

Pre-Experiment Resting Baseline Improves Motor Imagery BCI Decoding in a Robotic Hand Rehabilitation Context

Aniana Cruz¹, Gabriel Pires^{1, 2}

¹ Institute for Systems and Robotics, University of Coimbra, Coimbra, Portugal

² Engineering Department of the Polytechnic Institute of Tomar, Tomar, Portugal.

E-mail: anianabrito@isr.uc.pt

ABSTRACT: Motor imagery (MI)-based brain-computer interfaces (BCIs) rely on event-related desynchronization and synchronization (ERD/ERS), which are defined relative to a resting-state baseline. Therefore, baseline selection directly affects both the interpretation of the ERD/ERS patterns and MI classification performance. In this work, we systematically evaluate the impact of baseline definition and temporal window selection on ERD/ERS patterns and classification performance in a post-stroke rehabilitation protocol involving a robotic hand exoskeleton, considering both calibration and on-line sessions. Three resting-state baselines were compared: a long rest recorded before the experiment, a long rest recorded after task execution, and a short inter-trial rest commonly used in MI-BCI paradigms. The results show that the pre-experiment resting baseline yielded the strongest and most stable ERD patterns, as well as the highest classification performance in both calibration and online data. Rest recorded after task execution resulted in attenuated ERD and slight performance degradation. The short inter-trial rest led to substantial variability and performance loss.

INTRODUCTION

Motor imagery (MI), the mental simulation of movement, without its physical execution, has been increasingly used as a control signal in brain-computer interfaces (BCIs) across applications ranging from clinical uses, such as post-stroke rehabilitation [1], to non-clinical applications such as gaming [2]. Because MI activates the same neural networks involved in movement execution [3], MI-based BCIs exploit modulations of sensorimotor rhythms (SMRs) over motor cortical areas. During MI, a decrease in the power relative to the reference baseline is typically observed, referred to as event-related desynchronization (ERD), followed by an increase in power known as event-related synchronization (ERS) [4]. These modulations predominantly occur in the mu (8-13 Hz) and beta (14-30 Hz) frequency bands. In MI-BCI paradigms, the baseline generally corresponds to a relaxed resting-state condition, during which SMR power is maximized, thereby enhancing the discrimination between rest and MI states. Since ERD and ERS are defined as relative changes in band power with respect to the baseline, their estimation

is inherently dependent on the choice of baseline.

Several studies have shown that baseline selection plays a critical role in ERD/ERS computation and interpretation [5–7]. In [5], the authors investigated the impact of baseline selection on maximizing ERD by comparing static and dynamic baselines. Their results showed that subject-specific baseline selection could enhance mu-band power. Study [6] explores how two baseline conditions (observation of a static image and a period of stillness) affect mu-ERD/ERS modulation in toddlers during action-observation and action-execution tasks. The results showed that baseline effectiveness varied across subjects, although static-image observation generally produced stronger mu-ERD. In [7], the effects of four different baseline time-window selections were investigated with respect to both ERD/ERS modulation and BCI performance. The results indicated that ERD may be underestimated when the baseline is close to the upcoming trial, as it may include anticipatory neural activity. Contrarily, ERD can be overestimated when the baseline is selected near the end of the previous trial. Previous studies have analyzed baselines obtained in different conditions (e.g., different tasks, some artifacts), which may induce distinct mental states. In most MI-BCI paradigms, the baseline is defined as a short rest period between consecutive MI trials [7]. Because trials typically follow each other closely, such baseline windows may be contaminated by cue-related activity and may include subject concentration and anticipation, thereby distorting the estimated ERD. Incorrect ERD computation can negatively affect MI decoding performance, particularly in stroke patients, where typical contralateral lateralization of sensorimotor activation is frequently disrupted [8].

In this study, we consider a post-stroke rehabilitation scenario in which a BCI is used to control a robotic hand exoskeleton. To optimize this protocol, we systematically compare the impact of three different baselines on the performance of MI classification, namely: 1) a long clean resting state recorded before the beginning of the experiment; 2) a long resting period recorded after task execution; and 3) a short rest period recorded between MI trials.

MATERIALS AND METHODS

Participants: Ten healthy participants (6 female, mean age 26.6 years (std 4.9)) participated in the study. None of them had previous experience with MI-based BCI. Each participant signed an informed consent explaining the purpose of the study, the procedures, the possible risks, and the authorization for the publication of the results.

