

GUIDER: A GUI FOR SEMIAUTOMATIC, PHYSIOLOGICALLY DRIVEN EEG FEATURE SELECTION FOR A REHABILITATION BCI

E. Colamarino^{1,2}, F. Pichiorri², F. Schettini², M. Martinoia¹, D. Mattia², F. Cincotti^{1,2}

¹ Department of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

² Neuroelectrical Imaging and BCI Lab, Fondazione Santa Lucia IRCCS, Rome, Italy

E-mail: emma.colamarino@uniroma1.it

ABSTRACT: GUIDER is a graphical user interface developed in MATLAB software environment to identify electroencephalography (EEG)-based brain computer interface (BCI) control features for a rehabilitation application (i.e. post-stroke motor imagery training). In this context, GUIDER aims to combine physiological and machine learning approaches. Indeed, GUIDER allows therapists to set parameters and constraints according to the rehabilitation principles (e.g. affected hemisphere, sensorimotor relevant frequencies) and foresees an automatic method to select the features among the defined subset. As a proof of concept, we compared offline performances between manual, just based on operator's expertise and experience, and GUIDER semiautomatic features selection on BCI data collected from stroke patients during BCI-supported motor imagery training. Preliminary results suggest that this semiautomatic approach could be successfully applied to support the human selection reducing operator dependent variability in view of future multi-centric clinical trials.

INTRODUCTION

Brain-computer interfaces (BCIs) collect the neurophysiological correlates of the brain activity (e.g. the electroencephalogram, EEG) and process them with the aim of controlling external devices, bypassing the neuromuscular system, or providing the user with a feedback of specific processes occurring in the brain [1]. A growing application field of this technology regards rehabilitation and more specifically the improvement of motor recovery in stroke patients [2]. In this context EEG-based BCIs monitor the modulation of brain activity induced by e.g. the imagination of movement. In fact, motor imagery (MI) practice elicits event-related desynchronization that occurs within EEG frequency bands (alpha and beta) and primarily over the scalp in sensorimotor cortical regions contralateral to the imagined part of the body. The introduction of BCI technology in assisting MI practice has been demonstrated to uncover the rehabilitative potential of MI, contributing to significantly better motor functional outcomes [3]. In order to reinforce a specific pattern, related to correct MI, appropriate choosing of BCI

control parameters (EEG features) is needed. In [3], BCI features, channels and frequencies, were identified according to a "manual" procedure (following EEG data analysis from the screening session). Namely, neurologists and/or therapists identify the features taking into account neurophysiological evidence and rehabilitation principles and basing on the visualization of matrices obtained from the statistical comparison between two conditions (task and rest). This procedure is highly dependent on the operator and is not suitable for the majority of therapists because it requires experience for visualizing patterns of desynchronization in that form.

With the aim to reduce the variability of this procedure in view of a wider employ of the BCI-based rehabilitation in stroke (e.g. for a multi-centric clinical trial), this study proposes the application of an algorithm to automatically choose EEG features.

As reported in [4], the genetic algorithm, the principal component analysis, the distinctive sensitive learning vector quantization and the sequential floating forward selection are the most common features selection methods used in BCI studies, especially the last one in sensorimotor rhythms-based BCI. In the same context, McFarland [5] proposed a stepwise multiple regression procedure to periodically update the features used to control cursor movement across training sessions. According to the stepwise algorithm, the feature that most reduces the residual variance (i.e., the variance not accounted for by target location) and does so with p-value less than 0.01 is added to the model. Additional features are then added in the same way. After each new addition, a backward stepwise regression removes any variables for which p-value is greater than 0.01. This process continues until no further features satisfy the addition/removal criteria.

As previously mentioned, the identification of "appropriate" control features, consistent with rehabilitation principles in terms of frequencies and areas, is a milestone in rehabilitation protocols supported by BCI technology. For this reason, this study proposes the application of the method in [5], as automatic method, suggesting an essential improvement for the rehabilitative field: the inclusion of the operator, neurologist and/or therapist and his neurophysiological

knowledge, in the features selection procedure. With this in mind and in view of a wider employ of the BCI-based rehabilitation in stroke, a user-friendly graphical interface was developed to guide the operator in the feature selection procedure and give him the possibility to define some constraints in which the automatic method has to run. In the overview [6] of publicly available software platforms for BCIs, the presented tool might match needs of rehabilitation BCI researchers orientated to a translational approach, from machine learning to physiology and vice-versa.

