Towards Closing the Feedback Loop of Human Intention Decoding for Gaze-Driven Assistive Robotics

Yinxuan Li, Bukeikhan Omarali, Jinpei Han, A. Aldo Faisal

***Abstract*— Accurate real-time error detection is critical for effective human-robot interaction (HRI) in assistive robotics, where unintended/erroneous actions can reduce user trust, increase frustration, and potentially cause harm. Conventional error recognition methods rely on external sensors and predefined user commands, limiting their applicability in unstructured environments. This work-in-progress paper proposes an EEG-based error detection framework leveraging error-related potentials (ErrPs) to identify incorrect robotic actions and enable immediate corrective responses. By integrating gaze-based intention recognition with real-time ErrP classification, our approach enhances interaction naturalness and adaptability. We employ a conformer model for online ErrP detection, overcoming the limitations of discrete classification in offline settings. Our experimental setup involves a gaze-driven assistive robotic system, where users perform object manipulation tasks while EEG signals are recorded to detect erroneous actions. Preliminary results suggest that our method improves error recognition accuracy and reduces false positives in real-world scenarios. Future work will focus on refining model robustness, optimizing real-time processing, and expanding applications to more complex assistive tasks.**

I. Introduction

Assistive robotic systems, such as exoskeletons [1] and prosthetic devices [2], offer significant potential as rehabilitation tools for individuals with motor impairments. Upper extremity motor impairments, commonly resulting from spinal cord injuries, amputations, degenerative diseases, or stroke [3], [4], present considerable challenges for affected individuals. These impairments can severely reduce quality of life by restricting the ability to perform essential daily activities.

A critical aspect of successful Human-Robot Interaction (HRI) is accurately recognizing the intention of patients [5]. This is particularly challenging in assistive robotics, given that disabled users have fewer potential options to input their commands. Brain-computer interfaces (BCI) [6], [7] have been successfully applied to control assistive devices with a limited degree of freedom, such as ”drive forward” or ”turn left” for Brain-controlled wheelchairs [8]. However, decoding a larger amount of high-level commands using BCI is challenging. An alternative control method is through the human gaze. The human gaze is an information-rich signal [9] that retains its goal-oriented functionality across most motor impairments and is naturally linked to the user’s action intentions. This enables building ”Zero UI” approaches, where natural gaze is decoded to high-order commands such as ”pouring a bottle” [10]. However, an accurate decoding of human intention from gaze is an extremely challenging task as one first needs to be able to distinguish between ”inspection” and ”intention” gazes [11] and then contextualize the map of the intention gaze given the gazed object, user’s current state, world state and history of actions [10].

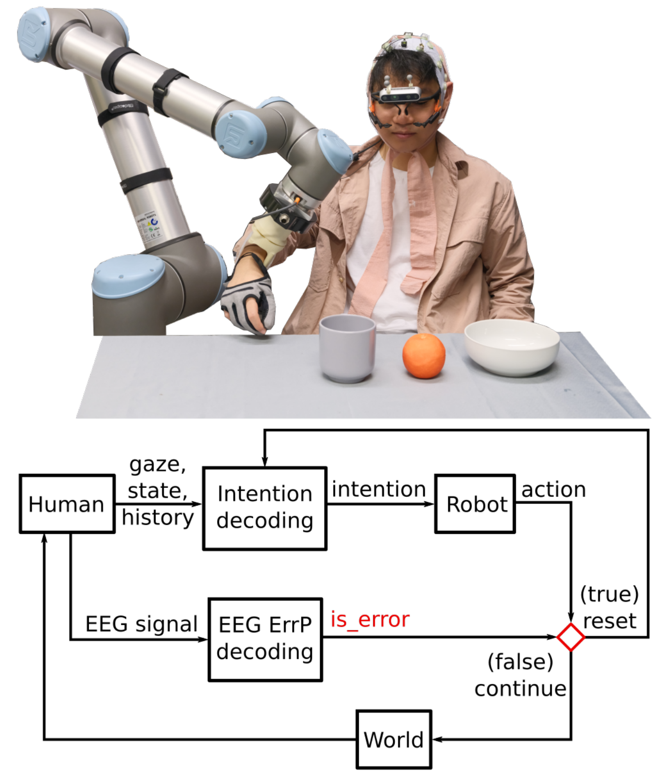

Fig. 1: Proposed EEG-based correction of incorrect actions for assistive robotics. Top: experimental setup - user is attached to an assistive robot, gaze tracking is used to infer the user’s intention, and EEG signals are collected to detect the user’s natural responses to incorrect robot actions. Bottom: A flowchart of the schematic of the proposed method.