EEG data acquisition: The EEG signal was recorded using a g.USBamp bioamplifier (g.tec medical engineering GmbH, Austria) from electrodes FC3, FCz, FC4, C3, Cz, C4, CP3, CP1, CPz, CP2, CP4, and Pz according to the international extended 10-20 standard system. The reference electrode was placed at the right earlobe, and the ground was located at FPz. The EEG signals were pre-processed using a notch filter at 50 Hz and a band-pass filter between 8–30 Hz and sampled at 256 Hz.

Experimental Protocol: The experiment consisted of three sessions (Fig. 1):

- **Calibration Session 1:** conducted to collect EEG data for model calibration. This session included a long resting-state recording before task execution, referred to as long baseline 1 (LB1), representing a clean and true resting-state condition. In addition, short baseline periods (SB) were recorded between MI tasks during the trials, representing task-dependent resting states that may include workload and fatigue effects.
- **Calibration Session 2:** used to increase the amount of training data and improve classifier robustness. At the beginning of this session, a second long resting-state was acquired, referred to as long baseline 2 (LB2), representing a resting-state potentially affected by cognitive fatigue after task execution. Similarly to Calibration Session 1, baseline SB was also recorded.
- **Online Session:** performed to evaluate the trained classifier in real time during MI-based robotic rehabilitation, including both visual feedback and somatosensory feedback delivered through actuation of the robotic hand exoskeleton.

The rehabilitation protocol included a robotic hand performing hand opening and hand closing movements. Accordingly, during the calibration sessions, participants were instructed to perform MI of the same movements with their right hand so that the imagined actions were aligned with those executed by the robotic hand. The calibration sessions started with a long resting-state recording lasting 90 s, during which participants were instructed to relax and remain at rest. This was followed by repeated trials (20 times), including relax, rest, and motor imagery periods. Each trial began with a 2-s relax period, during which participants were allowed to rest and be free. This was followed by a 6 s rest period, during which participants were asked to rest and minimize artifacts such as movement and blinking. Subsequently, a 6 s motor

imagery period was presented, during which participants performed motor imagery of the predefined movement (hand opening or closing, according to the visual cue). Participants were instructed to start performing the corresponding MI task as soon as the cue appeared and to continue until the cue disappeared. The total duration of each trial was approximately 16 s. Following the calibration sessions, an online session was conducted, consisting of relax and motor imagery periods. During the MI periods, classification was performed in real time using a model trained on calibration data from rest SB and MI periods. The classifier was applied using a sliding-window approach during the motor imagery period. Windows of 1.5 s length with a 1 s overlap were extracted, denoted as W^i , where $i=1,2,\dots,9$, with onset times shifted relative to the motor imagery onset (0.5 s, 1.0 s, 1.5 s, etc., as denoted in Fig. 1). At the end of each window, the system provided visual feedback: a green bar increased if MI was correctly detected. Simultaneously, somatosensory feedback was delivered through robotic hand actuation (trigger-based control). If at least one classification window was correctly identified, the robotic hand, called Hand Of Hope [9], executed the intended movement (opening or closing hand) (Fig. 2).

Feature Extraction and Classification: The 12 EEG channels were segmented into 1.5-second epochs beginning 0.5 seconds after the onset of each cue. All epochs were first band-pass filtered between 8 and 30 Hz using a Butterworth filter. Then, common spatial patterns (CSP) [10] with two spatial filters ($m = 2$) were applied for feature extraction. The variance (var_t) of each spatial filter projection was computed, followed by normalization and logarithmic transformation ($\log_{10} \left(\frac{\text{var}_t}{\sum_{t=1}^m \text{var}_t} \right)$). Finally, the resulting feature vectors were classified using a Linear Discriminant Analysis (LDA) classifier. Although two different MI tasks were included to increase participant engagement and reduce task monotony, the classifier performed binary classification between MI and rest, since previous studies have shown that this approach is more suitable for dealing with the weak contralateral activation commonly observed in stroke patients [11]. Consequently, the hand opening and hand closing conditions were combined into a single MI class. The training dataset consisted of 40 MI epochs and 40 rest epochs.

