



# Towards Moving Virtual Arms Using Brain-Computer Interface

Jaime Riascos<sup>(✉)</sup>, Steeven Villa<sup>(✉)</sup>, Anderson Maciel, Luciana Nedel, and Dante Barone

Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS),  
Porto Alegre, Brazil  
`{jarsalas,dsvsalazar,amaciel,nedel,barone}@inf.ufrgs.br`  
<http://inf.ufrgs.br/~jarsalas>, <http://inf.ufrgs.br/~dsvsalazar>,  
<http://inf.ufrgs.br/~amaciel>, <http://inf.ufrgs.br/~nedel>

**Abstract.** Motor imagery Brain-Computer Interface (MI-BCI) is a paradigm widely used for controlling external devices by imagining bodily movements. This technology has inspired researchers to use it in several applications such as robotic prostheses, games, and virtual reality (VR) scenarios. We study the inclusion of an imaginary third arm as a part of the control commands for BCI. To this end, we analyze a set of open-close hand tasks (including a third arm that comes out from the chest) performed in two VR scenarios: the classical BCI Graz, with arrows as feedback; and a first-person view of a human-like avatar performing the corresponding tasks. This study purpose is to explore the influence of both time window of the trials and the frequency bands on the accuracy of the classifiers. Accordingly, we used a Filter Bank Common Spatial Patterns (FBCSP) algorithm for several time windows (100, 200, 400, 600, 800, 1000 and 2000 ms) for extracting features and evaluating the classification accuracy. The offline classification results show that a third arm can be effectively used as a control command (accuracy > 0.62%). Likewise, the human-like avatar condition (67%) outperforms the Graz condition (63%) significantly, suggesting that the realistic scenario can reduce the abstractness of the third arm. This study, thus, motivates the further inclusion of non-embodied motor imagery task in BCI systems.

**Keywords:** Brain-Computer Interface · Virtual reality · Rubber hand illusion

## 1 Introduction

Along the years, researchers have sought different alternatives to allow human-machine communication. In this context, Brain-Computer Interface (BCI) plays

---

This study was partly funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and partly by CNPq.

a major role, motivated by overcoming the difficulties experienced by impaired people [9] or just by providing a non-mechanical user interface [5].

Motor Imagery BCI (MI-BCI) has been widely used and explored in active BCI [14]. MI-BCI employs sensorimotor rhythms (SMRs) that can be modulated voluntarily through the mental representation of physical motor actions. These patterns are called event-related de-synchronization (ERD/ERS) and have been successfully used for controlling different devices in BCI [5, 6, 9, 11].

As BCI is a relatively new research area, establishing metrics to objectively assess BCI systems (e.g., classification accuracy plus usability) is still a task to be done. Potential user identification is a mandatory step to design suitable BCI applications. However, this requires additional resources and developments of wearable recording equipment. Finally, the training and feedback should consider human factors and include the user inside the BCI loop through a more realistic, natural and intuitive training and feedback.

Skola and Liarnokapis [13] carried out a recent work using an embodied VR training for MI-BCI. A human-like avatar performs the motor actions in synchrony with the user’s actions. This neurofeedback-guided motor imagery training reports improvements in classification rates in comparison with the Graz paradigm. Even though it has not reached a significant difference, the authors report that ERD in VR subjects is stronger than the control group.

In that vein, this work studies an offline exploration of the classification of a third arm task in a BCI system. In order to explore as much as possible the differences in the classification of this task, several time windows were used in a Filter-Bank Common-Spatial Pattern (FBCSP) [1] for extracting features to train three classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA). Throughout such effort, we compared two training conditions: the traditional Graz paradigm and a realistic human-like feedback. In light of the results, we argue the feasibility of including the virtual third arm into a BCI system, and in line with the literature, the realistic training enhances the modulation of ERS/ERD patterns, and consequently, the performance of the user in motor imagery tasks.

## 2 Methods

### 2.1 EEG Signal Processing

**EEG Pre-processing.** The recorded data is imported and processed into EEGLAB (14.1) [4] (under Matlab 2017b). After down-sampling at 115 Hz, the signals are band-pass filtered at range 1–35 Hz using a finite impulse response (FIR) filter. Usually, a notch filter is used for line noise, but this method generally creates band-holes, and distortions close the cut-off frequency. Therefore, the Cleanline plugin, which uses multi-taper regression for removing sinusoidal artifacts, is used at 50–115 Hz instead. Likewise, Cleanraw plugin is set-up for rejecting bad channel. The rejected channels are then interpolated using a spherical function. Finally, EEG signals are re-referenced using common average referenced (CAR).

