Characterizing Expectation Mismatch in a Brain-Controlled Upper-Body Rehabilitation Exoskeleton

Satyam Kumar, Kanishka Mitra, *Graduate Student Member, IEEE*, Deland H. Liu, Hussein Alawieh, Frigyes Samuel Racz, *Member, IEEE*, Stefano Dalla Gasperina, Ashish D. Deshpande, and José del R. Millán, *Fellow, IEEE*

***Abstract*—Robot-assisted therapy has long promised to advance stroke rehabilitation by delivering intensive and personalized training, yet its clinical impact remains limited. Closing the sensorimotor loop with brain–computer interfaces offers a better strategy than passive mobilization, directly linking user intent to robotic assistance and potentially driving neuroplasticity. However, a brain-computer interface requires subject-specific calibration that is time-consuming and often impractical. Moreover, brain decoding remains error-prone due to variability of neural signals, thus resulting in unintended robot actions that could reduce engagement during closed-loop control. Here, we demonstrate that a decoder trained on an expert subject can be transferred to naïve users for online control of a rehabilitation exoskeleton in a rest-versus-reaching paradigm, a functional task with clinical relevance. We then characterize error-related potentials arising from expectation mismatches between brain commands and robot actions during closed-loop control. Finally, we show that these mismatches can be reliably decoded in a subject-independent framework (mean area under the receiver operating characteristic curve: 0.77), a crucial step toward rehabilitation scenarios where collecting subject-specific error-related potential data is challenging. Our findings highlight the potential for integrating real-time error-detection to enhance human–robot interaction by correcting unintended robot behaviors, which could significantly improve rehabilitation outcomes where accurate and contingent feedback is essential.**

***Index Terms*—Brain-machine interfaces, rehabilitation robotics, prosthetics and exoskeletons.**

I. Introduction

REHABILITATION robots aim to enhance the treatment of motor dysfunctions post-stroke; however, these systems have yet to yield clinically relevant recovery [1]. From a neurophysiological perspective, robot-assisted interventions adopt a bottom-up approach that tries to promote neural recovery by acting on the distal (physical) level [2]. Such an approach assists movements without necessarily engaging the brain, thus exerting limited influence on the neurophysiological processes that drive recovery after stroke. Hence, there is a critical need for rehabilitation robotics to incorporate top-down principles, where actuation at the distal level is mediated by sensorimotor activity. While previous studies have utilized electromyography (EMG) to detect residual muscle activity and trigger exoskeleton control [3], such methods are limited when muscle activity is minimal or absent, as is often the case for patients with severe stroke. An alternative is to rely on brain-computer interfaces (BCI) to directly convert cortical signals that carry the patient's motor intention into desired robot movements. If robot-assisted rehabilitation were driven by neurophysiological activity, it would optimize the effects of sensory feedback by promoting activity-dependent plasticity, where descending neural motor signals and ascending feedback are contingent on each other [4].

Motor imagery (MI) is a widely used BCI modality to control external devices such as robots [5]. During MI, a motor action is kinesthetically imagined without actually executing it, eliciting an event-related desynchronization of sensorimotor rhythms [6], [7]. Achieving reliable control of a MI-based BCI typically requires extensive closed-loop training, where users receive feedback to refine their control over time [8], [9], [10]. This training involves building decoders through user-specific calibration sessions followed by using these decoders for feedback to improve BCI control. Calibration-free BCIs, which utilize data from other subjects to build the decoder, eliminate the

Received 29 May 2025; accepted 26 October 2025. Date of publication 19 November 2025; date of current version 27 November 2025. This article was recommended for publication by Associate Editor E. Battaglia and Editor H. Yu upon evaluation of the reviewers' comments. This work was supported by Coleman Fung Foundation. *(Corresponding author: Satyam Kumar.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of The University of Texas at Austin under Application No. IRB-2020030073, and performed in line with the Declaration of Helsinki.

Satyam Kumar, Kanishka Mitra, Deland H. Liu, and Hussein Alawieh are with the Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: satyam.kumar@utexas.edu; mitrakanishka@utexas.edu; deland.liu@utexas.edu; hussein@utexas.edu).

Frigyes Samuel Racz is with the Department of Neurology, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: fsr324@austin.utexas.edu).

Stefano Dalla Gasperina is with the Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: stefano.dallagasperina@austin.utexas.edu).

Ashish D. Deshpande is with the Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712 USA, and also with Meta Reality Labs Research, Redmond, WA 98052 USA (e-mail: ashish@austin.utexas.edu).

José del R. Millán is with the Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA, also with the Department of Neurology, The University of Texas at Austin, Austin, TX 78712 USA, and also with the Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: jose.millan@austin.utexas.edu).

Digital Object Identifier 10.1109/LRA.2025.3634889

need for open-loop calibration sessions where users do not receive feedback on their BCI control. Calibration-free BCIs have been shown to promote long-term BCI training and skill acquisition [11]. Although transfer learning in BCIs is well established, it has been mainly explored in offline settings with a focus on algorithmic improvements [12], [13]. Here, we show how the expert-to-naïve decoder transfer paradigm, previously demonstrated online for left vs right hand MI [11], extends to functional arm movements relevant to rehabilitation.

Another key component in the BCI training is receiving congruent concomitant feedback through external devices such as functional electric stimulation [4], wheelchairs [9] and exoskeletons [14] during online BCI operation. However, due to inaccuracies in MI-based decoding, erroneous behavior of the exoskeleton could confuse the users and curb their motivation [15]—potentially hindering the closed-loop training process.

Previous works have shown that when humans perceive erroneous behavior of an external agent [16], [17], [18], subjects elicit error-related potentials (ErrPs). These ErrPs, originating from the anterior cingulate cortex, are reflected primarily in the fronto-central EEG and show dominant activity in the theta band [19]. Previous literature has not explicitly characterized or utilized ErrPs resulting from expectation mismatches specifically between intended MI and the executed robot actions during active MI-based rehabilitation exoskeleton control.