Performance Metrics - online data: In online operation, the classifier was evaluated only during MI tasks. Therefore, performance was assessed using metrics based on the true positive rate (TPR). Three metrics were considered: i) the sliding-window TPR (Win_{TPR}^i), which evaluates classification performance for each window W^i ; ii) the mean sliding-window TPR (Win_{TPR}), which is the average performance across all windows; and iii) the trial-level detection rate (Trial_{DR}), which evaluates whether at least one window was correctly classified within a trial. These metrics are defined as follows:

- (i) Win_{TPR}^i , the percentage of correctly classified W^i across all trials:

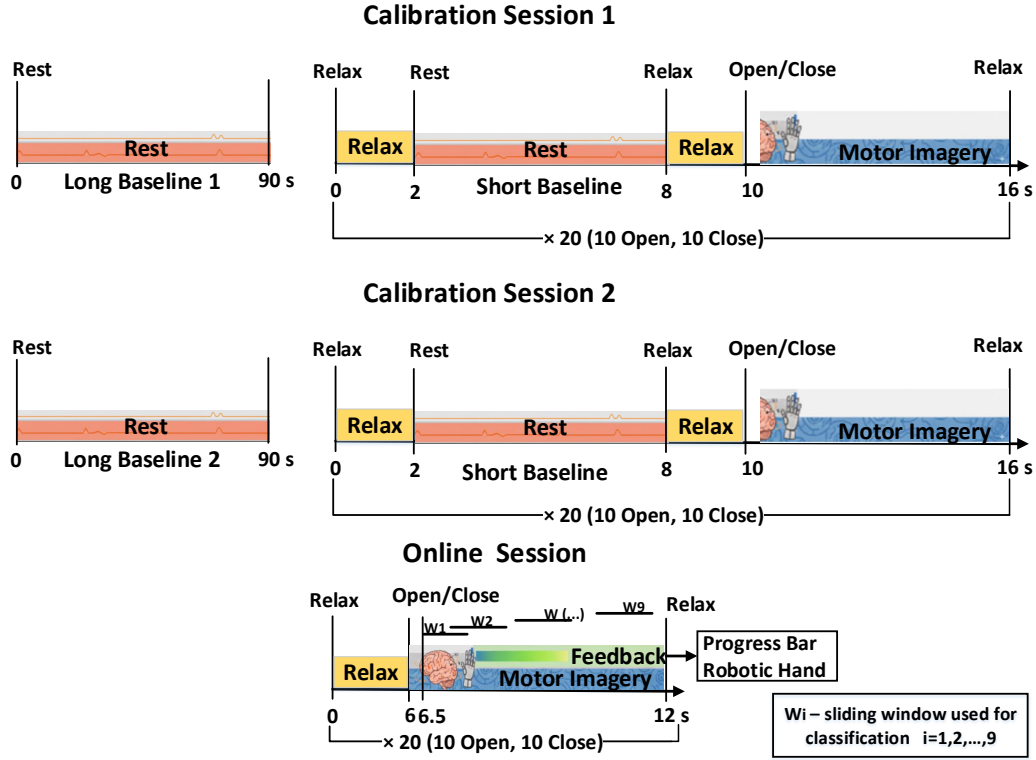


Figure 1: Schematic representation of the rehabilitation protocol, including the calibration sessions, and the online session. Two types of feedback were provided: a progressing bar (visual feedback) and robotic hand actuation (somatosensory feedback). Three baseline conditions were defined: long baseline 1 (LB1), recorded at the beginning of the experiment; long baseline 2 (LB2), recorded after Calibration Session 1; and short baseline (SB), recorded between MI trials.



Figure 2: Experimental setup including the EEG acquisition system, the visual interface, and the robotic hand (Hand Of Hope).

- (iii) $Trial_{DR}$, percentage of trials in which at least one window was correctly classified:

$$Trial_{DR} = \frac{1}{T} \sum_{t=1}^T \gamma_t \times 100 \quad (3)$$

where

$$\gamma_t = \begin{cases} 1, & \text{if at least one } W^i \text{ in trial } t \text{ is correct} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Performance Metrics - calibration data: The performance was computed using the sliding-window accuracy (Win_{Acc}^i), defined as:

$$Win_{Acc}^i = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where positives correspond to MI windows and negatives correspond to rest windows. Thus, whereas online performance is assessed only during MI periods, here classifier performance is assessed using both MI and rest windows.