This study aims to explore as much as possible the influence of both time window of the trials and the frequency bands on the accuracy of the classifiers. Therefore, we used a Filter Bank Common Spatial Patterns (FBCSP) algorithm [1] for several time windows (100, 200, 400, 600, 800, 1000 and 2000 ms) for extracting features and evaluating the classification accuracy. The bank with the MI frequency bands is comprised of  $\mu$  (8–12 Hz), low  $\beta$  (12–16 Hz), middle  $\beta$  (16–24 Hz), high  $\beta$  (24–30 Hz), and whole  $\beta$  (12–30 Hz) bands. The reason of splitting the  $\beta$  band into sub-bands is for getting enough variables for the FBCSP algorithm to work.

**Feature Extraction.** The ERS/ERD patterns are predominant in  $\mu$  (8–12 Hz), and  $\beta$  (13–30 Hz) rhythms and in its onset goes from 500 ms up to three seconds after the movement execution [12]. Inspired by these facts, we built a framework to obtain the best combination of window size, frequency band, and classifier for each user. For obtaining a pool of possible combinations, we ran the Filter Bank Common Spatial Pattern (FBCSP) algorithm [1] in seven time-window sizes of the signal (100, 200, 400, 600, 800, 1000 and 2000 ms)) and five frequency bands (8–12 Hz, 12–16 Hz, 16–24 Hz, 24–30 Hz, 12–30 Hz).

The method employs a greedy algorithm to heuristically find the best combination based on the classification error rates of all possibilities. We focused on the variability that exists across the users in their performance. So, with this approach, we were able to create a suitable and user-centered BCI classification.

The FBCSP approach has demonstrated successful performance in BCI applications [1]. This method extracts the most relevant spectral and spatial features using a CSP filter for each frequency band. CSP is one of the most known and widely used methods for extracting features in a two classes BCI application [3, 8]. CSP computes the project matrix  $\mathbf{W} \in \mathbb{R}^{c \times c}$  that linearly transforms the band-pass filtered data  $\mathbf{E} \in \mathbb{R}^{c \times t}$  into a spatial filtered signal  $\mathbf{Z} \in \mathbb{R}^{c \times t}$  (with  $c$  being the number of channels and  $t$  the EEG samples per channel) as follows:

$$\mathbf{Z} = \mathbf{W}^T \mathbf{E} \quad (1)$$

Thus, the power of  $\mathbf{Z}$  effectively discriminates two mental states (classes), maximizing the variance under one condition meanwhile it is minimizing for the other [3]. In order to get the most discriminative patterns, the first and last  $m$  ( $m=3$ ) columns of  $\mathbf{W}$  were used to create the spatial-filtered signal  $\mathbf{Z}$ . The  $m$ -dimensional feature vector is then formed from the logarithm of the normalized variance of  $\mathbf{Z}$ :

$$v_i = \log(\text{var}(Z_i)), \quad i = 1, 2, \dots, 2m. \quad (2)$$

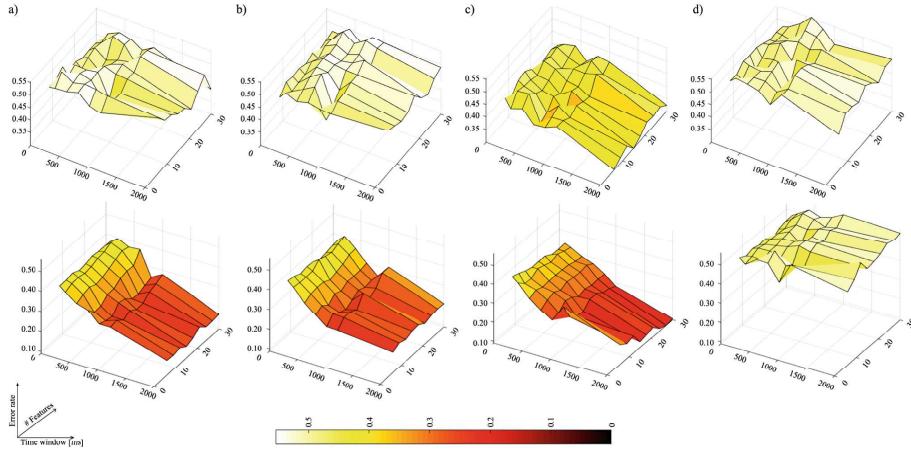
This results in 30 features (six CSP filters per each frequency bands) for each EEG trial in the specific window. From these features, the maximum Relevance Minimum Redundancy (mRMR) feature selection algorithm is used to extract the most relevant features [10]. This algorithm minimizes the redundancy meanwhile maximizes the relevance of the features using mutual information. As the size of  $v_i$  depends on the frequency bands, the number of selected features was

progressively increased (step = 3) from the minimum amount (6) up to the total size of  $v_i$ . Hence, the selected features for each time window is used to separately train three common BCI classifiers [7]: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Linear Discriminant Analysis (LDA).