To address this issue, we characterize the presence of ErrPs associated with the erroneous behavior of an upper-body rehabilitation exoskeleton during active MI-based control. Although prior studies [20], [21], detected ErrPs in rehabilitation settings, these involved subjects with residual limb activity performing exercises with support from the contralateral hand. In contrast, our work characterizes ErrPs when subjects rely solely on MI to trigger exoskeleton movements, a paradigm that could extend to stroke rehabilitation in individuals with little or no residual movement.

As detection of erroneously delivered commands is essential in BCI-controlled robotics, our goal in this study was to assess the feasibility of detecting ErrPs during the operation of MI-controlled upper-body exoskeleton. This would allow the BCI to correct unintended erroneous movements of the exoskeleton in rehabilitative scenarios, thereby enhancing contingent feedback and increasing patient engagement, both of which are critical aspects for maximizing recovery [22], [23]. Using a calibration-free MI-based BCI setup, we show that naïve healthy users could reliably operate an upper-body rehabilitation exoskeleton. Moreover, we demonstrate the presence of ErrPs associated with erroneous robot behavior. We then illustrate that these ErrPs can be reliably decoded in a subject-independent framework with high confidence. Although the present study was limited to healthy participants, these findings lay the groundwork for future translation into closed-loop rehabilitation, where ErrPs could be used to correct unintended robot actions during rehabilitation of paretic individuals.

In summary, this work makes three main contributions. First, we extend the concept of expert-to-naïve decoder transfer to a rest vs. reaching MI paradigm, going beyond traditional left–right MI tasks toward a more functional setting that is directly relevant to rehabilitation, where collecting calibration data is often impractical. Second, we characterize ErrPs that arise during closed-loop MI-based exoskeleton control—an aspect that is particularly important for rehabilitation, as reliable detection of expectation mismatches can help ensure contingent and meaningful feedback. Finally, we demonstrate that these ErrPs can be decoded in a subject-independent manner, which is crucial in rehabilitation contexts where collecting subject-specific calibration data may be difficult or infeasible.

II. Methods

In this section, we first describe the study design and experimental setup used to characterize ErrPs during BCI-controlled exoskeleton use, along with its main components, including the BCI system and the rehabilitation exoskeleton. Harmony, our rehabilitation exoskeleton, is a bilateral upper-limb device providing seven actuated degrees of freedom per arm to support anatomically aligned shoulder and elbow motion. We then outline the methodology for building the MI BCI decoder, followed by the methods used to characterize the presence of ErrPs. Next, we present the decoding frameworks employed to classify ErrPs at the single-trial level. Finally, we describe the metrics used to quantify classification performance for both MI and ErrP decoding.

A. Participants and Experimental Protocol

Nine non-disabled healthy participants with normal or corrected-to-normal vision participated in the study (mean age: 20.55 ± 1.01 years, all right-handed, 3 females). The experimental protocol was approved by the Institutional Review Board of The University of Texas at Austin (IRB-2020030073) and all participants provided written informed consent before conducting the measurements. Brain activity was collected from 59 scalp locations according to international 10-10 standard using a 64-channel eego EEG system (ANT Neuro, Germany) while the subjects were strapped to Harmony (Fig. 1). The four electrodes from the temporal and mastoid regions (T7, T8, M1, M2) were excluded due to their susceptibility to noise and artifacts. A fifth unused channel was a dedicated hanging electrooculogram (EOG) electrode. Instead, EOG signals were recorded using bipolar electrodes as described in [24] and used to identify and reject trials potentially contaminated by eye-movement artifacts. Subjects performed two recording sessions on two consecutive days. Day 1 consisted of an offline session (used to collect subject-specific calibration data, which are typically used to build BCI decoders) and an online session where closed-loop BCI feedback was provided. Day 2 consisted only of an online session. Each day consisted of 6–8 MI runs, 20 trials per run split evenly between rest (active imagination of remaining at rest) and right-hand reach MI. The number of runs varied with fatigue and fitting adjustments, while the total session time remained approximately constant. For online sessions, BCI feedback was provided based on the output of a decoder that was trained on EEG data from an expert subject rather than the decoder built on subject-specific calibration data. The latter decoder-transfer

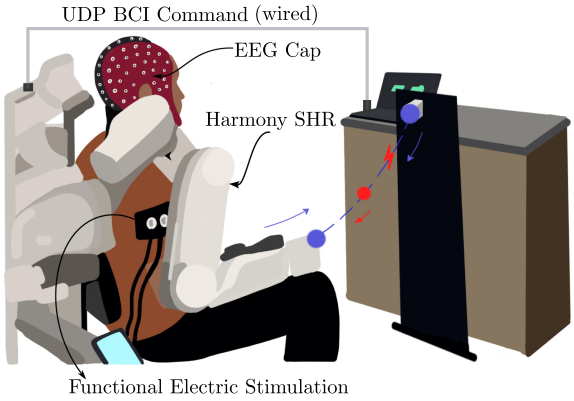

Fig. 1. **Experimental Setup.** Subjects operated an MI-BCI to control Harmony, an upper-limb exoskeleton, in reaching for objects. Successful delivery of reaching MI command made Harmony moved towards a goal location (purple dot). Failure to sustain reaching MI caused the robot to stop midway (red lightning bolt) and return to the starting position. Successful rest MI kept Harmony stationary. Failure to sustain rest MI triggered a robot movement to a preprogrammed location (red circle).

approach is detailed in [11]. Subject-specific calibration data (offline runs) were used only for offline benchmarking of decoding models, which is beyond the scope of the present study.

The expert subject was a participant who consistently achieved BCI classification accuracies above 85% on multiple MI experimental protocols. This subject only performed an offline session like all other participants and was therefore not included in the analyses reported in this study.