Neurophysiological analysis: The effect of baseline selection on ERD and ERS was evaluated offline, taking data recorded in the 3 sessions. Event-related spectral perturbations (ERSP) were computed using EEGLAB in the 8–30 Hz frequency band from the CSP projections. To reduce inter-session variability, CSP projections were

$$Win_{TPR}^i = \frac{1}{T} \sum_{j=1}^T \delta_j^i \times 100, \quad \delta_j^i = \begin{cases} 1, & \text{if } \hat{y}_j^i = y_j^i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where T corresponds to the total number of trials per window i (W^i), $\hat{y}_j^i$ and y_j^i denote the predicted and true labels of the W^i , respectively.

- (ii) Win_{TPR} computed as:

$$Win_{TPR} = \frac{1}{N} \sum_{i=1}^N Win_{Acc}^i \quad (2)$$

where $N = 9$ is the total number of windows.

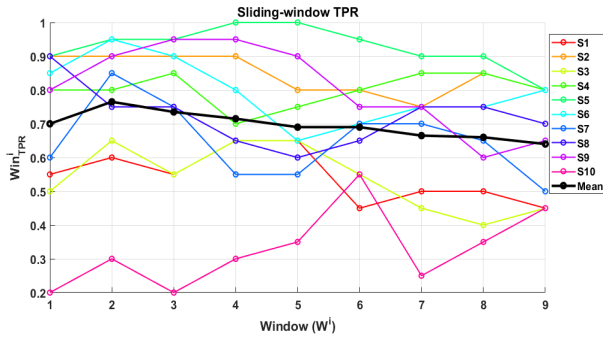


Figure 3: The sliding-window TPR (eq. 1) achieved during the online operation.

normalized by the standard deviation computed across all projected samples within each session. ERSP analysis focused on the minimum ERSP value across the four CSP projections. These values were computed separately for each baseline/MI window combination.

The spectral characteristics of each baseline condition were examined using the power spectral density (PSD), estimated with the Welch method. The relative band power was computed for the theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz) bands to quantify the contribution of each frequency band to the overall EEG activity, as follows:

$$RP_{band} = \frac{\sum_{f=f_1}^{f_2} PSD(f)}{\sum_{f=f_{min}}^{f_{max}} PSD(f)} \quad (6)$$

where $PSD(f)$ represents the power spectral density at frequency f , f_1 and f_2 denote the lower and upper frequency limits of the band of interest, and $f_{min} = 4$ and $f_{max} = 30$ define the total frequency range. The two motor cortex channels, C3 and C4, were analyzed since they are commonly used in MI-based BCI systems.

RESULTS

Online performance: Table 1 presents the mean TPR obtained during online operation. The results show that in most of the trials, participants achieved at least one correctly classified window, which successfully triggered the robotic hand. The $Trial_{DR}$ ranged from 85% to 100% across participants. For Win_{TPR} , performance was on average approximately 70%. The TPR obtained in each window is presented in Fig. 3 for each subject. As observed, participant S10 had difficulty performing MI, while participants S1 and S3 were able to sustain MI only for short periods. Some participants showed poorer performance in the later windows, possibly because the robotic hand executed the movement after the first correctly classified window, which may have reduced their motivation to continue the task.

Offline classification performance and ERD/ERS analysis: Offline analyses were carried out using the three baseline conditions: LB1, LB2, and SB. The classification problem was defined as MI versus rest, with the rest class corresponding to each baseline condition. Classi-

fication was performed separately for each baseline type and for each MI window onset combination. For the calibration data, classification performance was evaluated using a 5-fold cross-validation procedure. For data recorded during the online session, the classifier was trained on the calibration dataset and evaluated on the online data, thereby simulating real-time operation.

