

Volitional and Non-Volitional Motor Tasks Encode Distinct EEG Dynamics: Implications for Paradigm-Specific BCI Decoder Design

F. Morales-Magallón¹, E. Bojorges-Valdez¹, D.C. González-González²

¹ Engineering Studies for Innovation Department, Universidad Iberoamericana Ciudad de México, Mexico City, Mexico

² Geriatric Evaluation Center, Universidad Iberoamericana Ciudad de México, Mexico City, Mexico

E-mail: fernandomoralesm01@gmail.com, erik.bojorges@ibero.mx

ABSTRACT: This work explores differences in brain dynamics during the execution of voluntary and non-voluntary knee-joint movements (using an isokinetic dynamometer) by analyzing distance matrices constructed from a d -dimensional reconstructed phase-space of instantaneous band-power time series (iBPTs). Two condition-specific decoders were developed: a volitional (**Vol**) model, calibrated exclusively on active motor execution, and a non-volitional (**NVol**) model trained on machine-assisted movement. The nonlinear features derived from the phase-space distance matrix successfully encode condition-specific dynamical structures, yielding classification metrics consistent with current BCI standards, even for data acquired across a several-day interval. The **Vol** model achieved higher performance when classifying voluntary data (AUC: 0.772 ± 0.128) than non-voluntary data (AUC: 0.613 ± 0.159), whereas the **NVol** model demonstrated consistent performance across both conditions (AUC: 0.694 ± 0.119 for voluntary and AUC: 0.697 ± 0.100 for non-voluntary). A Linear Discriminant Analysis (LDA) classifier, combining the outputs of both models, revealed condition-dependent decoder contributions, confirming that volitional and non-volitional motor states elicit separable dynamical regimes in the EEG signal. These results indicate that although transfer learning from overt motor execution enables reliable classification of non-voluntary data, the underlying dynamical differences lead to degraded performance. Consequently, while calibration on overt movement may result in reduced accuracy in covert scenarios, the use of covert-movement data may provide more representative performance metrics for online implementations.

INTRODUCTION

A large body of EEG-based brain-computer interface (BCI) research relies on the extraction of features derived from Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), quantified across frequency bands associated with motor activity [1, 2]. The μ (8-13 Hz) and β (13-30 Hz) bands, localized predominantly over the sensorimotor cortex, are the most widely employed [2–6], given their well-established sensitivity to both motor and somatosensory processes. The

γ band has additionally been implicated in movement onset facilitation [7], broadening the spectral repertoire relevant to motor decoding.

A common assumption underlying many BCI systems is that the neural dynamics of motor imagery (MI) and motor execution (ME) are sufficiently similar—exhibiting comparable data distributions and dynamical processes, encompassing both cortical regions and frequency band dynamics—to support cross-condition generalization. That is, models trained on ME data can be transferred to MI tasks without substantial performance loss [8]. However, this assumption is increasingly challenged. ERD/ERS-based classifiers consistently plateau at approximately 70% accuracy [8], even without accounting for the BCI “illiteracy” population. This suggests a non-trivial gap between overt and imagined movement that spectral features alone cannot bridge. Assistive technologies, such as therapeutic robotics or human-assisted rehabilitation, allow for the execution of non-volitional limb movements that enhance neuroplasticity by promoting the activation of afferent pathways [9]. This scenario provides an MI-like environment where the user observes limb motion without corresponding efferent pathway activation; however, the mechanisms underlying overt and covert motion differ substantially [10]. We argue that these inter-condition differences constitute a fundamental dynamical distinction that linear spectral methods are structurally incapable of resolving.

Nonlinear dynamical approaches have been proposed as a means to capture motor-related EEG information that remains undetectable via conventional frequency-domain analyses [11, 12]. Methods grounded in phase-space reconstruction—such as Recurrence Quantification Analysis (RQA)—have demonstrated sensitivity to subtle, condition-specific changes in neural dynamics that elude linear features [11, 12].

Within this framework, the instantaneous band-power time series (iBPTs) provides a tractable representation that preserves the temporal structure of band-specific power fluctuations while enabling estimation of nonlinear features via distance matrices in reconstructed phase-space.