In the online experimental protocol, participants were presented with a sequence of visual cues. The protocol began with a 2 s fixation period, followed by a 1.5 s cue indicating the type of imagery to be performed; i.e., rest (active imagination of remaining at rest) or right-hand reaching MI. After the cue, participants received continuous feedback through a bar on the screen along with sensory threshold neuromuscular electric stimulation (stNMES; RehaMove3, Hasomed GmbH, Germany) on the triceps brachii proportional to the posterior probability estimated by the BCI decoder. Stimulation thresholds were individualized: the minimum was defined as the lowest perceptible current, and the maximum was set one step (0.5 mA) below the level that produced a visible contraction. During online operation, stimulation intensity was scaled between these thresholds in proportion to the classifier output. The BCI decoder estimated the posterior probability of each class (rest or reach) using a one-second sliding window of EEG signals, updated every 62.5 ms. These probabilities were accumulated using an exponential smoothing approach, and a trial ended when the accumulated probability for one class crossed a participant-specific threshold, at which point the corresponding command was delivered. Participants were required to reach this threshold to successfully deliver a BCI command within 7 s. In a given trial, participants could either surpass the correct threshold (hit), exceed the opposite class threshold (miss), or fail to reach any threshold within the 7 s window (timeout). Thresholds were participant-specific and adjusted across runs to provide a suitable challenge level, maximizing correct deliveries while minimizing misses and timeouts. Initial thresholds in each session were set using a brief mock online run, and subsequent adjustments were guided by the bar dynamics observed during each run to maintain an appropriate level of challenge.

Delivery of the rest MI class kept Harmony stationary while delivery of the reach MI command initiated the robot trajectory toward a predefined goal location (Fig. 1). Upon delivery of the MI command, the continuous feedback was replaced by discrete feedback, while Harmony executed the delivered command (rest or reaching), prompting the participant to sustain the MI for an additional 6 seconds. During the sustain phase of reaching MI, the stNMES stimulation intensity was set at the participant-specific maximum threshold. To successfully sustain MI, subjects had to maintain the accumulated posterior probability above a relaxed threshold relative to that required for BCI hits. If participants failed to sustain the appropriate MI, Harmony executed an erroneous action resulting in a mismatch between the participant's expectation and the robot's action, thus eliciting an ErrP. Precisely, during a rest trial subjects received stNMES and Harmony initiated a pre-programmed trajectory before returning to its original resting point. In a reaching trial, if the participant failed to sustain MI after successfully initiating the reaching movement, Harmony stopped mid-trajectory and then returned to its starting position. The main motivation behind employing discrete feedback during the sustain phase was to analyze scenarios where participants could not continuously anticipate failure (e.g., the bar moving towards the opposite threshold or timeout), as they might with continuous bar feedback, thereby creating a stronger expectation mismatch.

B. *Harmony SHR Exoskeleton*

We employed the Harmony SHR exoskeleton (Harmonic Bionics Inc., United States), a bimanual rehabilitation robot designed for clinical and neurological therapy applications [25]. Each arm of the device provides seven active degrees of freedom (DOFs): five at the shoulder complex, one at the elbow, and one for forearm pronation/supination at the wrist. This kinematic structure enables anatomically aligned upper-limb movements across a wide range of tasks. Harmony can deliver assistance, resistance, or perturbations to user-generated motions while recording detailed kinematic, kinetic, and end-effector pose data, making it well suited for studies of human–exoskeleton interaction.

In our current experimental protocol, neural activity decoded via the BCI system is used to initiate predefined robot-assisted tasks, forming a high-level neural trigger for movement assistance [26]. Once triggered, the Harmony exoskeleton executed task-specific trajectories using its baseline impedance controller. This control scheme corresponds to the so-called passive-triggered mode [26], where movement assistance is delivered after the wearer initiated the action through a high-level trigger (e.g., BCI, gaze, EMG, or residual force signals), thereby decoupling intention detection from low-level joint control.

The BCI system communicated with the exoskeleton through a Boolean signal $g_{\text{bci}} \in \{0, 1\}$. A value of $g_{\text{bci}} = 1$ indicated that assistance should be provided (i.e., start reaching trajectory), whereas $g_{\text{bci}} = 0$ signaled no assistance. Our protocol set $g_{\text{bci}} = 1$ both when a reach command was correctly detected

and when participants failed to maintain rest imagery, leading to unintended activations. Conversely, $g_{\text{bci}} = 0$ occurred when rest was correctly sustained and when participants failed to maintain reach motor imagery, causing the exoskeleton to stop mid-trajectory and return to the starting position. Together, these cases define the correct and erroneous interactions between the BCI system and the robot.

The overall control torque applied to Harmony is given by

$$\tau_{\text{control}} = \tau_{\text{dyn}} + \tau_{\text{shr}} + g_{\text{bci}} \cdot \tau_{\text{task}}, \tag{1}$$

where each term corresponds to a distinct control objective. The term τ_{dyn} compensates for the robot's gravity, inertia, and friction. τ_{shr} enforces scapular-humeral rhythm coordination by coupling humeral elevation and scapular rotation to preserve a natural motion ratio during arm elevation [27]. τ_{task} provides assistive torque that drives the reaching task, modulated by the Boolean gain g_{bci}, which enables assistance ($g_{\text{bci}} = 1$) or reverts to baseline transparent behavior ($g_{\text{bci}} = 0$).

C. Riemannian Geometry Framework in BCI

We used the Riemannian geometry framework [28] to decode reaching vs. rest MI. Let $X \in \mathbb{R}^{N_t \times N_c}$ represent a multi-channel EEG sample, where N_t denotes the number of time samples and N_c the number of EEG channels. The normalized covariance matrix $\Sigma \in \mathbb{R}^{N_c \times N_c}$ for this sample is calculated as

$$\Sigma = \frac{X^T X}{\text{trace}(X^T X)} \tag{2}$$

Since the matrix Σ is positive definite (PD), it resides on a differentiable Riemannian manifold. Distances between points on this manifold are computed using the concept of curved distances, known as geodesics [28].