## 2.2 Classification

We used two of the most popular linear approaches (SVM and LDA) and a non-parametric method (KNN). These methods were trained to classify independently four binary imaginary tasks: Third and Left hand (TH\RH); Third and Right hand (TH\RH); Third hand and Resting State (TH\RS); and Left and Right hand (LH\RH). The reader can notice that the real movements are not included in the classification. The intention of including the real movements in the experiment was to reduce the abstractness of the three imaginative tasks and have a fresh mental representation of the action.

The miss-classification error was computed using the usual k-fold cross-validation approach. This method randomly divides the data into  $k$  equal size partitions and uses  $k-1$  sets to train the model and one set to validate it. In this study, we used ten times the 10-fold cross-validation. Finally, the above classifiers were implemented using the Statistics and Machine Learning Toolbox of Matlab. Both SVM and LDA used the default parameters (linear kernel and standardize predictor data for SVM and LDA without hyperparameter optimization). KNN used a Euclidean distance with  $k = 5$ .



**Fig. 1.** Error rates over number of features and time window size for all users. Top: Graz condition. Bottom: Hands condition. The four binary classification are represented by (a) TH/LH; (b) TH/RH; (c) TH/RS; and (d) LH/RH.

### 3 Results

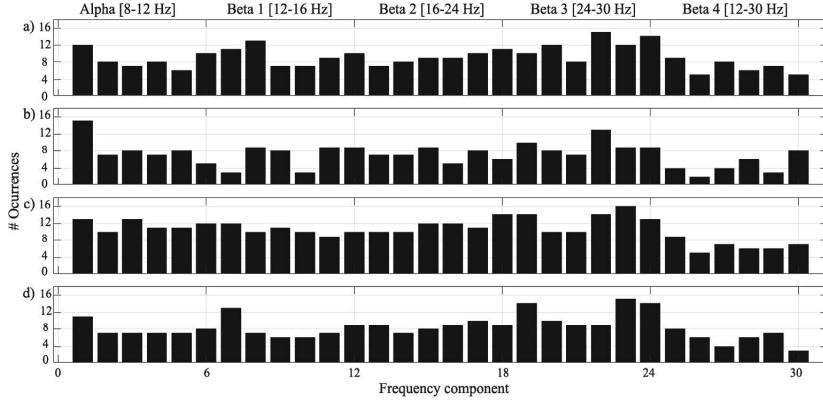
Figure 1 shows the error rates in function of the number of features and the size of the time windows for the KNN in each binary classification. We obtained these surfaces for both Graz (top) and Hands (bottom) conditions as well as for each classifier (SVM, KNN, LDA). Initially, we can see that Hands condition reached lower error rates than Graz for KNN. Likewise, the error rates are better in the classification that includes the third arm ((a), (b), (c) in the Fig. 1) than the left and right classification ((d) in the Fig. 1); in effect, the RH/LH classification reached error rates close to 0.5, which means a classification by chance. Finally, the tendency in both conditions is that as the size of the time window increases the error rates decrease but in the number of features the tendency is unclear.

Following the error mean values of each window\feature, we built a greedy algorithm to find the optimal global choice. Thus, we can obtain for each subject, the best combination of the three conditions (i.e. number of features (NF), size of the window (SW) and classifier (C)) to get the best miss classification rate (error). The Table 1 shows these values in Graz (G) and Hands (H) conditions among the subjects.

**Table 1.** The best combinations of number of features, time window and classifier and the error rates reached with them across the subjects for both conditions. The asterisks indicate the subjects that began the experiment with the Hands condition.