In this work, we evaluate the performance of paradigm-specific decoders, trained for volitional (**Vol**) and non-

volitional (**NVol**) limb control, in decoding movement-related cortical activity. Two independent CNN-Transformer models are trained on mutually exclusive datasets corresponding to each motor condition. The relative contribution of each decoder is subsequently quantified via a Linear Discriminant Analysis (LDA) model, providing a direct empirical measure of representational separability between **Vol** and **NVol** neural states. This framework offers both a principled test of the MI/ME equivalence assumption and a practical foundation for designing condition-aware decoders in BCI systems.

MATERIALS AND METHODS

Experimental Paradigm:

The experimental paradigm and data acquisition followed the protocol described in a previous work [13]. Fifteen healthy participants (eight males, seven females; mean age 23 ± 3 years) were recruited. The experimental protocol was approved by the Ethics Committee of Universidad Iberoamericana (Folio 221), and all participants provided written informed consent prior to enrollment.

Participants performed three knee movement exercises using an isokinetic dynamometer (System 4 Quick-Set, Biodex, USA): two volitional conditions at different load levels (10% and 5% of maximum voluntary torque) and one non-volitional (assisted) movement condition. Maximum torque was determined individually for each subject during isometric contractions in both flexion and extension prior to the experimental sessions.

Each participant completed three sessions, with each session adhering to a structured experimental protocol: Assessment of maximum isometric torque in flexion and extension; 4–6 cycles of volitional movement at 10% of maximum torque (**VM10**): 40 s of activity followed by 20 s of rest; 4–6 cycles of volitional movement at 5% of maximum torque (**VM5**): 40 s of activity followed by 20 s of rest; 4–6 cycles of non-volitional movement (**NVM**): 40 s of activity followed by 20 s of rest.

Signal Acquisition:

Electrophysiological and kinematic signals were simultaneously recorded using two synchronized g.USBamp amplifiers (g.tec Medical Engineering, Austria).

EEG signals were acquired with a 16-channel montage following the extended 10–20 international system, using the following electrode positions: FC3, FCz, FC4, CP3, C3, C1, Cz, C2, C4, CP4, P3, Pz, P4, O1, Oz, and O2 as seen in Fig 1, A. The ground electrode was placed at Fpz and the right earlobe served as the common reference.

The second amplifier was dedicated to kinematic signals from the isokinetic dynamometer, including joint angular velocity, position, and torque. This channel also recorded surface electromyography (EMG) signals to assess muscular activity and to verify the absence of voluntary muscle activation during the **NVol** condition.

All signals were digitized at a sampling rate of 4800 samples per second. Initial signal conditioning included a band-pass filter from 0.1 to 2000 Hz and a 60 Hz notch

filter to attenuate power-line interference. EEG signals were preprocessed using Independent Component Analysis to remove artifacts using the mne package [14, 15]. To increase the signal-to-noise ratio of the EEG signals, it was used a “common average reference”. EEG signals were subsequently downsampled to 500 Hz for further processing.

Feature Extraction:

EEG signals were processed through a filter bank of fourth-order Butterworth band-pass filters to extract three physiologically relevant frequency bands: μ , β , and γ . The filtered signals were subsequently squared sample-wise to obtain the Instantaneous Band Power Time Series (iBPTs) shown in Fig 1 B.

Phase-Space distance matrices:

Recordings were segmented into windows of 2 seconds length with no overlapping. Within each window, the band power time series of the seven selected channels (FC3, FCz, C3, C1, C2, Cz, CP3) were embedded using a time-delay embedding with $\tau = 100$ ms and embedding dimension $d = 3$, yielding a 21-dimensional phase-space representation according to Takens theorem [16]. Each space was built from the projection of the sample n , $n - \tau$ and $n - 2\tau$ of each channel. In total, three 21-dimensional spaces were built one per frequency band. Euclidean distances between all pairs of embedded points within each window were computed to produce an unthresholded weighted distance matrix. The matrices derived from the β , μ , and γ bands were then stacked to form a multichannel representation of brain dynamics, which served as the input for a deep learning architecture. Figure 1B illustrates representative examples of the phase-space and distance matrices derived via this method.