Figure 4 presents the classification performance and ERSP results obtained from the calibration data, averaged across all participants. Both classification accuracy and minimum ERSP values varied substantially with baseline selection and MI window onset. When LB1 was used as the baseline, classification accuracy was consistently high and stable across all MI windows, with Win_{Acc}^i ranging from 81 to 83.3%. This indicates robust separability between MI and resting-state conditions for all windows. Consistently, ERSP analysis revealed strong desynchronization during MI with low variability across window shifts. Using LB2 resulted in lower classification accuracy and increased variability to MI window onset, with accuracies ranging from 77.0% to 83.4%, suggesting reduced discriminability between MI and rest. ERSP values were also attenuated and exhibited greater temporal variability, particularly for later MI window onsets. A marked degradation in performance was observed when using SB as the baseline. In this condition, classification accuracy became highly dependent on the relative timing of baseline and MI windows. The highest accuracies (approximately 70.0–74.6%) were obtained for early MI window onsets (0.5–2.0 s). Similarly, ERSP analysis revealed stronger desynchronization for early MI windows and weaker ERSP values for later window onsets.

Results obtained from the online data showed similar trends but with increased variability (Fig. 5). LB1 again yielded the most stable performance across MI window onsets, with the highest and most consistent performance and ERSP values. For LB2, both classification performance and ERSP values were reduced in the online condition. When using SB, Win_{TPR}^i exhibited substantial variability, ranging from near-chance levels (52%) to high TPR of approximately 82%. The particularly low TPR observed for later MI window onsets may indicate that participants did not sustain motor imagery until the end of the cue period.

Spectral Analysis: The spectral analysis of the three baseline conditions shows consistent relative power distributions across the theta, alpha, and beta bands for the channels C3 and C4 (Fig. 6). The theta band presents relatively low power across all baseline conditions, with only minor variations. LB1 and LB2 have a slightly higher alpha relative power compared to SB, which may indicate a more relaxed resting state. In contrast, the SB condition shows relatively increased beta activity, possibly reflecting residual task engagement or motor anticipation between MI trials. Although consistent trends were observed across baseline conditions, Friedman tests revealed no significant effects for alpha power at C3 ($p = 0.097$) and C4 ($p = 0.895$), nor for beta power at

Table 1: The trial-level detection rate (eq. 3) and mean sliding-window TPR (eq. 2) obtained during the online operation.

Subjects	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean	STD
$Trial_{DR}(\%)$	95	100	100	90	100	100	100	100	100	85	97.0	5.4
$Win_{TPR}(\%)$	54.4	83.9	53.9	80.0	92.8	79.4	65.0	72.2	80.6	32.8	69.5	18.0

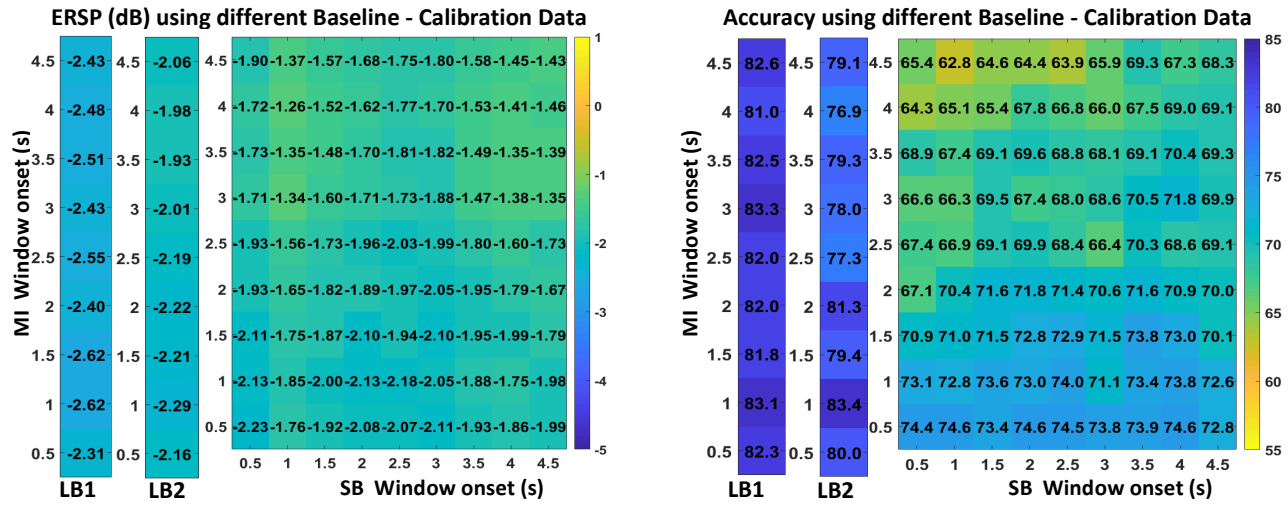


Figure 4: Confusion matrices for ERSP and classification performance from calibration data: Left - ERSP for different baseline conditions (LB1, LB2, SB (per window)) across MI window onsets. Right - Classification performance for the same baseline conditions.

