in the context of MI-based BCI. Performances of subjects 3, 7, and 9 are above 0.81, using our method. As mentioned, this threshold is considered as an almost perfect result, expecting that those subjects have a better control of the BCI in an online session than other subjects.

Nevertheless, in the other cases, especially for subjects 2 and 5, the expert supervised selection is not superior to an automatic method such as the CSSBP. It is all the more a drawback that their performance are already low, compared to other subjects. As shown in figure 8, differences between MI are not enough highlighted for subjects 2 and 5 to easily select the most relevant time-frequency intervals. It explains why the expert choose a standard time-frequency interval, encompassing  $\mu$  and  $\beta$  rhythms, for tongue motions of subject 2. Therefore in these cases an automatic machine learning method seems to be more adapted to select features that best discriminate the different MI.

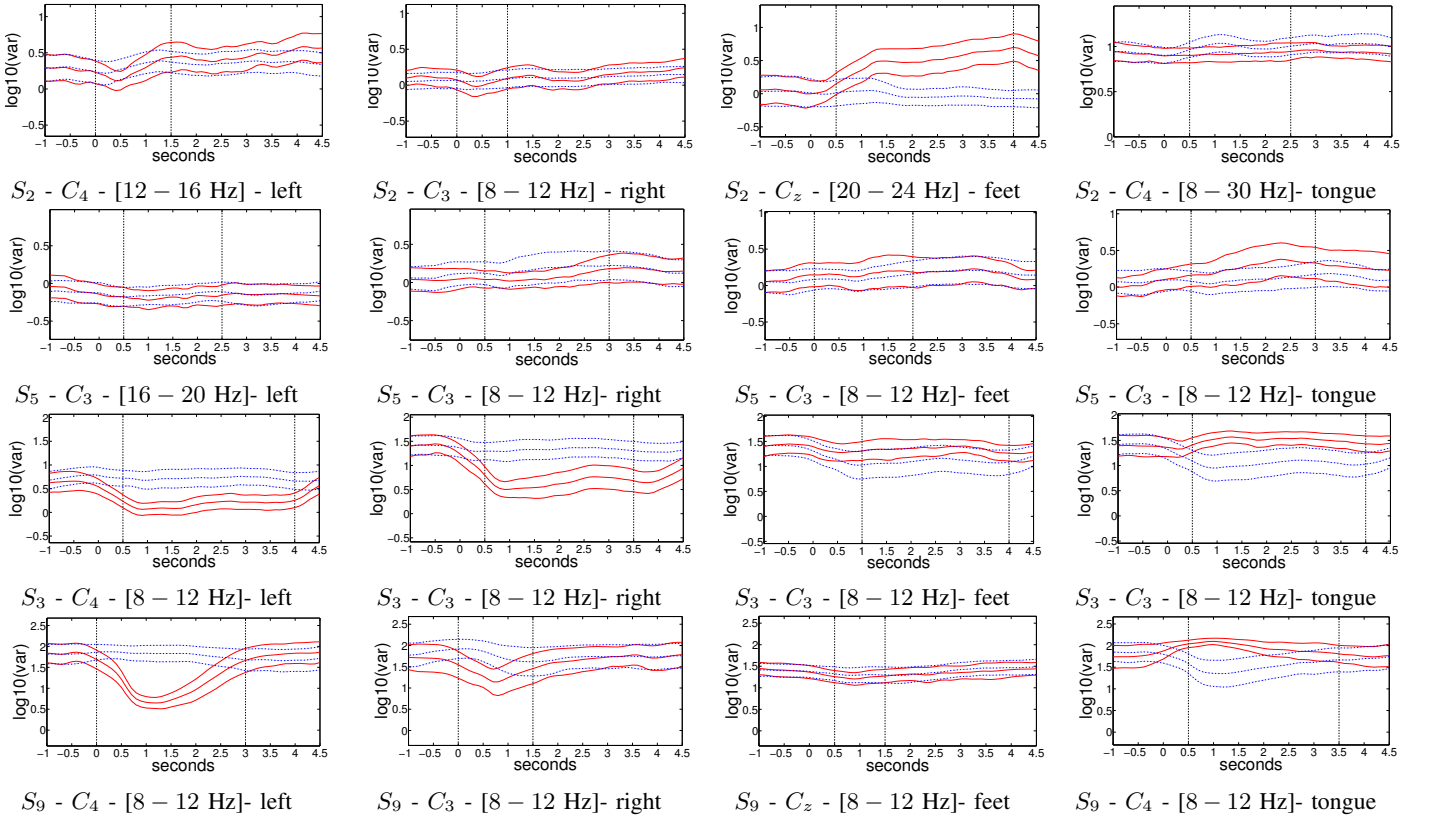


Fig. 8: First time-frequency feature for each MI reviewing and subjects 2, 3, 5, and 9, the vertical black dashed lines indicate the time interval

However, in order to assess the relevance of the features, the performance is not the only parameter to take into account. The sparsity of the feature space is also an important parameter to consider during the features selection step. Indeed the use of a sparse set of features enables to prevent the effect of over fitting, that an automatic method is more prone to. Although over fitting can obtain better performances on a single session, it decreases the stability of the features over time. Moreover the selection of the most discriminant features, using an automatic method, can be due to temporary artefacts that disappear in the next session. That is why we expect by selecting a set of time-frequency intervals based on neurophysiological knowledge, that features are more relevant and stable.