Model Architecture:

Each paradigm-specific decoder utilized a CNN-Transformer architecture designed to extract dynamical structures from the distance-pattern image representation.

The input image was processed by a convolutional patch extractor comprising three successive Conv2D–Batch Normalization blocks. The first two Conv2D layers applied 32 filters each with 3×3 kernels, same padding, and linear activations; the third applied 15 filters with a 3×3 kernel and ReLU activation. A MaxPooling2D layer with pool size 4×4 then down-sampled the feature maps, projecting them into a sequence of non-overlapping patch embeddings with projection dimension 15.

A positional encoding layer was applied to the patch sequence to preserve spatial order. The encoded sequence was then passed through three stacked Transformer blocks, each comprising a Layer Normalization layer, a Multi-Head Self-Attention (MHSA) module with $h = 2$ attention heads. This design enables the model to capture global dependencies across distant regions of the distance pattern matrix, reflecting the complex temporal dynamics inherent in the EEG signal.

Following the final Layer Normalization, a Global Average Pooling layer aggregated the sequence into a fixed-

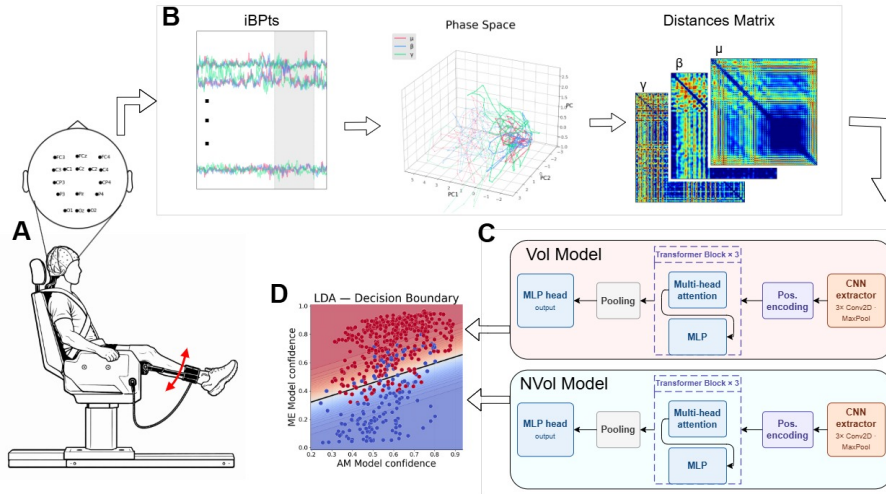


Figure 1: Schematic of the BCI pipeline. **A.** Experimental setup showing the isokinetic machine, right leg movement, and EEG cap with electrode configuration. **B.** Signal processing and feature extraction: cleaned EEG signals are used to compute iBPTs for the μ , β , and γ bands; the center panel shows the phase-space representation of the d -dimensional embedding used to extract Euclidean distances and Distances Matrix of each band. **C.** Two paradigm-specific models: Volitional (**Vol**, top) and Non-volitional (**NVol**, bottom), sharing the same architecture but trained on different datasets. **D.** Linear Discriminant Analysis (LDA) showing the contribution of both models to the true label.

length feature vector. This vector was passed through two fully connected Dense-Dropout blocks before a final Dense layer with softmax activation for binary classification (Activity vs. Rest).

Paradigm-Specific Training:

Two independent decoders were trained on mutually exclusive datasets: the **Vol** model was trained exclusively on volitional motor task recordings (**VM10** condition), while the **NVol** model was trained exclusively on non-volitional passive movement recordings (**NVM** condition). This paradigm-specific training strategy was adopted to preserve the distinct neural representations associated with each motor condition, rather than conflating them into a single shared representation.

Data from sessions S1 and S2 were used exclusively for model development. Files were partitioned at the window level into training (65%), validation (15%), and internal test (20%) subsets using stratified random sampling, with the validation subset used for early stopping and hyperparameter monitoring. Session S3 was held out entirely prior to any model development and served as the independent evaluation set for all reported performance metrics, ensuring no overlap between training data and final assessment.

Both the **Vol** and **NVol** decoders were trained once using pooled data across all subjects, yielding subject-independent models. This design choice reflects the assumption that the dynamical structures captured by the distance-pattern representation are consistent across individuals.