In addition to intention recognition, real-time error recognition is critical for effective rehabilitation and adoption in assistive robotics [12]. The ability to detect and correct errors

JH and BO are with the Brain & Behaviour Lab, Department of Computing, Imperial College London; (email:j.han20@imperial.ac.uk). Yinxuan Li is with the Brain & Behaviour Lab, Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom. (email:yinxuan.li21@imperial.ac.uk). A. Aldo Faisal is with the Brain & Behaviour Lab, Department of Computing and Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom. Chair in Digital Health & Data Science, University of Bayreuth, Bayreuth, Germany. (email: aldo.faisal@imperial.ac.uk). We acknowledge funding from the UKRI Turing AI Fellowship to AAF (EP/V025449/1).

in real time enhances safety, improves system effectiveness, and fosters user satisfaction. Conventional HRI systems identify errors by analyzing differences between expected and actual outcomes based on the robot's actions and user feedback. These systems commonly utilize external sensors, including cameras [13], [14], verbal feedback [15], [16], etc., to facilitate error detection. While such approaches perform well in controlled environments where the robot operates within a predefined set of user commands and states, they often rely on pre-defined commands and structured interactions, hence they generalize poorly to unstructured, dynamic scenarios [17] where user's state and feedback may be ambiguous or difficult to interpret [18].

While BCI has limited applications for direct control, it provides a promising approach for real-time error detection by identifying error-related potentials (ErrPs) in users' electroencephalogram (EEG) signals. ErrPs are neural responses that occur when an individual perceives an error, whether self-generated or originating from an external system, such as an assistive robot. Recent studies have explored EEG-based error detection for real-time correction in HRI tasks, such as correcting robot trajectories using ErrPs [19], and improving BCI speller accuracy through double ErrP detection [20]. Unlike traditional methods that rely on predefined responses, EEG-based error detection systems offer direct access to brain activity, enabling adaptive interaction without explicit user input. This seamless and minimally intrusive approach has the potential to improve communication efficiency in assistive robotics [19], [21].

Recent works have shown that deep learning models such as Convolutional Neural Networks [22]–[24], Graph Neural Networks [25] outperformed traditional machine learning models in multiple BCI tasks. However, these approaches rely on pre-defined, discrete time windows that assume the error event occurs around the center. Although effective in offline settings, this assumption limits their ability to generalize to real-life scenarios that require continuous online detection [26]. As demonstrated in [27], such models often exhibit high false positive rates in pseudo-online or online settings, making them unsuitable for reliable error detection in HRI systems.

Data imbalance, feature sparsity, and variable response times across participants [26] pose significant challenges for ErrP detection in real-world HRI scenarios. We previously proposed a conformer model [28] that integrates a pre-trained CNN encoder with a transformer block employing causal attention to capture long-term dependencies in EEG signals. Experimental results validated using publicly available datasets [29] demonstrated that our model outperformed classical discrete methods and sequential deep learning approaches in both inter-session and inter-subject cross-validation tasks, notably achieving substantial reductions in false positive rates. However, validation in controlled datasets does not necessarily guarantee equivalent performance in practical, real-world settings. Therefore, in this study, we aim to demonstrate the robustness and adaptability of our EEG conformer model within an actual assistive robotic environment [30]. By evaluating its performance under real-world operational conditions, we seek to further verify the model's capability, thus substantiating its applicability and effectiveness in enhancing safety and responsiveness in complex human-robot interaction scenarios.

This paper presents our method for real-life deployment of ErrP signal detection for error detection and correction in a gaze-driven assistive robotic setup. Here we present our experimental setup, training data collection pipeline and the ErrP detection algorithm.