We employ the commonly used distance metric on the manifold of PD matrices, namely the affine invariant Riemannian metric. The affine invariance property of this metric asserts that the Riemannian distance δ_r between two covariance matrices $[\Sigma_1, \Sigma_2]$ remains unchanged, even after applying the affine transformation (W) to both matrices independently [12]

$$\delta_r(\Sigma_1, \Sigma_2) = \delta_r(W^T \Sigma_1 W, W^T \Sigma_2 W) \tag{3}$$

where $\delta_r(\Sigma_1, \Sigma_2)$ is the Riemannian distance between two PD matrices Σ_1 and Σ_2. This affine invariance property enables the effective transport of matrices on the manifold without distorting their distributions, making it a robust tool for handling non-stationarities in EEG samples [11], [12].

D. Decoding Model and Expert Decoder Transfer

For decoding, we used a Minimum Distance to Mean (MDM) classifier, a common model in Riemannian geometry-based BCI. During the training phase, class-specific Riemannian means (or prototypes) of each class are estimated from the covariance matrices. Then, during the testing phase, the estimated covariance matrix of an incoming sample is classified by identifying the class whose prototype has the minimum Riemannian distance to that sample.

We used a recently proposed subject-independent online feedback training framework [11] that utilizes an MDM decoder of an expert MI-BCI subject in performing the reach vs. rest task. Due to inter-subject and inter-session variability, the covariance features across different subjects are shifted on the Riemannian manifold [12], [13]. The key component of the online subject-independent training framework is to minimize the distribution shift across subjects. Precisely, the feature spaces of the expert and the new subject are matched by applying whitening transformations that bring the centers of both distributions to the same point on the Riemannian manifold. The transformation for the expert (T^E) is estimated as the Riemannian mean of the covariance features from all her/his available data, pooled over both classes/tasks. Next we apply the affine transformation on their covariance data (Σ_i^E)

$$\tilde{\Sigma}_i^E = (T^E)^{-\frac{1}{2}} \Sigma_i^E (T^E)^{-\frac{1}{2}'} \tag{4}$$

The MDM decoder is built on the affine-transformed covariance matrices $(\tilde{\Sigma}_i^E)$ of the expert data. For an incoming i^{th} covariance sample (Σ_i^s) of a new subject s, the affine transformation matrix (T_i^s) is estimated per sample incrementally (see [11] for details). We then recenter the i^{th} covariance sample (Σ_i^s) using the recentering transform T_i^s in (4). Finally, the recentered covariance sample is classified using the MDM expert decoder.

E. Neurophysiological Analysis of ErrPs

To characterize ErrPs, we performed grand-average analyses in the time and time–frequency domains to capture both temporal and spectral characteristics, followed by decoding analysis to assess their sample level discriminability.

1) Time Domain Analysis: For the time domain analysis of ErrPs, we first employed a non-causal 4th-order Butterworth bandpass filter with cutoff frequencies [1, 10] Hz. The preprocessed EEG signals were then divided into epochs of [-0.5, 1.5] s relative to the mismatch event. We used a rejection scheme based on EOG channels to discard epochs that might be contaminated with ocular or facial activity [24]. Additionally, any epoch in which EEG amplitude exceeded 70 μV in any channel was deemed artifact-contaminated. In total, 36.46% of epochs were removed. We categorized 'error' epochs based on the presence of expectation mismatch, which could either be due to Harmony unexpectedly stopping while users are trying to sustain reaching MI or due to Harmony unexpectedly initiating a movement while users try to sustain rest MI. Sustain trials without the above-mentioned events are labeled as 'correct' trials (including both successfully sustained reaching and resting trials). Finally, we also analyzed expectation mismatches during the initial initiation phase, i.e., when participants failed to correctly initiate the corresponding BCI command. For time domain analysis, we computed the grand average as the mean EEG signal across all participants and trials within each condition ('error' and 'correct') at electrode FCz (located at the fronto-central midline in the 10–10 system), since ErrP components are prominently observed over fronto-central sites such as FCz and Cz [16].

2) Time-Frequency Analysis: Time-frequency decomposition of single-trial EEG data was performed using complex Morlet wavelets. We applied a wavelet with 6 cycles and a frequency resolution of 0.1 Hz to achieve higher precision in the frequency domain compared with shorter wavelets (e.g., 3–4

cycles), consistent with prior literature [29]. Following decomposition, we squared the magnitudes of the wavelet-transformed data and averaged them across trials to calculate the average total power for each frequency at electrode FCz, as ErrPs are linked to increased fronto-central theta power [19]. The power at each time-frequency point was then normalized in decibels (dB) relative to the baseline interval of [-0.2, -0.1] s before the event. To compute the theta power ratio between error and correct class associated with ErrPs, we averaged the ratio of baseline normalised power between 'error' and 'correct' trials within the [0, 1] s window post-event. This yielded the ratio of power across frequencies, which we subsequently averaged over the theta band range of [4, 8] Hz.

3) Sample-Level Decoding: Unlike MI, where sample-level covariance captures sensorimotor modulations, event-related potentials (ERPs) are predominantly characterized by their temporal features. A sample-level covariance structure may not adequately capture these detailed temporal characteristics. To cope with this limitation, we use extended covariances, where grand average ERPs are concatenated with sample-level trials [12], [30]. Precisely, let $X \in \mathbb{R}^{N_t \times N_c}$ be a EEG sample and $X_{\text{ERP}} \in \mathbb{R}^{N_t \times N_c}$ be a grand averaged ERP (i.e., template signal); an extended trial structure can be formed by concatenating the sample level EEG and grand averaged ERP as $X_{\text{ext}} = [X_{\text{ERP}}, X]$.

