Accelerating Reinforcement Learning via Error-Related Human Brain Signals

Suzie Kim
Dept. of Artificial Intelligence
Korea University
Seoul, Republic of Korea
sz_kim@korea.ac.kr

Hye-Bin Shin
Dept. of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea
hb_shin@korea.ac.kr

Hyo-Jeong Jang
Dept. of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea
h_j_jang@korea.ac.kr

***Abstract*—In this work, we investigate how implicit neural feedback can accelerate reinforcement learning in complex robotic manipulation settings. While prior electroencephalogram (EEG)-guided reinforcement learning studies have primarily focused on navigation or low-dimensional locomotion tasks, we aim to understand whether such neural evaluative signals can improve policy learning in high-dimensional manipulation tasks involving obstacles and precise end-effector control. We integrate error-related potentials decoded from offline-trained EEG classifiers into reward shaping and systematically evaluate the impact of human-feedback weighting. Experiments on a 7-DoF manipulator in an obstacle-rich reaching environment show that neural feedback accelerates reinforcement learning and, depending on the human-feedback weighting, can yield task success rates that at times exceed those of sparse-reward baselines. Moreover, when applying the best-performing feedback weighting across all subjects, we observe consistent acceleration of reinforcement learning relative to the sparse-reward setting. Furthermore, leave-one-subject-out evaluations confirm that the proposed framework remains robust despite the intrinsic inter-individual variability in EEG decodability. Our findings demonstrate that EEG-based reinforcement learning can scale beyond locomotion tasks and provide a viable pathway for human-aligned manipulation skill acquisition.**

***Index Terms*—brain-computer interfaces, human-robot interaction, reinforcement learning, electroencephalogram;**

I. INTRODUCTION

Reinforcement learning from human feedback (RLHF) has emerged as a promising paradigm for aligning agent behavior with human intention [1]–[3], enabling autonomous agents to learn complex skills without reliance on manually engineered reward functions. Previous studies have explored the integration of various forms of explicit feedback, such as preference labels or corrective demonstrations [4], [5], and implicit feedback, such as body gestures or facial expressions [6]–[8]. Recently, non-invasive neural signals like electroencephalogram (EEG) [9], [10] have emerged as a new form of implicit feedback that can potentially offer a more natural and low-effort communication channel for human–robot collaboration. Among these signals, error-related potentials (ErrPs) [11]—cortical responses elicited when humans perceive mistakes—have demonstrated clear utility as evaluative cues within adaptive brain-computer interfaces (BCI) and robotic control systems [12]. Prior studies have shown that ErrP-based feedback can accelerate learning in navigation or locomotion tasks by providing moment-by-moment assessments of behavioral appropriateness [13]–[15], even when explicit reinforcement is unavailable.

Despite these advances, a critical gap remains in understanding whether neural evaluative feedback can scale to more challenging robotic domains. Existing ErrP-driven RL frameworks have largely operated in low-dimensional settings involving discrete action spaces or simple locomotion behaviors, where the control demands are modest and the role of trajectory optimality is limited [16]. High-DoF manipulation tasks, in contrast, require precise coordination of arm kinematics, continuous control across a multi-joint workspace, and robust handling of obstacles and spatial constraints [17], [18]. Such tasks amplify both the complexity of credit assignment and the consequences of suboptimal exploration, thereby raising fundamental questions regarding the reliability, stability, and utility of neural feedback under rich sensorimotor dynamics.

To address this challenge, we investigate how implicit neural evaluative signals can accelerate reinforcement learning in a 7-degree-of-freedom robotic manipulation task involving obstacle-rich lifting and placement [19]. Our approach builds upon the core insight that ErrP-driven evaluative feedback can modulate exploration and guide policy refinement even when the reward landscape is sparse [20]. The robustness of such neural-feedback systems builds upon foundational progress in machine learning and deep learning methods, including convolutional and recurrent neural networks [21], [22], and various different algorithms that laid the groundwork for modern signal processing and pattern recognition [23]–[26]. By decoding ErrPs using an offline-trained EEGNet [21] classifier and mapping these neural responses into scalar shaping rewards, we enable the agent to leverage human internal evaluations during training without explicit human intervention.

Unlike prior work focused on navigation settings, our study examines how the weighting of neural feedback influences the learning trajectory of a high-dimensional manipulation policy, and whether optimal performance emerges at intermediate

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant, funded by the Korea government (MSIT) (No. RS-2019-II190079, Artificial Intelligence Graduate School Program, Korea University).