II. Methods

A. Experimental setup

Our experimental setup, as shown in Fig. 1, is based on [10], here we provide a brief overview. The user wore a Bioservo Carbonhand soft robotic glove and was securely yet detachably connected to a UR10 robot via an electromagnet and a torque sensor. Eye-tracking data were collected using Pupil Labs' Pupil Core system. An Intel RealSense D435i RGB-D camera and passive optical trackers were mounted on the eye tracking glasses. OptiTrack Flex 13 motion tracking cameras captured the user's head pose.

To facilitate gaze detection and object recognition, the user's gaze was mapped onto the RGB-D camera video stream, with objects in the field of view detected and labeled using Mask R-CNN. A machine-learned classifier determined whether the user's gaze indicated intention or inspection. This gaze information was then passed to an action-grammar controller to select appropriate assistive actions. Robotic motion planning adhered to user comfort constraints, assessed via Rapid Upper Limb Assessment (RULA) [31], ensuring that generated trajectories remained collision-free and ergonomically suitable.

EEG data were recorded continuously during trials using a Brain Product actiChamp amplifier and a Classic 32Ch actiCAP EEG cap. The robot's actions and timestamps at the beginning and end of each action are recorded as event triggers.

B. ErrP Decoding Model and EEG Channel Selection

As intrinsic neural responses of erroneous robotic action, ErrPs have two key components: error-related negativity (ERN) and error positivity (Pe). The ERN typically occurs 50–100 ms after an erroneous response, predominantly over fronto-central scalp regions. This is followed by Pe, which appears between 200–500 ms over centro-parietal areas [32]–[34].

To ensure reliable online ErrP detection, we adapted our state-of-the-art pseudo-online ErrP detection EEG conformer model (see full details of the model [28]) and deployed it in a real-life online testing scenario. As shown in Fig 2, the model concatenates a pre-trained CNN encoder with a causal transformer layer: the CNN encoder captures the local spatial-temporal features from short EEG windows, while the transformer models long-range temporal dependencies across the sequence. Combined with label smoothing and class balancing strategies, this architecture significantly reduces the false positive rate by effectively leveraging both short- and long-range information for more accurate ErrP detection.

To maximize our system's applicability in real-world deployment, we prioritized usability and efficiency over the full-scale setups typically used in EEG experimental research. Instead of relying on the standard 32-channel 10–20 EEG configuration, we limited our setup to eight channels that are most relevant to ErrP detection. As shown in Fig. 2, we selected Fz, FC1, FC2, Cz, C3, C4, CP1, and CP2, with FP1 as the reference and FPz as the ground [33]. These channels were selected to preserve prediction accuracy while significantly reducing the number of electrodes. This reduction not only minimizes setup time and improves user comfort but also lowers hardware and computational costs, thereby enabling more efficient real-time processing for low-latency error correction in practical applications beyond controlled laboratory environments.

C. Error Correction Pipeline

The proposed closed-loop error correction pipeline is illustrated in the flowchart in Fig. 1. After decoding the user's intention based on gaze history and the current robot state, the robot begins executing the corresponding action (e.g., picking up an orange). During execution, the system continuously monitors the output of the online ErrP classifier. If an ErrP is detected, indicating that the user has perceived the action as erroneous, the robot immediately interrupts the task and returns to the previous state prior to initiating the action.

This rollback mechanism allows the user to re-specify their intention via gaze, enabling the robot to reattempt the task. This pipeline facilitates low-latency correction of misinterpreted commands, enabling successful task completion without manual intervention or verbal feedback. Ultimately, it demonstrates the potential of leveraging implicit brain signals for seamless error recovery in assistive human-robot interaction.

III. Experimental Protocol

We are looking to recruit N=20 right-handed healthy participants with no history of neurological disorders.

The experiment consists of two parts: **Data Collection** for ErrP detection model training and **Evaluation with online ErrP detection** based correction.