Assessment of Model Contribution:

The scalar output of each paradigm-specific decoder was normalized via z-score standardization on the training data and the held-out validation session, data was used as a two-dimensional feature vector for a Linear Discrimi-

nant Analysis (LDA) fusion classifier, trained per subject on sessions training data for S1 and S2. The resulting discriminant coefficients provide a direct measure of the relative contribution of each decoder to the final classification decision, computed as:

$$w_k = \frac{|\beta_k|}{|\beta_{\text{Vol}}| + |\beta_{\text{NVol}}|} \times 100\% \quad (1)$$

where β_k denotes the LDA coefficient associated with decoder k . LDA was evaluated on the held-out session S3, consistent with the evaluation protocol of the paradigm-specific decoders.

Statistical Analysis:

Classification performance was assessed using three metrics computed per subject on the held-out session S3: Area Under the ROC Curve (AUC), accuracy (ACC), and precision. These metrics were evaluated independently for the **Vol** model and **NVol** model.

Pairwise comparisons between **Vol** and **NVol** were performed using the Wilcoxon signed-rank test for paired samples ($\alpha = 0.05$), with subjects as the pairing unit. Four comparisons were conducted: **Vol** vs **NVol** within each motor condition (**NVM**, **VM10**), and cross-condition comparisons for each model separately. A Bonferroni correction was applied to all p-values to account for multiple comparisons.

Cross-condition comparisons within each model (**VM10** vs **NVM** for **Vol**, and **VM10** vs **NVM** for **NVol**) were performed independently using the Wilcoxon signed-rank test ($\alpha = 0.05$), with no correction for multiple comparisons applied.

Relative LDA weights on each model (**Vol** and **NVol**) were compared using the Wilcoxon signed-rank test for paired samples. Comparison was applied independently for **NVM** and **VM10** conditions.

RESULTS

Figures 2a and 2b illustrate the model predictions on the validation session of a representative subject, for the **NVM** and **VM10** conditions respectively. In both cases, the condition-specific model yielded higher performance on its corresponding task, while exhibiting reduced metrics when applied to the alternative condition.

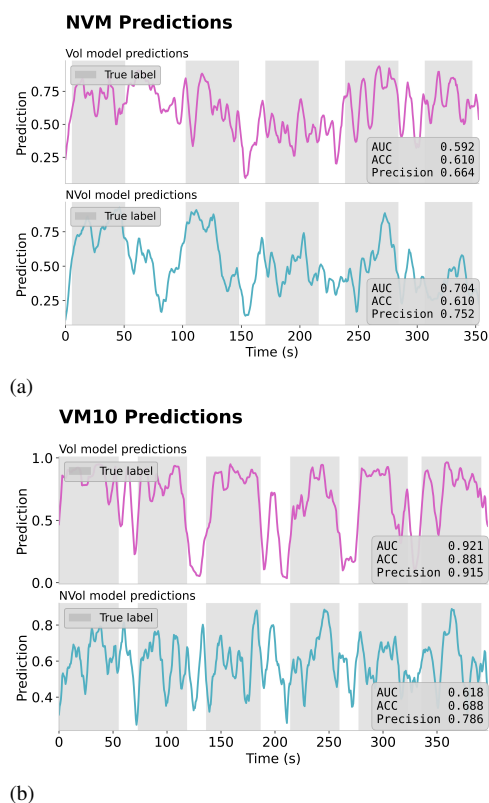


Figure 2: Model predictions on validation recordings. Dark magenta corresponds to the **Vol** model predictions and dark cyan to the **NVol** model predictions. In each plot, the gray shaded region represents the true activity label. AUC, Accuracy, and Precision scores are shown at the bottom.

Figure 3 presents the within-condition validation metrics for the **NVM** and **VM10** recordings across all subjects using both models. Across all metrics, the **NVol** model achieves superior performance in classifying **NVM** data, whereas the **Vol** model performs better on **VM10** data. This behavior demonstrates that models constructed for a specific task are well-suited for similar data, even when those data were not acquired on the same day. The figure also shows the cross-dataset comparison of model performance; p-values are reported after Bonferroni correction for multiple comparisons, where the **Vol** model approaches statistical significance in AUC, with higher mean values observed for the corresponding model dataset pairs. However, no statistically significant differences were found.