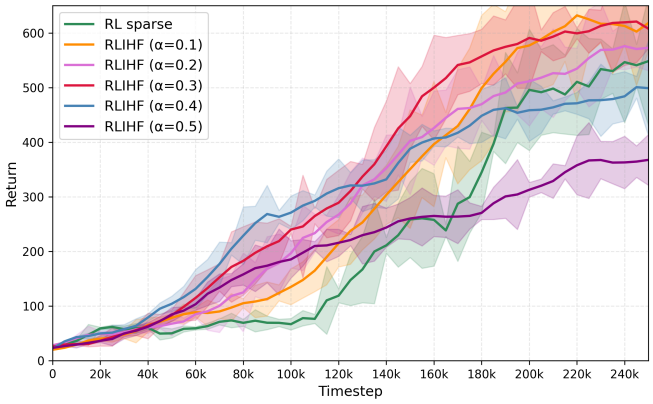

Fig. 1. Effect of human-feedback weighting on learning performance. Episodic return curves for sparse RL and RLIHF across different feedback weights (α). Moderate weights ($\alpha = 0.1 - 0.3$) consistently accelerate learning, while excessively large weights reduce final performance.

feedback strengths. In this study, we systematically vary the human-feedback weight to characterize how feedback strength modulates learning stability. Beyond episodic return metrics, we introduce manipulation-specific evaluation criteria including path efficiency and collision statistics to capture aspects of motion quality and safety that are critical for robotic manipulators. We further conduct subject-level robustness analyzes in twelve participants, each supplying independent EEG-derived evaluative signals. This enables us to quantify inter-individual variability and determine whether implicit neural feedback remains effective even for subjects with relatively low decoder accuracy. Across extensive evaluations, we find that small to moderate feedback weights yield consistent improvements in learning speed, trajectory optimality, and collision reduction. Notably, even though sparse-reward agents eventually learn the task, the addition of implicit neural feedback substantially accelerates their convergence by providing supplementary evaluative signals throughout training. Overall, this study provides new evidence that neuroadaptive reinforcement learning can scale beyond locomotion, offering a principled pathway toward human-aligned robotic skill acquisition in realistic manipulation settings.

II. METHODS

A. Overview

Our framework integrates implicit neural evaluative signals into the reinforcement-learning process of a continuous-control manipulation agent. Building upon this foundation, the agent is trained using a reward function that combines sparse task-level reinforcement with an additional human-feedback term derived from decoded ErrPs. To obtain this feedback, raw EEG activity is processed through a pretrained ErrP decoder, which outputs the probability that the human observer would perceive the agent's current action as erroneous. This probabilistic neural evaluation is then transformed into a scalar reward and integrated with the sparse environment reward, enabling the learning agent to incorporate human internal assessments during policy optimization.

TABLE I
COMPARISON OF SUCCESS RATE, PATH EFFICIENCY, AND MEAN COLLISION BETWEEN SPARSE AND RLIHF REWARD SETTINGS.

Method	Success Rate	Path Eff.	Mean Collision
RL sparse	0.22 ± 0.38	0.64 ± 0.24	2.54 ± 9.49
RLIHF ($\alpha = 0.1$)	0.33 ± 0.43	0.65 ± 0.24	2.98 ± 9.91
RLIHF ($\alpha = 0.2$)	0.30 ± 0.43	0.71 ± 0.24	3.74 ± 15.79
RLIHF ($\alpha = 0.3$)	0.37 ± 0.45	0.69 ± 0.24	2.41 ± 9.03
RLIHF ($\alpha = 0.4$)	0.26 ± 0.42	0.77 ± 0.25	3.40 ± 15.72
RLIHF ($\alpha = 0.5$)	0.14 ± 0.32	0.73 ± 0.27	1.86 ± 15.17

B. ErrP Decoder Pretraining

To construct a subject-specific neural feedback model, we first pretrain an ErrP decoder using EEG data obtained from twelve participants in a human–robot interaction setting. The raw EEG signals are bandpass-filtered in the 1–20 Hz range to isolate slow cortical dynamics associated with event-related responses, downsampled to reduce redundancy, and re-referenced to minimize channelwise bias. Each recording is segmented into fixed-length epochs time-locked to the moment the participant observes the robot's action outcome, capturing the characteristic post-stimulus window in which error-related potentials reliably emerge. These uniformly preprocessed epochs form the input samples used for decoder training.

For each participant, an EEGNet-based convolutional model is trained using a leave-one-subject-out (LOSO) strategy. In this scheme, the decoder for a given subject is trained on the epochs obtained from the remaining eleven participants, ensuring subject-independent generalization at test time. EEGNet's temporal convolutions extract frequency-selective dynamics, while its depthwise spatial filters capture distributed neural patterns associated with error perception. The resulting decoder outputs a continuous probability estimating whether the observed action would elicit an ErrP for that subject. During reinforcement learning, preprocessed epochs are streamed sequentially and passed through the corresponding pretrained decoder, producing temporally aligned neural evaluations that can be directly incorporated into the reward function.