Data Collection. Participants perform a "pick and place" task. An object (orange) is placed in front of the participant on a table. The participant is asked to indicate intent to grasp the orange using their gaze. Then the robotic arm performs the grasp action. Next, the participant is asked to indicate intent to place the and place the orange onto a predetermined marked position on the table. The robot then places the object. Hence the the user expects to perform the sequence "pick object, place object".

Before the experimental run, participants are guided through 10 practice trials of the "pick and place" task to familiarize themselves with the system. Each experimental run consists of 30 trials. Unknown to the participant, the

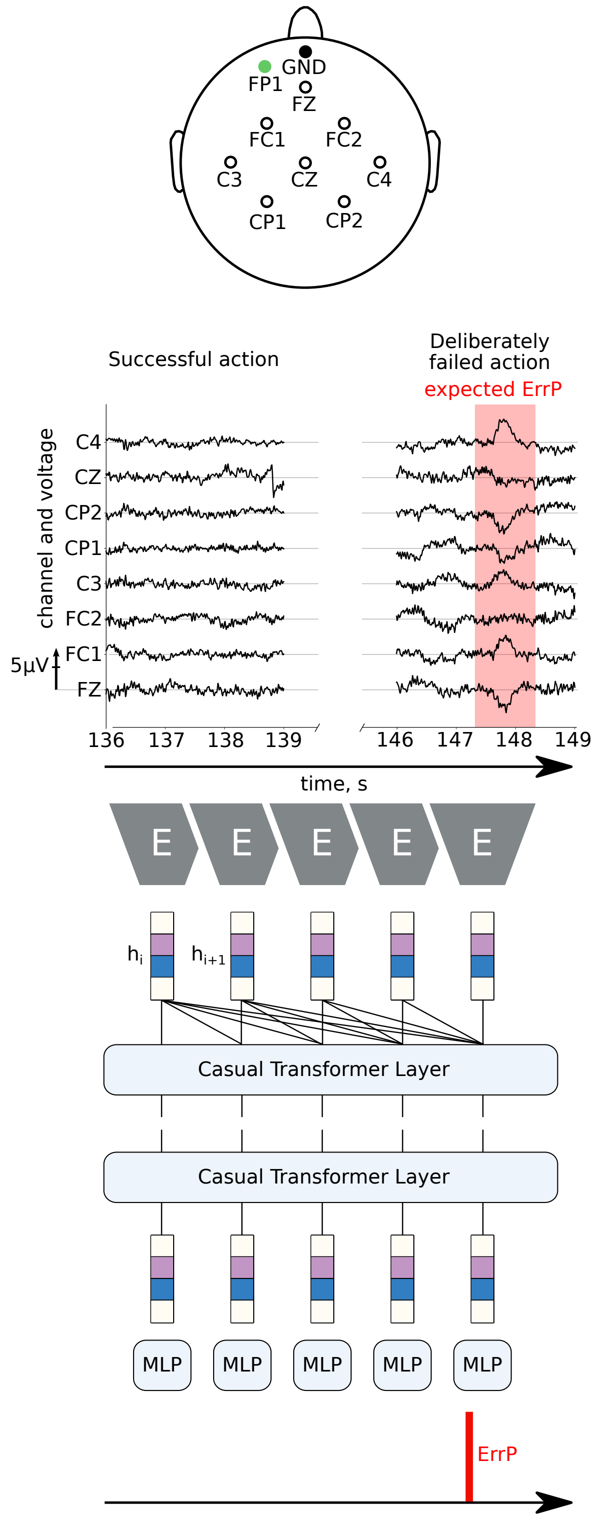

Fig. 2: Sample BCI recordings of a successful action, and deliberately failed action and the Casual EEG Conformer Model. The red shaded area is the expected ErrP signal, E - encoder blocks, h - tokenized feature vector.

robot is programmed to deliberately fails the "pick" action in 20% of trials (i.e. 6 trials) and the "place" action in another 20% of trials (i.e. 6 trials). The sequence of successful and deliberately failed action sequences is randomized across participants. In the "fail pick" scenario, the robot performs the pick action at a 10 cm offset distance from the object - i.e. the robot misses the object. The trial then terminates, and the experimenter resets the scene. In the "fail place" scenario the robot places the object at the exact position it was picked from. Participants are also instructed to press a button whenever they perceive an error made by the robot. This button press generates a trigger in the EEG recording, which is used to label the timing of ErrPs for training the online detection model. After each trial, the experimenter resets the object position.