Table 1 presents a cross-condition numerical comparison, evaluating how each model generalizes when applied to a recording outside its training condition. The **NVol** model

exhibited consistent performance across both **VM10** and **NVM** recordings; the only statistically significant difference was found in Precision, reflecting slightly higher precision on its native **NVM** task. In contrast, the **Vol** model showed a more pronounced performance gap between conditions: while it performed well on its native **VM10** recording, its metrics declined notably on **NVM** data, with statistically significant differences found in AUC and ACC metrics. This suggests limited transferability of the learned parameters to the non-volitional condition.

Table 1: Comparison of model performance across **VM10** and **NVM** recordings. Values are mean $\pm$ SD. P-values from Wilcoxon signed-rank test; no correction for multiple comparisons applied.

Model	Metric	VM10	NVM	pval
Vol	AUC	0.772 ± 0.128	0.613 ± 0.159	0.0166
	ACC	0.692 ± 0.105	0.588 ± 0.099	0.0245
	Precision	0.825 ± 0.097	0.789 ± 0.088	0.3258
NVol	AUC	0.694 ± 0.119	0.697 ± 0.100	0.9032
	ACC	0.644 ± 0.091	0.630 ± 0.087	0.8077
	Precision	0.785 ± 0.073	0.833 ± 0.060	0.0494

Figure 4 displays the relative LDA weights, reflecting the contribution of each model to the fusion classifier across both recording conditions. A statistically significant difference between model contributions was observed in both conditions, with the **VM10** recording yielding a higher level of significance ($p < 0.05$) than the **NVM** recording. Notably, for the **VM10** data, all weights from the **Vol** model achieve higher values than those of the **NVol** model. Similarly, for the **NVM** data, the **NVol** weights are larger than those of the opposing model, although this trend is not absolute like that observed in the **Vol-VM10** pairing.

DISCUSSION

The results suggest that the two paradigm-specific decoders capture distinct dynamical signatures of motor-related cortical activity derived from phase-space analysis. As illustrated in Figure 3 and Table 1, model-data comparisons yield superior performance when each model is applied to its respective training condition—specifically, the **NVol** model on **NVM** data and the **Vol** model on **VM10** data. In general, the **NVol** model exhibits less dispersion across all evaluated metrics, suggesting that it maintains a more stable performance when capturing both volitional and non-volitional motor dynamics.

The cross-condition comparison in Table 1 further supports this interpretation: **NVol** predictions remained largely consistent across both **VM10** and **NVM** conditions, implying that the dynamical patterns captured during passive movement overlap substantially with those present during active execution. In contrast, the **Vol** model showed a more pronounced performance decline when transferred to **NVM**, showing statistical signifi-

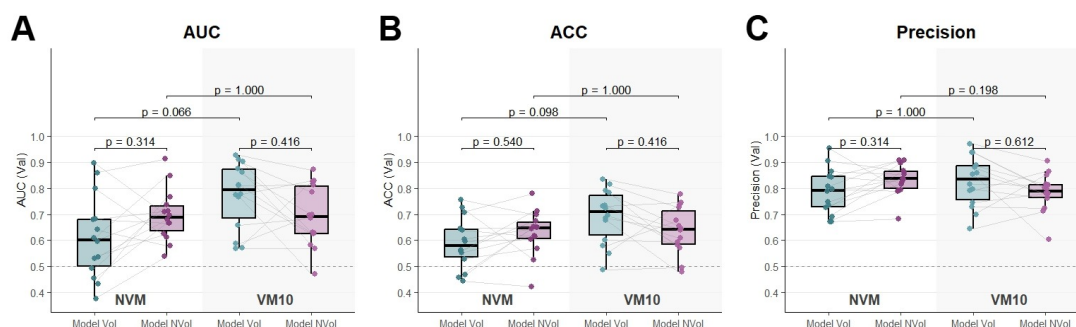


Figure 3: Validation metrics for **Vol** (cyan) and **NVol** (magenta) models evaluated on **NVM** and **VM10** datasets. P-values for between-group comparisons are reported after Bonferroni correction. Panels A, B, and C illustrate AUC, Accuracy, and Precision metrics, respectively. Individual data points represent subject-specific evaluations, while boxplots characterize the distribution across each metric.