METHODS

Description and operating procedure: GUIDER is a graphical user interface (GUI) for semiautomatic and physiologically driven EEG features selection. It was designed and developed in MATLAB R2015a (The MathWorks, Inc., Natick, Massachusetts, USA) and runs until MATLAB R2011a. GUIDER allows users to interact with BCI data through a graphic interface without needing to use MATLAB syntax. Calling GUIDER in the MATLAB command window launches the first screen (Fig. 1) of the tool.

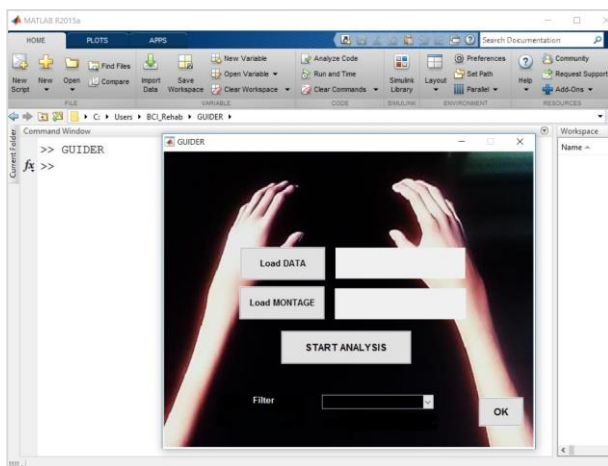


Figure 1: MATLAB main window and typical screen shot of GUIDER (right side).

The buttons “Load DATA” and “Load MONTAGE” allow importing BCI data (GUIDER supports several file formats) and montage files, respectively. More than one data file could be processed: files are concatenated. The “START ANALYSIS” button launches the following modules,

- (a) data conditioning module which applies spatial and frequency filtering,
- (b) feature extraction module which employs methods to estimate the signal spectrum,
- (c) statistical analysis module in which comparison tests of two conditions, i.e., two tasks, task vs rest, are implemented,
- (d) visualization module that gives output a matrix, channels and bins, where the statistical comparison index value of any feature is shown in a colour tint. A colour bar shows the colour range.

At the end of these analyses, the user visualizes pattern of desynchronization in the form of statistical index matrices, each one obtained using a different spatial filtering [7], e.g., common average reference (CAR), bipolar filters, surface Laplacian filters (other filtering options can be implemented). The pop-up menu in the bottom part of GUIDER main window allows the choice of filtering option (how many and which filters). At this point, the operator is required to define topographical and spectral constraints taking into account neurophysiological evidence and rehabilitation principles. This step may be guided by GUIDER or manual (Fig. 2). According to the first modality, the operator checks which hemisphere and channels to include in the analysis. In the manual modality, instead, operator selects rectangular areas in the statistical matrix after inserting the number of areas of interest in the edit-text box in GUIDER. The “OK” button closes all figures and opens the figures of statistical matrices, according to the number and the type of filtering earlier chosen, allowing the user to select rectangular areas accordingly (Fig. 3).

The values of features belonging to the areas selected manually or by the guided modality, are the inputs for the features selection algorithm: the stepwise regression (SW). This algorithm identifies an optimal subset of predictor variables (i.e. the features) and assigns weights to them in order to build an effective regression model to evaluate the relationship between the predictors and the dependent variable (here equivalent to subject’s intention). During each iteration, the algorithm adds or removes a feature from the classification model in order to obtain a combination of features ensuring a good classification performance. GUIDER implements it using the MATLAB ® function `stepwisefit.m`.

The optimal features and their weights are saved both as text file and external parameter file.