Using (2) we can estimate the trace normalised covariance of the extended trial, whose block structure is

$$\Sigma_{\text{ext}} = {X_{\text{ext}}}^T X_{\text{ext}} = \begin{bmatrix} X_{\text{ERP}}^T X_{\text{ERP}} & X_{\text{ERP}}^T X \\ X^T X_{\text{ERP}} & X^T X \end{bmatrix} \tag{5}$$

Since ErrPs are known to be spatially confined (Fig. 2), we selected a subset of electrodes (FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CP2, CP4) for decoding. While the extended covariance approach effectively captures the temporal and spatial characteristics of single-trial ErrPs, it does not fully address the issue of variability in feature space, which manifests in non-stationary single-trial ErrPs and inter-subject differences. Given our experimental setting with a low number of trials, cross-subject ErrP classification becomes essential. Similar to [12], we applied the recentering framework in (4), where the extended covariances of training and test set are independently recentered. For comparison, we also evaluated classification performance without the recentering step (referred as template matching) to assess its impact on addressing inter-subject variability.

F. Classification Metrics

Here, we describe the metrics used to assess online MI BCI performance and sample-level ErrP classification.

1) BCI Online Performance: For MI classification we used a 22-channel montage similar to [10], [11]. During the online sessions, the MI feedback was provided using the expert decoder on 1-second samples with a step size of 1/16 seconds. Here we report Cohen's kappa (κ) [31] value on those 1-second samples for each session and every subject. In particular, we report the command delivery performance of the BCI to control Harmony based on hits and misses and normalised to timeouts using

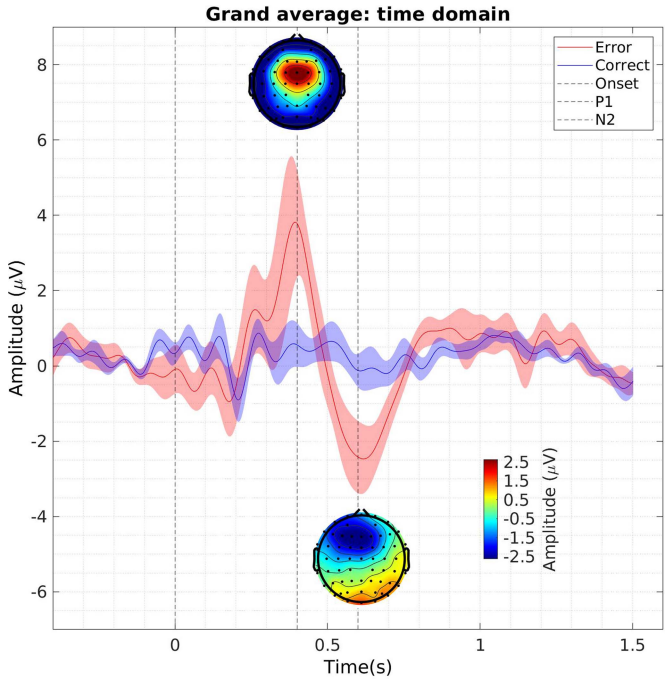

Fig. 2. **Grand average ErrPs.** Time-locked grand average of erroneous and correct classes across different subjects ($n = 8$) at the FCz channel, with shaded regions indicating the standard error of the mean. T=0, vertical black dashed line, corresponds to the onset of the expectation mismatch event. Insets display the topography of the grand average EEG amplitude for erroneous epochs averaged across a 50 ms window centered around the identified positivity and negativity peaks of the ErrP; i.e., 0.40 and 0.60 s.

normalised kappa value [10], [11]. Finally, we also report the accuracy of the BCI commands delivered to Harmony during the sustain phase.

2) ErrP Classification: We report the Area Under the Curve (AUC) score of the receiver-operator characteristic for evaluating the classification of erroneous trials compared with correct trials. Given that our experimental protocol design led to a small sample setting, we report the ErrP classification using a Leave-One-Subject-Out Cross-Validation (LOSOCV) scheme. Additionally, in the Riemannian framework described earlier, the grand averaged ErrP was estimated across subjects in the training fold and used to build the extended trial structure. At test time, the same grand averaged ErrP was utilized to construct the extended trial structure on the testing subject's data to mimic the LOSOCV scheme.

III. Results

In this section, we first present the online MI BCI performance, followed by the characterization of ErrPs in the temporal and spectral domains. Finally, we report the subject-independent single-trial decoding performance of ErrPs.

A. Online BCI Performance

Analysis of sample-level online BCI performance using the expert decoder across the two online sessions reveals that, in the first session, the group-level κ value across subjects was

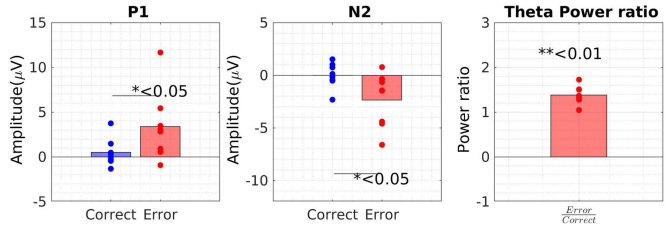

Fig. 3. **Neural markers of expectation mismatch.** Panels present the various neural markers of expectation mismatch for correct and erroneous classes. Each dot corresponds to the average neurophysiological marker for a subject, aggregated across corresponding types of trials. p-values for P1, and N2 were estimated using paired t-tests, while the significance of theta power ratio was assessed using a right-tailed t-test relative to one. $(*)$ and $(**)$ indicates the significant p-values.

0.1273 ± 0.1385, while performance increased to 0.1989 ± 0.1717 in the second session. A paired t-test indicated this improvement in BCI classification performance as statistically significant ($p = 0.0438$). Moreover, we also observe a statistically significant ($p = 0.0053$) improvement in κ of BCI command delivery to Harmony for initiating the rest or reach commands from the first online session (0.2584 ± 0.2262) to the second one (0.4144 ± 0.2618). Finally, we observe improved control during the sustain phase as accuracy of sustaining the delivered BCI commands increased from the first online session ($60.54 \pm 19.40\%$) to second one ($67.34 \pm 17.30\%$); however, this increase was statistically insignificant ($p = 0.1328$).