Sample recordings of successful and deliberately failed actions are shown in Fig. 2 - the red shaded area shows the expected ErrP signal.

Evaluation with online ErrP detection. We evaluate the ErrP detection by performing the "pick and place" task as well as "pick, pour and place". The former is the same as the one described above, with the same deliberate fails, except if a fail is detected by ErrP detection - the robot interrupts the erroneous action and returns to the previous state, such that the participant can indicate the action intention again and perform the action successfully.

We add an extra task to inspect how ErrP detection generalizes between different potential actions and objects. In the "pick, pour and place" task, participants are asked to pick a cup pour it into a bowl and place the cup back on the table at a designated location. Similar to the "pick and place" task each action deliberately failed 20% of the time with the robot interrupting the erroneous action and returning to the previous state, such that the participant can indicate the action intention again and perform the action successfully.

IV. Conclusion and Future Work

This work proposes a real-time EEG-based error detection framework that closes the loop in assistive robotic control by leveraging ErrPs to identify and correct unintended actions. By integrating gaze-based intention recognition with our online EEG conformer model for error detection, our approach enhances user-robot interaction by improving the system safety and fostering user trust, ultimately increasing its reliability and applicability in real-world settings.

Our ongoing work involves data collection from a diverse participant pool to evaluate system performance across individuals. Future directions include refining the ErrP detection model to improve robustness in dynamic environments, optimizing real-time processing for low-latency error correction, and extending our approach to more complex assistive tasks. Additionally, we aim to explore transfer learning techniques to adapt pre-trained ErrP models to individual users, enhancing personalization and generalization.

References

[1] Z. Li, Z. Huang, W. He, and C. Su, "Adaptive impedance control for an upper limb robotic exoskeleton using biological signals," *IEEE Trans. on Industrial Electr.*, 2017.

[2] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O. C. Aszmann, "The extraction of neural information from the surface emg for the control of upper-limb prostheses: Emerging avenues and challenges," *IEEE Trans. on Neur. Sys. and Rehab. Eng.*, 2014.

[3] J. Zariffa, A. Curt, M. C. Verrier, M. G. Fehlings, and S. Kalsi-Ryan, "Predicting task performance from upper extremity impairment measures after cervical spinal cord injury," *Spinal Cord*, 2016.

[4] D. Cattaneo, I. Lamers, R. Bertoni, P. Feys, and J. Jonsdottir, "Participation restriction in people with multiple sclerosis: Prevalence and correlations with cognitive, walking, balance, and upper limb impairments," *Arch. of Phys. Med. and Rehab.*, 2017.

[5] G.-Z. Yang, R. Riener, and P. Dario, "To integrate and to empower: Robots for rehabilitation and assistance," *Sci. Robot.*, 2017.

[6] A. B. Ajiboye, F. R. Willett, D. R. Young, W. D. Memberg, B. A. Murphy, J. P. Miller, B. L. Walter, J. A. Sweet, H. A. Hoyen, M. W. Keith *et al.*, "Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration," *The Lancet*, vol. 389, no. 10081, 2017.

[7] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. Van Der Smagt *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, 2012.

[8] T. Carlson and J. d. R. Millan, "Brain-controlled wheelchairs: a robotic architecture," *IEEE Robotics & Automation Magazine*, vol. 20, no. 1, pp. 65–73, 2013.

[9] W. W. Abbott and A. A. Faisal, "Ultra-low-cost 3D gaze estimation: An intuitive high information throughput compliment to direct brain-machine interfaces," *J. of Neur. Eng.*, vol. 9, no. 4, 2012.

[10] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," *ICRA*, 2019.

[11] P. Festor, A. Shafti, A. Harston, M. Li, P. Orlov, and A. A. Faisal, "Midas: Deep learning human action intention prediction from natural eye movement patterns," *arXiv:2201.09135*, 2022.