Proof-of-Concept: EEG dataset from three patients (subacute stroke patients with right-sided lesions, involved in previous studies at IRCCS Fondazione Santa Lucia, [3] for an extended description) were used to compare semiautomatic and manual procedure in terms of both features selection (channels and frequencies) and classification performance. Screening session’s data of each patient were analysed to identify the control features. Patients were instructed by the therapist to perform the movement (grasping and finger extension) imagination of their affected hand. During the initial screening session EEG signals were collected from 61 electrodes according to an extension of the 10–20 International System with 200 Hz as sampling frequency; scalp signals were referenced to the linked-ear signal. Each run consisted of 15 trials related to motor imagery task (grasping and finger extension) and 15 trials of baseline rest (9s each). Trials were randomly presented within a run. For each trial, we analysed the 4 epochs corresponding to the 4 seconds during which the patients performed the MI task. In the screening session, the subjects were not provided with any feedback of their brain activity.

To compare the procedures, the number of features automatically identified (in the constraints imposed by operator using GUIDER) is coherent with the number of features manually selected just by observing the statistical index matrix. These two control features drive the visual feedback to therapist and patient during the training sessions. In the training session the patients sat in a chair/wheelchair and their hands were covered by a white sheet on which a dedicated software projected a realistic visual representation of the patient's hands. The therapist instructed the patients to imagine the movement (grasping and finger extension) of the affected hand and they received a feedback when the trial was successful. The feedback consisted in the replication of the imagined movement by the virtual hand.

Data Analysis: The spectral analysis was performed on EEG data epochs (1s long) corresponding to MI task. The Maximum Entropy Method (16th order model) was employed to estimate amplitude spectrum with a resolution of 2 Hz and considering not overlapped epochs. All possible features in a reasonable range (i.e.,

0-36 Hz in 2 Hz bins) were extracted and analysed. A feature vector, spectral amplitude at each bin for each channel, was extracted from each epoch. For each feature a contrast was performed to assess statistically significant modulation induced on a specific feature. To this aim, the coefficient of determination R-square, i.e., the proportion of the total variance on the feature samples accounted for by target position, was computed for each feature (dependent variable).

Validation: In order to compare the classification performance achieved with both human and semiautomatic selection, Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve was assessed in auto-validation and cross-validation condition: screening data (previously exploited to obtain the control features and weights) and training data were used as testing dataset, respectively. Separately, outputs of the stepwise regression and the weights assigned by neurologists and/or therapists (conventionally, each one of -0.5) were the input of a linear classifier for the computation of the score used to calculate AUC values.

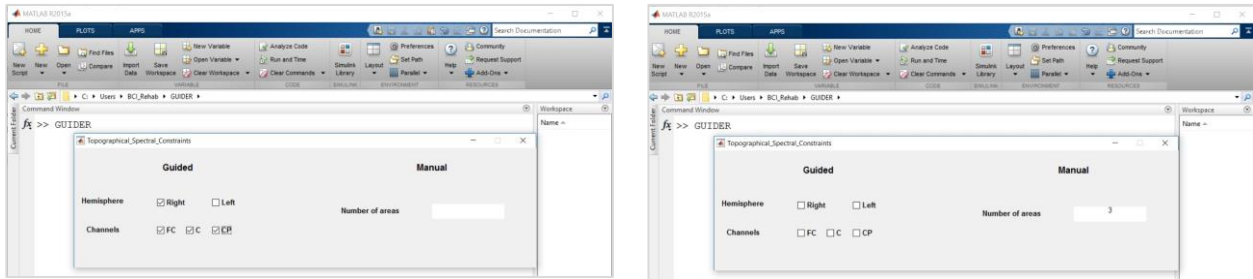


Figure 2: Screen shot of the window for the definition of topographical and spectral constraints. In the left side, guided procedure: the operator just checks the hemisphere and channels (in the sensorimotor area) to include in the analysis. In the right side, manual procedure: the operator just writes in the box the number of rectangular areas he would like to select in the statistical matrix that it will following open.