B. Neurophysiological Correlate of Expectation Mismatch

All reported ErrP analyses and figures correspond to data from session 2, which we focused on for two reasons: a) Subjects exhibited improved performance in the MI tasks during day 2, resulting in more successful command deliveries and more frequent initiation of sustain trials. b) After post-hoc rejection of artifact-contaminated epochs, the number of trials in session 2 was higher than in session 1. Of note, one participant was excluded from the analysis for not having both error and correct trials on session 2 (resulting in n= 8 participants). Across all participants, the total number of ErrP trials was 66 erroneous and 145 correct in session 1, while 137 erroneous and 317 correct in session 2.

Fig. 2 presents the grand average time domain plots across different subjects for both the erroneous and correct classes. Notably, within each class, two distinct types of actions are grouped together: for the erroneous class, these are 'failure to sustain resting' and 'failure to sustain reaching'; for the correct class, these are 'successful sustain of resting' and 'successful sustain of reaching'. The observed patterns include a positivity around 0.4 seconds (P1), followed by a negativity around 0.6 seconds (N2). However, the grand-average analysis did not reveal characteristic ErrP components during the initiation phase (i.e., failure to correctly initiate the BCI command); therefore, these trials were excluded from the plotted averages and subsequent ErrP analyses.

Post-hoc statistical analysis (Fig. 3) using paired t-tests revealed a significantly higher P1 amplitude during error trials across subjects, with the amplitude for correct trials being $0.4883 \pm 1.5332\,\mu V$ and for incorrect trials $3.3628 \pm 3.8827\,\mu V$ ($p = 0.017$). All eight subjects exhibited a higher P1 amplitude for the error class.

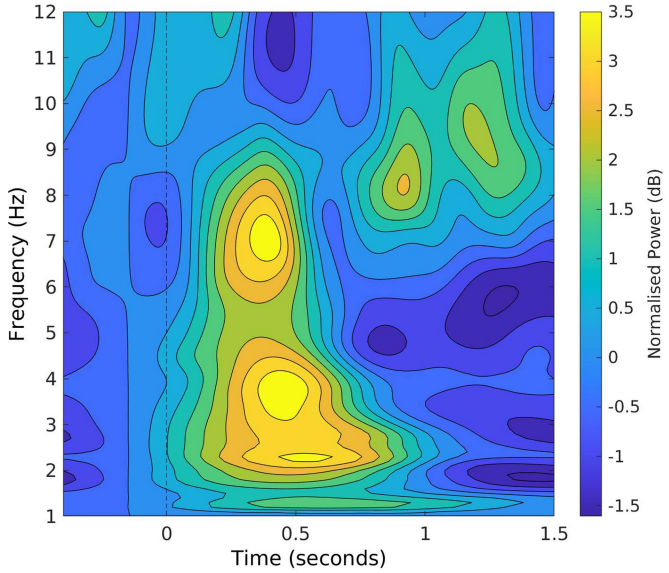

Fig. 4. **Time-frequency decomposition of ErrPs.** Difference in time-frequency representations between erroneous and correct trials across subjects (n = 8). The figure highlights changes in the theta band power ([4,8] Hz) over time ([-0.4, 1.5] s) relative to event onset at the FCz channel.

TABLE I
PERFORMANCE METRICS FOR EACH PARTICIPANT

Subject ID	Template matching	Template matching + Recentering
1	0.6000	0.6207
2	0.7836	0.7967
3	0.7899	0.8311
4	0.6556	0.5333
5	0.9321	0.9383
6	0.7612	0.7969
7	0.6702	0.7362
8	0.9627	0.9800
Mean	0.7694	0.7791
Std Dev	0.1287	0.1495

Statistical analysis of the N2 showed a significantly lower amplitude in error trials than in correct trials ($p = 0.023$), with values of $-0.0148 \pm 1.1732\,\mu V$ for correct trials and $-2.3444 \pm 2.5567\,\mu V$ for error trials. Seven out of eight subjects exhibited a lower N2 amplitude for the error class.

Theta power analysis (Fig. 4) revealed that theta power in error class was higher in all eight subjects compared with correct class (Ratio: 1.3815 ± 0.20). A right-tailed statistical t-test, using 1 as the threshold, indicated a significantly higher theta power in error than correct classes ($p = 0.0005$).

C. Decoding Error-Related Potentials

Table I displays the AUC performance metrics for all participants individually. Using the template matching approach in the Riemannian framework, all the participants showed an AUC

score better than chance level of 0.5 (0.7694 ± 0.1287) using the decoder trained on the data of the rest of the cohort.The purpose of employing the recentering framework was to address between-subject non-stationarities, which often result in shifts within feature spaces [12]. When integrating the recentering approach with template matching, the group-level AUC performance reached 0.7791 ± 0.1495. Although this did not result in statistically significant improvement, it is noteworthy that 7 out of 8 subjects demonstrated an increase in AUC over template matching.