Results presented in figure 7 are useful to analyse performances of each subject. We observe that subjects for whom classifiers have the highest kappa scores perform better than other subjects. That is why subjects 1, 3, 7, 8, and 9, who have at least three classifiers with performance above 0.81, perform better than subjects 2, 4, 5, and 6. Figure 8 is interesting to bring explanation, from a neurophysiological point of view, about classifiers performances for the two subjects with the best (3 and 9) and the worst (2 and 5) performances. Differences between MI are more pronounced for subjects 3 and 9 compared to subjects 2 and 5, confirming results in figure 7.

Figure 8 brings also some interesting neurophysiological

informations about ERD/ERS patterns. During right hand MI, an ERD is elicited on the contralateral hemisphere (C3), confirming the literature. An ERD pattern is observed during left hand MI on the contralateral hemisphere (C4), except for EEG signal of subject 5 which elicited a desynchronization on the ipsilateral side (C3). During feet MI, no ERD is observed on EEG signal from Cz as expected, except for subject 9 who has a low desynchronization. In the other hand, for subject 2 a strong ERS is observed from the same signal, allowing a better discrimination between classes. It partly explains the good result (0.79) of classifier for feet MI compared to other classifiers. For subjects 3 and 5, a low ERS is observed on EEG signal from C3, showing no clear differences between feet MI and other motions in the latter case. For subject 3, the difference observed between feet MI and other MI is mainly due to the strong desynchronization elicited during right hand MI in the same signal (C3). During tongue MI an ERS is observed on EEG signals from C3 and C4 respectively for subjects 3 and 5, and 9, as mentioned by [15]. This synchronization is useful to select the most relevant time-frequency intervals. Nevertheless for subject 2, no ERD/ERS are observed during tongue MI in EEG signal filtered from each frequency bands. That is why the expert chose a standard time-frequency interval, as used in other studies [7].

## V. CONCLUSION

We described a method which allows an expert to select time-frequency intervals that he/she considers as the most relevant in order to discriminate different classes in the context of MI-based BCI. Our method is adapted to each subject and is performed off-line, using a set of EEG signals recorded during a cue-based paradigm. The human expert is able to select time-frequency intervals by reviewing a set of curves plotted for each MI. According to our results, CSP performances depend on the selected time-frequency intervals, confirming the literature. Our method is also interesting to analyse specific variation of brain activity elicited during MI, by gathering neurophysiological informations from EEG signals. Such informations are useful to better understand performances of each subject. Finally, a graphical user interface is developing to easily implement the proposed method by an expert.

## ACKNOWLEDGMENT

The authors would like to thank Fabien Lotte to provide his Matlab toolbox with CSP algorithms. They also thank the reviewers for their helpful comments and advices.

## REFERENCES

- [1] J.-J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.
- [2] T. O. Zander, C. Kothe, S. Welke, and R. M., "Enhancing human-machine systems with secondary input from passive brain-computer interfaces," in *4th International BCI Workshop & Training Course*, Graz, Austria, 2008.
- [3] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.
- [4] B. Benjamin, G. Dornhege, M. Krauledat, M. Schröder, J. Williamson, R. Murray-Smith, and K.-R. Müller, "The berlin brain-computer interface presents the novel mental typewriter hex-o-spell," 2006.
- [5] G. Blanchard and B. Blankertz, "Bci competition 2003-data set iia: spatial patterns of self-controlled brain rhythm modulations," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1062–1066, 2004.
- [6] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcspp) for brain-computer interface," in *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on*. IEEE, 2007, pp. 204–207.
- [7] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 2, pp. 355–362, 2011.
- [8] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcspp) in brain-computer interface," in *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. IEEE International Joint Conference on. IEEE, 2008, pp. 2390–2397.
- [9] H.-I. Suk and S.-W. Lee, "Subject and class specific frequency bands selection for multiclass motor imagery classification," *International Journal of Imaging Systems and Technology*, vol. 21, no. 2, pp. 123–130, 2011.
- [10] Y. Liu, H. Zhang, M. Chen, and L. Zhang, "A boosting-based spatial-spectral model for stroke patients' eeg analysis in rehabilitation training," 2015.
- [11] H. Raza, H. Cecotti, and G. Prasad, "Optimising frequency band selection with forward-addition and backward-elimination algorithms in EEG-based brain-computer interfaces," in *International Joint Conference on Neural Networks*, 2015, pp. 1–7.
- [12] D. J. McFarland, "The advantages of the surface Laplacian in brain-computer interface research," *International Journal of Psychophysiology*, vol. 93, no. 3, pp. 271–276, 2015.
- [13] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust eeg single-trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41–56, 2008.
- [14] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, pp. 136–142, 2008.
- [15] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.
- [16] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.