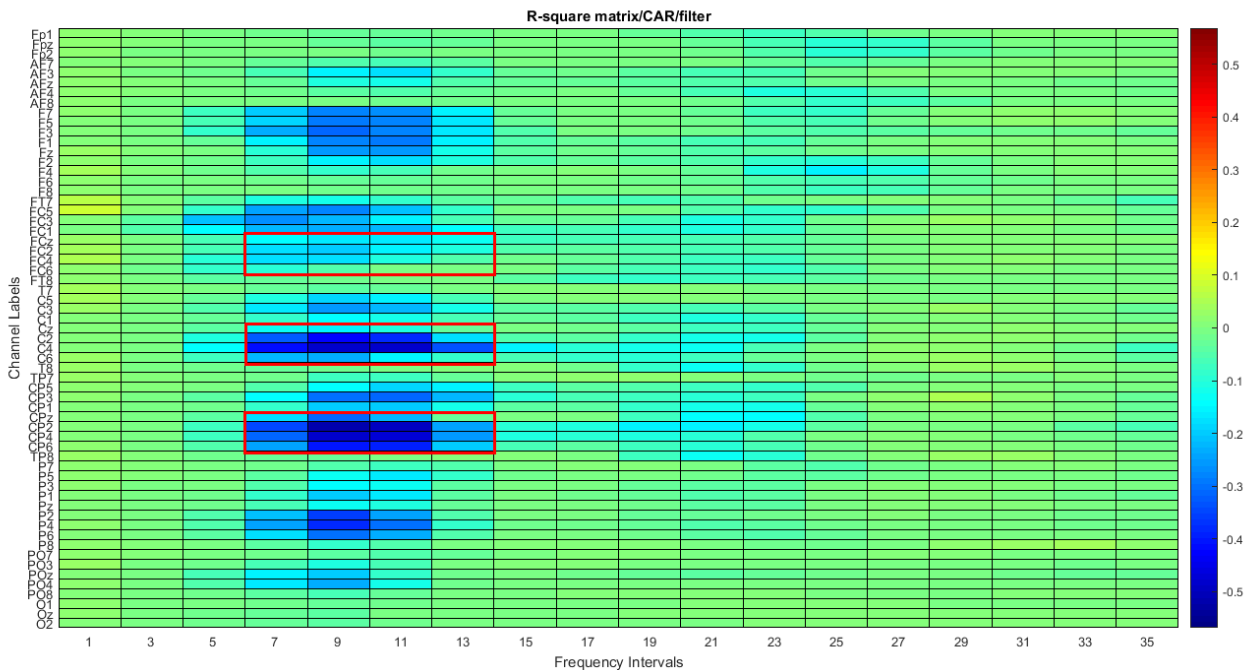


Figure 3: R-square matrix (channels and frequency intervals) obtained from EEG data analysis collected during the screening session and filtered using the filter chosen in GUIDER pop-up menu (e.g. CAR). The red rectangular areas (e.g. three areas as those written in Figure 2 right panel) are those selected by the operator (according neurophysiological evidence and rehabilitation principles) for the features selection using the GUIDER manual procedure.

RESULTS

Figure 4 shows the graphic output of the GUIDER operating procedure until topographical and spectral constraints definition. It displays for Subject 1 (S01) the R-square values of all features (61 channels and 18 frequency bins) after the filtering selection (in this case, CAR). The relevant control features selected, just based on R-square matrix visualization, by an expert neurophysiologist are reported in Table 1 for all three patients. The same operator, using the areas selection procedure in GUIDER, selected some rectangular areas: three for S01, from FCz to FC6, from Cz to C6, from CPz to CP6, all ranged from the fourth and to the seventh bin in the R-square matrix. Two optimal, coherent with the number of features selected just based on R-square matrix visualization, features identified by the stepwise algorithm in those areas are in Table 1.

The classification performance (AUC) in both auto-validation and cross-validation condition using the features identified in manual and semiautomatic (GUIDER) procedure are summarized in Table 2.

Table 1: Control features identified during BCI tasks at the screening session (for three subjects with right-sided lesions) by an expert neurophysiologist (manual procedure) and by the semiautomatic procedure implemented by GUIDER. For each feature EEG channel and frequency are reported (for each procedure) in the left and right columns, respectively.