IV. Discussion

In this study, we aimed to characterize the presence of ErrPs associated with erroneous behaviors performed by a BCI-controlled exoskeleton. The performance of our online BCI, which directly controlled the exoskeleton behavior, showed a statistically significant improvement in classification and command delivery between the first and second sessions. Similar to the previous finding that an MI-BCI decoder trained on data of a single expert can enable immediate online BCI control for naïve subjects [11], our study showed that subjects achieved better than chance level performance (kappa>0) in online sample-level MI classification as well as command delivery performance using an expert decoder. However, consistent with prior literature, early-stage performance remained low, as is typical for naïve participants beginning with MI-based BCI control who gradually improve with longitudinal training [10], [11]. Although accuracy in sustaining triggered commands improved in the second session, it was still insufficient for reliable BCI control. This highlights the need for complementary modalities, such as ErrP detection, to correct erroneous robot actions and provide more congruent feedback. As hypothesized, the expectation mismatch created by the exoskeleton's unexpected behavior (i.e., robot actions not aligning with the participant's intent) generated ErrPs over prefrontal brain regions. Grand averages of the error-class trials in FCz exhibited physiologically valid modulations consistent with those reported in the literature [17], [19] (Fig. 2). The statistical analysis (Fig. 3) further confirmed that these modulations are specific to the error class, showing significant amplitude differences between the error and correct classes. Additionally, the time-frequency decomposition of the differences in the grand averages of the error and correct classes at FCz (Fig. 4) revealed localized theta band activity (4–8 Hz) between 200 ms and 800 ms post-trigger onset, which coincided temporally with the P1 and N2 components, as expected for ErrPs. We also observed an error-related negativity (ERN) which typically occurs before the P1 peak and around 100–200 ms after error onset [16], [19]. However, we did not observe a significant difference in ERN between the error and correct classes. The exact temporal characteristics of ErrPs can vary across tasks and paradigms due to the influence of the mental task on device operation [32].

Collecting task-specific calibration data for building subject-specific ErrP decoders is cumbersome and prone to provide wrong feedback in the early stages of subject's training. By using data from other subjects who performed the same task, we could build a subject-independent decoder that efficiently recognizes erroneous actions performed by the brain-controlled exoskeleton for new subjects. Consistent with existing literature, we also find that incorporating domain adaptation via the recentering framework enhances classification performance by addressing inter-subject variability [11], [12].

A. Limitations and Future Work

While our study establishes the presence of expectation mismatch-based events—and the elicitation as well as robust recognition of ErrPs—in an MI-BCI paradigm that controls a rehabilitation robot, there are several limitations that need to be addressed in future work. The study was limited to a sample size of 9 subjects, and the expectation mismatch analysis focused on the second online session. Although this session had a higher number of erroneous trials compared with the first, the overall trial count remains relatively low, which is a consequence of the task design. Expanding the study to include more subjects and additional closed-loop recordings could help validate and strengthen the findings.

Furthermore, increasing the number of online sessions might provide deeper insights into the morphology of ErrP signals, as longitudinal training has been shown to improve MI-BCI control, potentially revealing how ErrP signal characteristics evolve with greater expectation mismatch. Interestingly, our analysis did not reveal the characteristic ErrP pattern in the erroneous class during the initial class mismatch, unlike the mismatch observed in the sustain phase. This could be attributed to the continuous visual feedback provided during the task initiation phase, which may have led subjects to anticipate an unsuccessful trial execution. In contrast, during the sustain phase, the absence of continuous bar feedback for MI may have created a clearer expectation mismatch, thereby eliciting a classical ErrP.

Although these preliminary experiments were conducted with healthy participants, acknowledging that stroke patients may exhibit different neurophysiological responses, this study provides a foundation for future clinical application and validation. We hypothesize that although stroke patients may exhibit altered neural dynamics and face greater challenges in MI decoding, the similarity of characteristic ErrP patterns in individuals with stroke [20], [21] supports the potential of using ErrPs to correct unintended behaviors of rehabilitation robots, thereby compensating for decoding inaccuracies and enhancing rehabilitation outcomes. These prior works successfully decoded ErrPs, although they did so in a passive BCI setting with patients who retained residual limb activity. In contrast, our approach could extend these findings by enabling ErrP-based correction during MI-driven control, making it applicable even for patients with little or no residual movement.

While this study demonstrated successful ErrP decoding in an offline setting, real-time ErrP decoding is crucial for correcting the robot's erroneous actions in practical settings [17], [18], [33]. Furthermore, unlike the current design, where the reaching trajectory was fixed across subjects, it is essential to account for individual differences in preferred trajectories. Previous works have shown that different subjects have varying preferences for

robot trajectories [18], [33], [34], making it crucial to generate individual trajectories for rehabilitation so as to enhance subject's embodiment of the rehabilitation robot, and, thus, eliciting stronger ErrPs.

Finally, we acknowledge the absence of EMG monitoring as a limitation. In actual rehabilitation protocols, tracking muscular activation alongside BCI control is important, since excessive reliance on brain-driven assistance without improvement in residual muscular activity during motor tasks could hinder reinforcement of natural motor pathways [35].

V. Conclusion

In this work, we showed how naïve subjects could control an MI-BCI upper-limb exoskeleton to reach for objects using a subject-independent BCI built with data from a single expert. Critically, we have characterized the presence of ErrPs when the rehabilitation exoskeleton exhibited erroneous behavior in an online BCI setting. Our analysis showed the presence of significant physiological modulations associated with ErrPs. We demonstrated that these ErrPs could be reliably decoded in a subject-independent framework, thereby enhancing the reliability of BCI-controlled robots. By combining calibration-free control with error characterization and detection, our study takes a step toward brain–robot interfaces that are practical and clinically relevant for neurorehabilitation.

References

[1] J. Mehrholz, M. Pohl, T. Platz, J. Kugler, and B. Elsner, "Electromechanical and robot-assisted arm training for improving activities of daily living, arm function, and arm muscle strength after stroke," *Cochrane Database Systematic Rev.*, no. 9, 2018.

[2] J.-M. Belda-Lois et al., "Rehabilitation of gait after stroke: A review towards a top-down approach," *J. Neuroengineering Rehabil.*, vol. 8, pp. 1–20, 2011.

[3] N. Ho et al., "An EMG-driven exoskeleton hand robotic training device on chronic stroke subjects: Task training system for stroke rehabilitation," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, 2011, pp. 1–5.

[4] A. Biasiucci et al., "Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, 2018.

[5] L. Tonin and J. D. R. Millán, "Noninvasive brain-machine interfaces for robotic devices," *Annu. Rev. Control Robot., Auton. Syst.*, vol. 4, no. 1, pp. 191–214, 2021.