[12] N. Lingg and Y. Demiris, "Building trust in assistive robotics: Insights from a real-world mobile navigation experiment," in *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 2023, pp. 1–7.

[13] P. Trung, M. Giuliani, M. Miksch, G. Stollnberger, S. Stadler, N. Mirnig, and M. Tscheligi, "Head and shoulders: automatic error detection in human-robot interaction," in *Proceedings of the 19th ACM international conference on multimodal interaction*, 2017, pp. 181–188.

[14] M. Stiber, R. Taylor, and C.-M. Huang, "Modeling human response to robot errors for timely error detection," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 676–683.

[15] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in bert: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction," in *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2016, pp. 493–500.

[16] E. Rosen, D. Whitney, M. Fishman, D. Ullman, and S. Tellex, "Mixed reality as a bidirectional communication interface for human-robot interaction," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 431–11 438.

[17] M. Awais and D. Henrich, "Human-robot collaboration by intention recognition using probabilistic state machines," *19th International Workshop on Robotics in Alpe-Adria-Danube Region (RAAD 2010)*, pp. 75–80, 2010.

[18] P. Ferrez and J. Millán, "Error-related eeg potentials generated during simulated brain–computer interaction," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 923–929, 2008.

[19] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE ICRA*. IEEE, 2017, pp. 6570–6577.

[20] A. Cruz, G. Pires, and U. J. Nunes, "Double errp detection for automatic error correction in an erp-based bci speller," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 26, no. 1, pp. 26–36, 2017.

[21] S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning–using error-related potentials for real world human-robot interaction," *Scientific reports*, vol. 7, no. 1, pp. 1–16, 2017.

[22] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *JNE*, vol. 15, no. 5, 2018.

[23] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Hum Brain Mapp*, vol. 38, no. 11, pp. 5391–5420, 2017.

[24] J. Han, X. Gu, and B. Lo, "Semi-supervised contrastive learning for generalizable motor imagery eeg classification," in *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2021, pp. 1–4.

[25] J. Han, X. Wei, and A. A. Faisal, "Eeg decoding for datasets with heterogenous electrode configurations using transfer learning graph neural networks," *JNE*, 2023.

[26] X. Gu, F. Deligianni, J. Han, X. Liu, W. Chen, G.-Z. Yang, and B. Lo, "Beyond supervised learning for pervasive healthcare," *IEEE Reviews in Biomedical Engineering*, 2023.

[27] Z. Ren, X. Xia, Y. Tang, B. Zhao, C. P. Wong, and D. Xiao, "Asynchronous detection of erroneous behaviors in human-robot interaction with eeg: A comparative analysis of machine learning models," *bioRxiv*, pp. 2023–09, 2023.

[28] J. Han, Y. Li, X. Gu, and A. A. Faisal, "An eeg conformer model for error feedback during human-robot interaction," in *2025 IEEE ICRA*. IEEE, 2025.

[29] N. Kueper, K. Chari, J. Bütefür, J. Habenicht, T. Rossol, S. K. Kim, M. Tabie, F. Kirchner, and E. A. Kirchner, "Eeg & emg dataset for the detection of errors introduced by an active orthosis device," *Frontiers in Human Neuroscience*, vol. 18, p. 1304311, 2024.

[30] A. Shafti and A. A. Faisal, "Non-invasive cognitive-level human interfacing for the robotic restoration of reaching & grasping," in *2021 10th International IEEE NER*. IEEE, 2021, pp. 872–875.

[31] L. McAtamney and N. Corlett, "Rapid upper limb assessment (rula)," in *Handbook of human factors and ergonomics methods*, 2004.

[32] R. Chavarriaga, A. Sobolewski, and J. d. R. Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces," *Frontiers in Neuroscience*, vol. 8, p. 208, 2014.

[33] L. Wang, Y. Gu, G. Zhao, and A. Chen, "Error-related negativity and error awareness in a go/no-go task," *Scientific Reports*, vol. 10, no. 1, p. 4026, 2020.

[34] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain–computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, p. 011001, 2019.