Subject	Control feature	Manual procedure		GUIDER procedure	
S01	1	CP4	9 Hz	CP2	9 Hz
	2	C4	9 Hz	C4	11 Hz
S02	1	C2	21 Hz	C2	21 Hz
	2	CP2	23 Hz	Cz	25 Hz
S03	1	C4	19 Hz	C4	19 Hz
	2	CP4	19 Hz	CPz	21 Hz

Table 2: AUC values computed in auto-validation condition (namely, on data previously exploited to obtain the control features and weights) and in cross-validation condition (data from a rehabilitation training session used as testing dataset) for manual and semiautomatic procedure.

Subject	Validation (V)	Manual procedure	GUIDER procedure
S01	Auto-V	0.91	0.94
	Cross-V	0.88	0.88
S02	Auto-V	0.76	0.79
	Cross-V	0.74	0.74
S03	Auto-V	0.75	0.82
	Cross-V	0.70	0.65

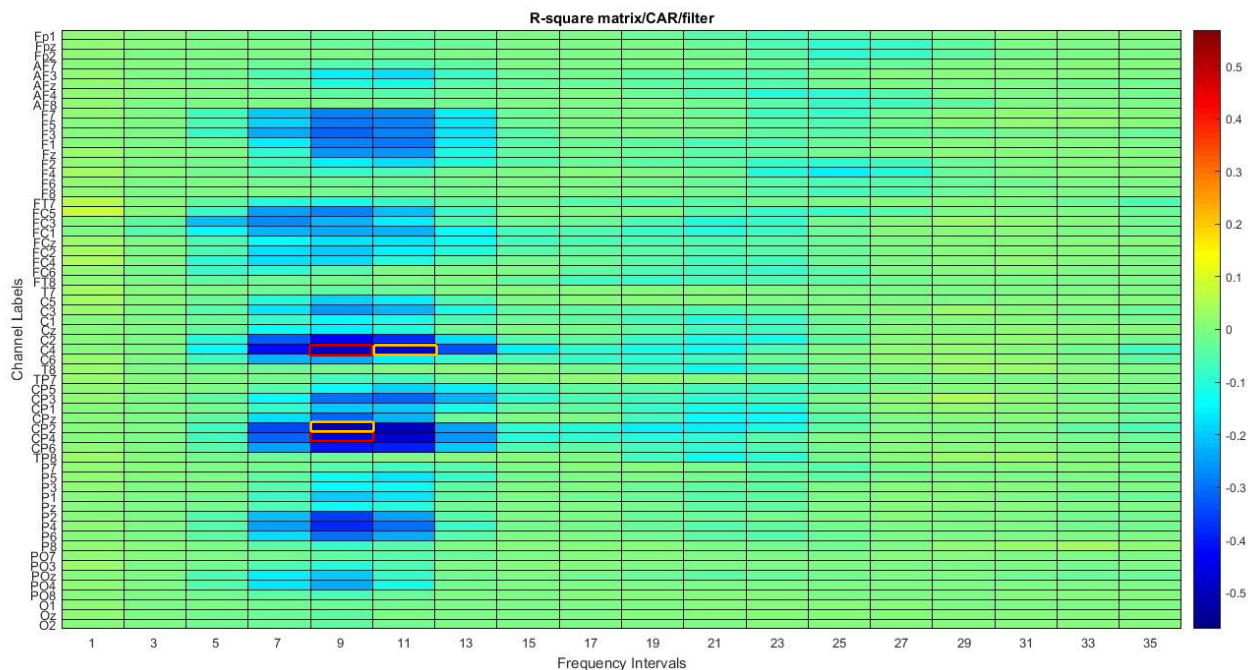


Figure 4: R-square matrix (channels and frequency intervals) obtained from EEG data analysis collected during the screening session from a subacute stroke patient with right-sided lesions (S01). The red (channels CP4 and C4 at 9 Hz) and yellow rectangles (channels CP2 and C4 at 9 Hz and 11 Hz, respectively) are features selected (for the rehabilitation training phase) by an expert neurologist and by GUIDER, respectively.