[6] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[7] D. H. Liu, S. Kumar, H. Alawieh, F. S. Racz, and J. D. R. Millán, "Personalized μ-transcranial alternating current stimulation improves online brain–computer interface control," *J. Neural Eng.*, vol. 22, no. 1, 2025, Art. no. 016037.

[8] S. Perdikis and J. D. R. Millán, "Brain-machine interfaces: A tale of two learners," *IEEE Syst., Man, Cybern. Mag.*, vol. 6, no. 3, pp. 12–19, Jul. 2020.

[9] L. Tonin et al., "Learning to control a BMI-driven wheelchair for people with severe tetraplegia," *iScience*, vol. 25, no. 12, 2022, Art. no. 105418.

[10] H. Alawieh et al., "Electrical spinal cord stimulation promotes focal sensorimotor activation that accelerates brain-computer interface skill learning," *Proc. Nat. Acad. Sci. USA*, vol. 122, no. 24, 2025, Art. no. e2418920122.

[11] S. Kumar, H. Alawieh, F. S. Racz, R. Fakhreddine, and J. D. R. Millán, "Transfer learning promotes acquisition of individual BCI skills," *PNAS Nexus*, vol. 3, no. 2, 2024, Art. no. pgae076.

[12] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.

[13] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2390–2401, Aug. 2019.

[14] M. Tavella, R. Leeb, R. Rupp, and J. D. R. Millán, "Towards natural non-invasive hand neuroprostheses for daily living," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, 2010, pp. 126–129.

[15] T. Sollfrank et al., "The effect of multimodal and enriched feedback on SMR-BCI performance," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 490–498, 2016.

[16] P. W. Ferrez and J. D. R. Millán, "Error-related EEG potentials in brain-computer interfaces," in *Toward Brain-Computer Interfacing*, G Dornhege, J. D. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, Eds., Cambridge, MA, USA: MIT Press, 2007.

[17] C. Lopes-Dias, A. I. Sburlea, and G. R. Müller-Putz, "Online asynchronous decoding of error-related potentials during the continuous control of a robot," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 17596.

[18] I. Batzianoulis et al., "Customizing skills for assistive robotic manipulators, an inverse reinforcement learning approach with error-related potentials," *Commun. Biol.*, vol. 4, no. 1, 2021, Art. no. 1406.

[19] R. Chavarriaga, A. Sobolewski, and J. D. R. Millán, "Errare machinale est: The use of error-related potentials in brain-machine interfaces," *Front. Neurosci.*, vol. 8, 2014, Art. no. 208.

[20] A. Kumar, E. Pirogova, S. S. Mahmoud, and Q. Fang, "Classification of error-related potentials evoked during stroke rehabilitation training," *J. Neural Eng.*, vol. 18, no. 5, 2021, Art. no. 056022.

[21] A. Kumar, Q. Fang, J. Fu, E. Pirogova, and X. Gu, "Error-related neural responses recorded by electroencephalography during post-stroke rehabilitation movements," *Front. Neurorobot.*, vol. 13, 2019, Art. no. 107.

[22] N. Hogan et al., "Motions or muscles? Some behavioral factors underlying robotic assistance of motor recovery," *J. Rehabil. Res. Develop.*, vol. 43, no. 5, 2006, Art. no. 605.

[23] F. A. Bright, N. M. Kayes, C. Cummins, L. M. Worrall, and K. M. McPherson, "Co-constructing engagement in stroke rehabilitation: A qualitative study exploring how practitioner engagement can influence patient engagement," *Clin. Rehabil.*, vol. 31, no. 10, pp. 1396–1405, 2017.

[24] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. D. R. Millán, "The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLoS Biol.*, vol. 16, no. 5, 2018, Art. no. e2003787.

[25] B. Kim and A. D. Deshpande, "An upper-body rehabilitation exoskeleton harmony with an anatomical shoulder mechanism: Design, modeling, control, and performance evaluation," *Int. J. Robot. Res.*, vol. 36, no. 4, pp. 414–435, 2017.

[26] S. Dalla Gasperina, L. Roveda, A. Pedrocchi, F. Braghin, and M. Gandolla, "Review on patient-cooperative control strategies for upper-limb rehabilitation exoskeletons," *Front. Robot. AI*, vol. 8, 2021, Art. no. 745018.

[27] B. Kim and A. D. Deshpande, "Controls for the shoulder mechanism of an upper-body exoskeleton for promoting scapulohumeral rhythm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, 2015, pp. 538–542.

[28] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *Proc. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 629–636.

[29] R. M. Reinhart and J. A. Nguyen, "Working memory revived in older adults by synchronizing rhythmic brain circuits," *Nature Neurosci.*, vol. 22, no. 5, pp. 820–827, 2019.

[30] F. S. Racz, R. Fakhreddine, S. Kumar, and J. D. R. Millán, "Riemannian geometry-based detection of slow cortical potentials during movement preparation," in *Proc. 11th Int. IEEE/EMBS Conf. Neural Eng.*, 2023, pp. 1–5.

[31] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[32] I. Iturrate, L. Montesano, and J. Minguez, "Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces," *J. Neural Eng.*, vol. 10, no. 2, 2013, Art. no. 026024.

[33] F. Iwane, A. Sobolewski, R. Chavarriaga, and J. D. R. Millán, "EEG error-related potentials encode magnitude of errors and individual perceptual thresholds," *iScience*, vol. 26, no. 9, 2023, Art. no. 107524.

[34] F. Iwane, A. Billard, and J. D. R. Millán, "Inferring individual evaluation criteria for reaching trajectories with obstacle avoidance from EEG signals," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 20163.

[35] F. Cincotti et al., "EEG-based brain-computer interface to support post-stroke motor rehabilitation of the upper limb," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 4112–4115.