Subject	Condition	TH-RH				TH-LH				TH-RS				LH-RH			
		NF	SW	C	Error	NF	SW	C	Error	NF	SW	C	Error	NF	SW	C	Error
1*	G	6	2000	KNN	<b>0.18</b>	6	2000	KNN	<b>0.18</b>	6	800	LDA	<b>0.24</b>	9	2000	LDA	0.42
	H	9	2000	LDA	<b>0.17</b>	12	2000	KNN	0.28	15	1000	SVM	0.27	12	2000	SVM	0.41
2	G	6	800	SVM	0.26	24	100	KNN	0.40	9	2000	SVM	<b>0.16</b>	6	2000	KNN	0.30
	H	6	2000	KNN	<b>0.19</b>	6	2000	LDA	<b>0.13</b>	24	2000	SVM	<b>0.24</b>	21	800	KNN	0.39
3	G	6	800	LDA	0.42	6	800	KNN	0.41	12	2000	KNN	0.35	21	600	KNN	0.45
	H	12	1000	SVM	0.40	9	2000	KNN	0.34	12	2000	KNN	0.27	6	100	KNN	0.48
4*	G	30	100	KNN	0.42	6	800	LDA	0.41	24	1000	KNN	0.37	6	100	SVM	0.46
	H	15	100	KNN	0.47	6	400	LDA	0.45	18	600	KNN	0.37	6	1000	SVM	0.36
5	G	30	800	SVM	0.31	6	100	LDA	0.46	12	600	KNN	0.28	15	100	KNN	0.45
	H	12	2000	KNN	0.34	24	800	SVM	0.41	15	400	KNN	0.42	12	2000	KNN	0.38
6*	G	21	2000	KNN	0.40	6	100	LDA	0.42	18	2000	SVM	0.26	6	800	KNN	0.45
	H	18	2000	SVM	<b>0.10</b>	6	2000	SVM	<b>0.16</b>	30	2000	KNN	<b>0.09</b>	21	400	SVM	0.42
7	G	15	2000	SVM	0.38	18	800	SVM	0.38	21	400	SVM	0.37	21	2000	KNN	0.41
	H	21	200	LDA	0.28	30	1000	KNN	0.38	9	1000	SVM	0.34	6	800	SVM	0.44
8*	G	6	800	LDA	0.36	9	800	SVM	0.40	18	2000	SVM	0.31	6	200	SVM	0.46
	H	9	800	SVM	0.39	12	800	LDA	0.41	6	800	LDA	0.35	21	800	LDA	0.37
9	G	12	800	KNN	0.38	6	100	SVM	0.48	27	2000	KNN	0.34	15	600	LDA	0.40
	H	21	1000	KNN	0.31	9	1000	SVM	<b>0.22</b>	27	800	KNN	0.37	9	800	LDA	0.37
10*	G	12	800	KNN	0.32	9	800	LDA	0.34	6	2000	KNN	<b>0.22</b>	6	100	SVM	0.46
	H	6	800	SVM	<b>0.25</b>	6	2000	KNN	<b>0.21</b>	9	2000	LDA	<b>0.22</b>	6	2000	SVM	0.36

We merged the data of each run into a single dataset for thus training the classifiers. The table shows that there are several variations across the subjects



**Fig. 2.** Relevant frequency components over all subjects. (a) TH\LH; (b) TH\RH; (c) TH\RS; d) LH\RH.

and conditions. Only in TH/RS task, the number of features was widespread. Otherwise, the distribution was concentrated from six up to 21, precisely, the Fig. 2 shows the histogram of selection of each frequency component over all subjects for the four binary classifications. In the figure can be seen that frequency components in the  $\mu$  and  $\beta_3$  were selected for most of the participants, i.e., the clear peaks are inside these ranges. Meanwhile, the whole  $\beta$  band ( $\beta_4$ ) was unused, showing how the  $\beta$  sub-bands can be used more than the whole band itself. In the other side for the size of the time window, the classification that includes the third arm (TH-RH, TH-LH and TH-BS) was superior of the 800 ms up to 2 s (maximum time). Finally, the KNN was the most used among the participants and conditions.

The mean error obtained over all participants ( $n = 10$ ) in Hands condition was approx 32% (i.e., TH-LH:  $0.29 \pm 0.03$ , TH-RH:  $0.29 \pm 0.03$ , TH-BS:  $0.29 \pm 0.03$ , and LH-RH:  $0.39 \pm 0.01$  and around 36% in Graz condition (i.e., TH-LH:  $0.34 \pm 0.02$ , TH-RH:  $0.38 \pm 0.02$ , TH-BS:  $0.29 \pm 0.02$ , and LH-RH:  $0.42 \pm 0.01$ ). Pairwise comparison using paired Wilcoxon signed rank test with Bonferroni correction reveals a significant ( $V = 0$ ,  $p$ -value = 0.048) between conditions. Likewise, Dunn's Kruskal-Wallis Multiple Comparisons with Bonferroni correction show significant difference among groups, exactly in both TH-LH ( $z = -2.49$ ,  $p$ -value = 0.03) and TH-BS ( $z = -3.92$ ,  $p$ -value = 0.0003) against LH-RH for Graz, and in Hands only for TH-BS ( $z = -2.44$ ,  $p$ -value = 0.04).