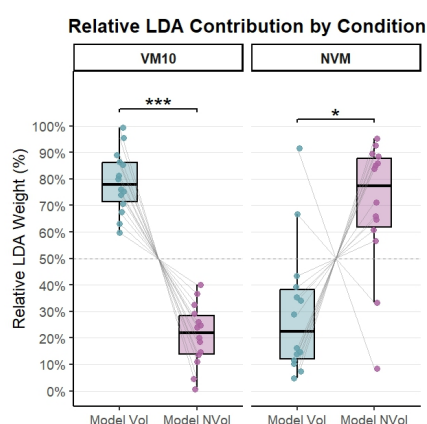


Figure 4: Relative LDA contribution by condition. The contribution of each model, **Vol** (dark cyan) and **NVol** (dark magenta), is shown for the two recordings, **VM10** and **NVM**. Triple asterisk (***) denote $p < 0.001$ and single (*) $p < 0.05$.

cance for AUC and ACC ($p = 0.0166$ and $p = 0.0245$ respectively). This indicates a greater difficulty in generalizing beyond its specific training condition.

The LDA weights shown in Figure 4 corroborate that each model exerts a greater influence when classifying its respective data, suggesting that the underlying dynamics of the two experimental conditions are distinct. In the **VM10** data, the LDA classifier relied predominantly on the **Vol** model, whereas in the **NVM** data, the contributions from both models were more balanced.

This asymmetry supports the hypothesis that a model trained on the dynamical patterns of volitional motor execution differs fundamentally from one trained on non-volitional movement, with the primary distinction residing in the involvement of efferent pathways. While both the **NVM** and **VM10** conditions involve afferent communication that elicits motor cortex responses, these results demonstrate that the presence of efferent communication significantly impacts classifier performance. The overall lower performance observed in the **NVM** condition relative to **VM10** may be attributed to the absence of these volitional motor commands and the associated sensorimotor feedback loop. Without a gen-

erated volitional drive, the stable cortical attractor that both models attempt to track may be disrupted; consequently, any resting-state epoch in which the subject incidentally reaches a dynamically stable cortical state could confound the model output, potentially destabilizing classification metrics independently of true motor intent. Given that efferent pathways are intrinsically linked to volitional control, these findings suggest that BCI paradigms must incorporate explicit feedback to effectively engage volitional mechanisms and establish robust dynamical regimes.

The LDA fusion model carries an inherent limitation in this context: by operating on the scalar outputs of each decoder, it cannot fully capture the nonlinear structure of the underlying feature distributions. Nevertheless, it provides a valuable interpretability tool, allowing the relative contribution of each decoder to be quantified directly from the discriminant coefficients.

Taken together, these findings indicate that the **Vol** model is better suited to capturing the dynamics of active motor execution, while the **NVol** model generalizes more robustly across conditions. Consequently, BCI paradigms calibrated on overt limb movement may exhibit performance degradation when transferred to scenarios lacking efferent pathways. The reduced performance observed in both models under the **NVM** condition supports the hypothesis that passive paradigms involve a disruption of the stable cortical attractor associated with overt movement. This effect may be further exacerbated by the absence of closed-loop sensorimotor feedback, a limitation that could be observed in motor imagery-based BCI contexts.

Additionally, the classification results demonstrate that the proposed method utilizing RQA-like features constitutes a reliable framework for the development of BCI systems. Its performance on unseen data, acquired several days after the initial calibration phase, aligns with current performance standards and suggests a low incidence of BCI “illiteracy” among participants. Mean AUC values were 0.772 ± 0.128 for the **Vol-VM10** pair and 0.697 ± 0.100 for the **NVol-NVM** pair. Figures 2a and 2b show representative outputs for each model, demon-

strating reliable control within a hypothetical BCI environment.

The relatively small sample size is taken as a limitation on this study, which may constrain statistical power and limit the generalization of the findings, particularly given the inherent inter-subject variability in cortical dynamics that could not be fully characterized.