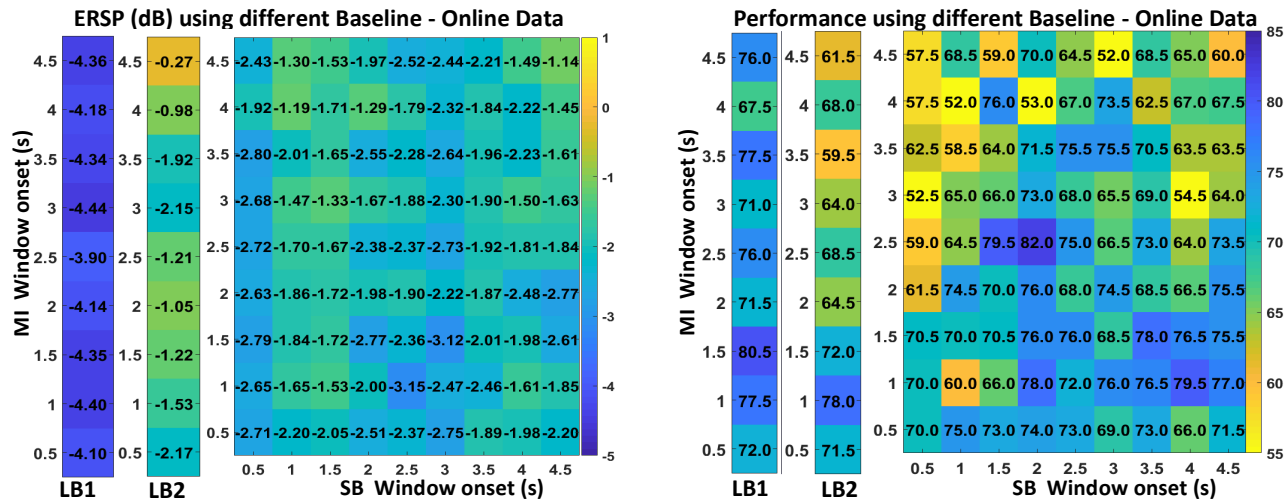


Figure 5: ERSP and classification performance for different baseline conditions (LB1, LB2, SB (per window)) across MI window onsets during the online phase.

C3 ($p = 0.121$) and C4 ($p = 0.895$).

DISCUSSION

Baseline definition and temporal window selection substantially influence both ERSP and classification performance, with important implications for online BCI operation. Using LB1 consistently yielded the strongest and most stable desynchronization, which translated into the highest and most stable performance in both calibration and online data. In contrast, LB2 exhibited attenuated ERSP and moderate performance degradation, suggesting that rest recorded after task execution

is partially influenced by residual cognitive engagement, thereby reducing discriminability from MI. The most pronounced effects were observed when using SB, where classification accuracy strongly depended on the selected time window onset. Although high accuracies could be achieved by carefully selecting MI and baseline timing window, performance was unstable. A repeated-measures ANOVA revealed significant differences among baseline conditions in the calibration data ($F = 10.6$, $p = 0.0009$). Post-hoc pairwise comparisons indicated that LB1 achieved significantly higher accuracy than SB ($p = 0.0005$), and no significant differences were ob-

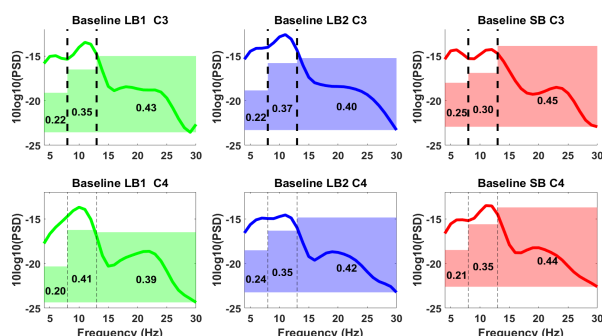


Figure 6: Grand average of power spectral density for channels C3 and C4 during the three resting conditions. The shaded areas indicate the theta, alpha, and beta frequency bands. The values inside each band correspond to the relative power for that frequency range.