In summary, the Hands condition significantly outperforms (0.32) the Graz (0.36) in the classification. Intriguingly, the classification of TH-LH was better than the other two motor imagery conditions (TH-RH, LH-RH) in both conditions (Graz = 0.34, Hands = 0.29).

#### 4 Conclusion

This study investigated the possibility of using an imaginary third arm and the differences of the EEG patterns and classification rates of using a realistic

visual feedback. The benefits presented by this feedback are reflected in the enhanced of the ERS signals that consequently produces an improvement of the classification. Supernumerary MI-BCI systems are prominent and possible uses should be explored, especially for VR applications, where customized avatars could be controlled using imaginary non-body signals.

The high error rates reached by the left-right hand classification (LH/RH) in both conditions (Hands:  $0.39 \pm 0.01$ , Graz:  $0.42 \pm 0.01$ ) could suggest that the inclusion of the third arm would cause the reduction of its accuracy because the users could interpret either left and right hand as the third arm. Whereas, the third arm is distinguished from the left hand than the right with better accuracies. It could support the fact that the TH task follows the activity based on the handedness. Unfortunately, all of our subjects were right-handed, so we can not evaluate the handedness thoroughly in this experiment.

Unfortunately, this work lacks in studies about ownership of the third-arm in both subjectively, with questions about the sense of agency and sense of ownership, and objectively, using galvanic skin response (GSR), following the work of Bashford and Mehring [2]. These data could give some insights regarding the use of supernumerary BCI and how it could be used in real application, coming from the answers of the users. Also, it would be necessary to perform the experiment with left-handed people, in order to study the handiness of the third arm.

## References

1. Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 2390–2397 (2008). <https://doi.org/10.1109/IJCNN.2008.4634130>
2. Bashford, L., Mehring, C.: Ownership and agency of an independent supernumerary hand induced by an imitation brain-computer interface. PLoS One **11**(6), 1–15 (2016). <https://doi.org/10.1371/journal.pone.0156591>
3. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.: Optimizing spatial filters for robust EEG single-trial analysis. IEEE Sig. Process. Mag. **25**(1), 41–56 (2008). <https://doi.org/10.1109/MSP.2008.4408441>
4. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**(1), 9–21 (2004). <https://doi.org/10.1016/j.jneumeth.2003.10.009>
5. Gert, P., Leeb, R., Faller, J., Neuper, C.: Brain-computer interface systems used for virtual reality control. In: Kim, J.J. (ed.) Virtual Reality, Chap. 7, pp. 1–19. InTech (2011)
6. He, B., Baxter, B., Edelman, B.J., Cline, C.C., Ye, W.W.: Noninvasive brain-computer interfaces based on sensorimotor rhythms. Proc. IEEE **103**(6), 907–925 (2015). <https://doi.org/10.1109/JPROC.2015.2407272>
7. Lotte, F., et al.: A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. J. Neural Eng. **15**(3), 031005 (2018)

8. Lotte, F.: A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In: Miranda, E.R., Castet, J. (eds.) Guide to Brain-Computer Music Interfacing, pp. 133–161. Springer, London (2014). [https://doi.org/10.1007/978-1-4471-6584-2\\_7](https://doi.org/10.1007/978-1-4471-6584-2_7). <https://hal.inria.fr/hal-01055103>
9. Neuper, C., Muller, G., Kubler, A., Birbaumer, N., Pfurtscheller, G.: Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment. *Clin. Neurophysiol.* **114**(3), 399–409 (2003). [https://doi.org/10.1016/S1388-2457\(02\)00387-5](https://doi.org/10.1016/S1388-2457(02)00387-5)
10. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1226–1238 (2005). <https://doi.org/10.1109/TPAMI.2005.159>
11. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. *Proc. IEEE* **89**(7), 1123–1134 (2001). <https://doi.org/10.1109/5.939829>
12. Pfurtscheller, G.: Quantification of ERD and ERS in the Time Domain, pp. 89–105, 6th edn. Elsevier B.V., Netherlands (1999). Revised edition
13. Skola, F., Liarokapis, F.: Embodied vr environment facilitates motor-imagery brain-computer interface training. *Comput. Graph.* **75**, 59–71 (2018). <https://doi.org/10.1016/j.cag.2018.05.024>
14. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**(6), 767–791 (2002). [https://doi.org/10.1016/S1388-2457\(02\)00057-3](https://doi.org/10.1016/S1388-2457(02)00057-3)