DISCUSSION

Identifying the optimal control features is a milestone in rehabilitation protocols supported by BCI technology. In contrast to other fields of application where optimal cursor control is pursued, in a rehabilitation context the aim is also to reinforce the appropriate sensorimotor activation in terms of both topographic and spectral characteristics. Therefore, the feature selection procedure requires knowledge coming of neurophysiology and rehabilitation principles as well as expertise in visualizing pattern of desynchronization in the form of statistical index matrices. The manual procedure is highly dependent on the operator and is currently restricted to researchers with experience in the BCI field. Therefore, the aim of GUIDER is twofold: first, to reduce the intra- and inter- operator variability of feature selection supporting the procedure with a semiautomatic method but without giving up to neurophysiological principles that characterize the rehabilitation; second, to facilitate this procedure for therapists without experience with BCIs.

GUIDER could be a (user-friendly) tool to support even non-expert operators in the reproducible identification of control features, since it considers both neurophysiological and machine learning approaches.

However, in view of a wider employ of GUIDER, several limitations must be addressed in the near future. First, although involved operators anecdotally considered GUIDER a user-friendly tool, the needs of the target group in terms of usability haven't been evaluated yet according to the user centered design approach. Second, the implementation in MATLAB environment, which is subjected to licensing issue, will be considered and overtaken as a next step.

The preliminary results suggest that the features identified by GUIDER are close to those chosen by experienced operators (manual procedure): e.g., CP4 vs CP2 at the same frequency and the same channel (C4) at neighboring bins (9 Hz and 11 Hz) for Subject 01. Furthermore, both procedure's outputs are congruent with the physiological evidences. Also in terms of classification performance, the procedures give indices (values of AUC) comparable in cross-validation condition and higher in the GUIDER application than in manual procedure in auto-validation. Hence, the choices of neurologists could be reproducible by a semiautomatic method that includes the operator and his neurophysiological knowledge in the procedure.

CONCLUSION

The introduction of GUIDER and its application in a BCI rehabilitation context suggest that it is feasible to support

the operators during the procedure of features selection with a user-friendly tool. GUIDER employs a semiautomatic method and takes into account neurophysiological evidence and rehabilitation principles. Performances are as good as manual selection, and GUIDER allows reproducibility of the procedure. The latter is a prerequisite for planning large multi-centric clinical trials, including a larger number of patients with several different operators, ensuring the comparability of BCI results among centers and thus increasing the generalizability of the results.

ACKNOWLEDGEMENTS

This work was funded in part by the Sapienza University of Rome - Progetti di Ateneo 2015 (C26A15N8LZ).

REFERENCES

- [1] Wolpaw J and Wolpaw EW. *Brain-Computer Interfaces: Principles and Practice*. Oxford, 2012.
- [2] Cincotti F, Pichiorri F, Arico P, Aloise F, Leotta F, de Vico Fallani F et al. EEG-Based Brain-Computer Interface to Support Post-Stroke Motor Rehabilitation of the Upper Limb. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2012, 4112–15*.
- [3] Pichiorri F, Morone G, Petti M, Toppi J, Pisotta I, Molinari M et al. Brain-computer interface boosts motor imagery practice during stroke recovery. *Annals of Neurology* 2015; 77 (5): 851–65.
- [4] Bashashati A, Fatourehchi M, Ward RK, Birch GE. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering* 2007; 4(2): R32-R57.
- [5] McFarland DJ, Sarnacki WA, Wolpaw JR. Electroencephalographic (EEG) control of three-dimensional movement. *Journal of Neural Engineering* 2010; 7(3).
- [6] Brunner C, Andreoni G, Bianchi L, Blankertz C, Breitwieser S, Kanoh S et al. *BCI Software Platforms*. In: *Towards practical Brain-Computer Interfaces*. Springer, 2013.
- [7] McFarland DJ, McCane LM, David SV and Wolpaw JR. Spatial Filter Selection for EEG-Based Communication. *Electroencephalography and Clinical Neurophysiology* 1997; 103 (3): 386–94.