served between LB1 and LB2 ($p = 0.6$). The comparison between LB2 and SB approached significance ($p = 0.05$). In contrast, no significant differences between baseline conditions were found for the online data ($F = 0.7$, $p = 0.51$). Across all baseline conditions, performance was higher for calibration data than for online data. However, these metrics are not directly comparable, since calibration performance was evaluated using accuracy, and online performance was assessed using TPR. Nevertheless, the lower online performance is consistent with the performance degradation reported in [12]. This study showed that when feedback does not match the participant's expectations, it can lead to frustration and disrupt focus. Furthermore, in our setup, the robotic hand actuation triggered after the first correctly classified window may have also reduced participants' engagement in motor imagery, considering they had already achieved their goal.

CONCLUSION

This work systematically investigated the impact of baseline definition and temporal window selection on both classification performance and ERD/ERS, under calibration and online conditions for post-stroke rehabilitation. The results consistently demonstrated that a long resting period recorded before task execution provides the best overall performance among the three baseline conditions. It constituted the most stable baseline, producing strong ERD together with the highest and most robust performance across both calibration and online data. In contrast, long rest recorded after task execution led to reduced ERD magnitude and degraded classification performance, likely due to residual cognitive or task-related engagement. The greatest performance degradation was observed when using a short rest period between MI trials as the baseline, possibly due to contamination by residual sensorimotor activity. Although this baseline is commonly used in BCI systems, these findings highlight the importance of baseline selection in achieving reliable and robust motor imagery decoding, especially in online BCI operation.

ACKNOWLEDGEMENTS

This work was supported by the Portuguese Foundation for Science and Technology under Grants UpReGain (<https://doi.org/10.54499/2023.13809.PEX>), BCI4ALL (COMPETE2030-FEDER00842800, doi: <https://doi.org/10.54499/2023.17977.ICDT>), and <https://doi.org/10.54499/UID/PRR2/00048/2025>.

REFERENCES

- [1] Baniqued PDE et al. Brain-computer interface robotics for hand rehabilitation after stroke: A systematic review. *Journal of neuroengineering and rehabilitation*. 2021;18(1):15.
- [2] Ntetska A, Mimou A, Tzimourta KD, Angelidis P, Tsipouras MG. A deep learning approach to classifying user performance in bci gaming. *Electronics*. 2025;14(24):4974.
- [3] Sharma N, Baron JC. Does motor imagery share neural networks with executed movement: A multivariate fmri analysis. *Front. Hum. Neurosci*. 2013;7:564.
- [4] Pfurtscheller G, Da Silva FL. Event-related eeg/meg synchronization and desynchronization: Basic principles. *Clinical neurophysiology*. 1999;110(11):1842–1857.
- [5] Tangwiriyasakul C, Verhagen R, Van Putten MJ, Rutten WL. Importance of baseline in event-related desynchronization during a combination task of motor imagery and motor observation. *Journal of Neural Engineering*. 2013;10(2):026009.
- [6] Piazza C, Visintin E, Reni G, Montirosso R. The effect of baseline on toddler event-related mu-rhythm modulation. *Brain Sciences*. 2021;11(9):1159.
- [7] Rimbart S, Trocellier D, Lotte F. Impact of the baseline temporal selection on the erd/ers analysis for motor imagery-based bci. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2023, 1–4.
- [8] Jones TA. Motor compensation and its effects on neural reorganization after stroke. *Nature Reviews Neuroscience*. 2017;18(5):267–280.
- [9] Ho N et al. An emg-driven exoskeleton hand robotic training device on chronic stroke subjects: Task training system for stroke rehabilitation. In: 2011 IEEE international conference on rehabilitation robotics. 2011, 1–5.
- [10] Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*. 2000;8(4):441–446.
- [11] Cruz A, Kuzmanoski M, Pires G. Optimizing bci rehabilitation protocols for stroke: Exploring task design and training duration. In: 2025 IEEE 8th Portuguese Meeting on Bioengineering (ENBENG). 2025, 57–60.
- [12] Pires G et al. A new error-monitoring brain-computer interface based on reinforcement learning for people with autism spectrum disorders. *Journal of neural engineering*. 2022;19(6):066032.