CONCLUSION

The nonlinear features of the iBPs, specifically the distance matrix derived from the phase-space embedding, are capable of encoding condition-specific dynamical structures sufficient to construct a functional BCI. Both the **Vol** and **NVol** models successfully captured the underlying iBPs dynamics, yielding classification metrics consistent with current BCI standards. Nevertheless, it is important to emphasize that, given the inherently distinct dynamical regimes of volitional and non-volitional movement, each model was calibrated on fundamentally different dynamical parameters, and this divergence has a measurable effect on cross-condition prediction performance.

The **Vol** and **NVol** models captured distinct cortical dynamics, with the **Vol** model achieving optimal performance during volitional execution and the **NVol** model generalizing more consistently across both experimental conditions. This suggests that BCI calibration based on overt motor execution results in performance degradation when transferred to other states, whereas training on covert motor movement yields more consistent, albeit lower, results. Linear Discriminant Analysis (LDA) further revealed that the relative contribution of each decoder is condition-dependent, confirming that volitional and non-volitional motor states elicit separable dynamical regimes within the EEG signal.

The overall performance degradation observed under non-volitional conditions suggests that the absence of efferent motor drive and closed-loop sensorimotor feedback disrupts the stable cortical attractor that underlies reliable decoding. This has implications on BCI training paradigms, as it highlights the importance of model calibration and questions whether the decoders trained on motor execution transfer seamlessly to assisted or even imaginary-based conditions as the intrinsic dynamics of these signal changes.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Research Office and the Department of Engineering Studies for Innovation at Universidad Iberoamericana Ciudad de México for the financial support provided to attend this conference.

REFERENCES

[1] Pfurtscheller G. Central beta rhythm during sensorimotor activities in man. *Electroencephalogr Clin Neurophysiology*. 1981.

- [2] Pfurtscheller G, Brunner C, Schlögl A, Lopes da Silva F. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*. 2006;31(1):153–159.
- [3] Szurhaj W et al. Basic mechanisms of central rhythms reactivity to preparation and execution of a voluntary movement: A stereoelectroencephalographic study. *Clinical Neurophysiology*. 2003;114(1):107–119.
- [4] McFarland DJ, Miner LA, Vaughan TM, Wolpaw JR. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topography*. 2000;12(3):177–186.
- [5] Muthuraman. Source analysis of beta-synchronisation and cortico-muscular coherence after movement termination based on high resolution electroencephalography. *PLOS ONE*. 2012.
- [6] Stolk A et al. Electrographic dissociation of alpha and beta rhythmic activity in the human sensorimotor system. *eLife*. 2019;8:e48065.
- [7] Ulloa JL. The control of movements via motor gamma oscillations. *Frontiers in Human Neuroscience*. 2022;15.
- [8] Jayalakshmi Srinivasan K, Periasamy P, Gunasekaran S. Motor imagery and motor execution: A narrative review of electroencephalographic (EEG) signatures, methodological consistency, and translational applications. *Cureus*. 2025;17(9):e93011.
- [9] Yu Y et al. Enhanced neuroplasticity and gait recovery in stroke patients: A comparative analysis of active and passive robotic training modes. *BMC Neurology*. 2025;25(1):239.
- [10] Edelman BJ et al. Non-invasive brain-computer interfaces: State of the art and trends. *IEEE Reviews in Biomedical Engineering*. 2024;1–25.
- [11] Marwan N, Meinke A. Extended recurrence plot analysis and its application to ERP data. *International Journal of Bifurcation and Chaos*. 2004;14(02):761–771.
- [12] Pitsik E et al. Motor execution reduces EEG signals complexity: Recurrence quantification analysis study. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2020;30(2).
- [13] Morales-Magallón F, Bojorges-Valdez E. Intended and non-volitional knee joint movements elicit distinct functional brain networks. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2025, 1–6.
- [14] Gramfort A et al. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*. 2013;7(267):1–13.
- [15] Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*. 2007;34(4):1443–1449.
- [16] Takens F. Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, Warwick 1980. Springer Berlin Heidelberg: Berlin, Heidelberg, 1981, 366–381.