C. Human Feedback Reward Integration

To incorporate neural evaluative signals into the reinforcement-learning process, we map the decoder's output probability to a scalar reward term and blend it with the environment's task-level reward. At each timestep, the pretrained ErrP decoder produces a probability:

$$p_t \in [0, 1], \tag{1}$$

indicating the likelihood that the human observer would judge the agent's most recent action as erroneous. To ensure that uncertain predictions contribute minimally to the reward signal and that highly confident error detections penalize the agent proportionally, we transform this probability using a centered mapping:

$$r_{\text{hf}}(t) = 0.5 - p_t. \tag{2}$$

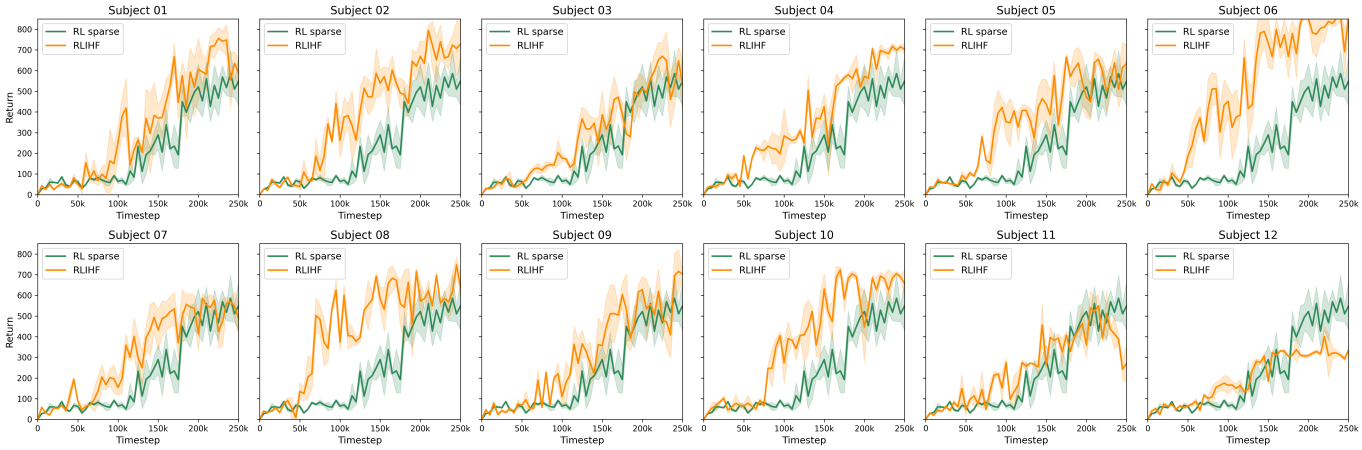

Fig. 2. Cross-subject RLIHF evaluation across 12 participants. For nearly all subjects, RLIHF accelerates learning relative to the sparse-reward baseline, demonstrating robustness to individual variability in EEG decoder accuracy.

This formulation assigns positive reinforcement to actions unlikely to evoke perceived error ($p_t < 0.5$) and negative reinforcement when the decoder detects a high likelihood of error ($p_t > 0.5$).

The final reward supplied to the RL agent combines this neural feedback component with the sparse task-level reward generated by the environment. Let $r_{\text{env}}(t)$ denote the sparse environment reward, which reflects only success events and collision penalties. The total reward used for policy learning is computed as a weighted sum:

$$r_{\text{total}}(t) = r_{\text{env}}(t) + \alpha\, r_{\text{hf}}(t), \tag{3}$$

where $\alpha > 0$ controls the relative influence of neural feedback. This design allows ErrP-derived evaluations to guide policy refinement without overwhelming task-specific reward structure. By integrating neural feedback in this manner, the agent receives dense evaluative signals even when explicit environmental rewards are sparse, thereby accelerating learning while maintaining compatibility with standard off-policy algorithms such as Soft Actor–Critic [27].

III. EXPERIMENTS

A. Experimental Setup

All experiments were carried out in a simulated 7-DoF robotic manipulation environment built on the MuJoCo physics engine and implemented through the robosuite framework [19]. The robot operated within a cluttered tabletop workspace populated with multiple static obstacles, and was required to reach toward a designated object and transport it to a predefined goal region. To support these evaluations, we employed a modified version of robosuite's Lift task, in which additional obstacles and a fixed goal area were incorporated to increase task complexity and to more closely reflect real-world manipulation constraints.

To analyze robustness under human-specific feedback conditions, we evaluated our framework using EEG data obtained from a public HRI-ErrP dataset [11] collected from twelve human participants observing robot actions. Each subject's data was used to train an independent EEG decoder, and identical reinforcement-learning experiments were repeated for all twelve decoders to examine inter-individual variability. All models were implemented in Python, and policies were trained for 250,000 timesteps using identical training and evaluation protocols across all conditions.

B. Effect of Human-Feedback Weighting on Policy Learning

To investigate how strongly implicit neural feedback should influence the agent's reward, we varied the feedback weight α and compared learning performance across conditions. As shown in Fig. 1, low-to-moderate values ($\alpha = 0.1, 0.2, 0.3$) consistently accelerated the growth of episodic return relative to the sparse baseline, demonstrating that neural feedback effectively compensates for insufficient task-level rewards. Among these, $\alpha = 0.3$ achieved the most stable learning curve and the highest final return, suggesting that moderate weighting strikes a balance between providing additional evaluative guidance and maintaining stable policy updates. These effects were further reflected in Table I, where success rate peaked at $\alpha = 0.3$, confirming that well-calibrated human feedback enhances both sample efficiency and task completion.

In contrast, higher weighting values ($\alpha \geq 0.4$) generated a distinct learning pattern: although early gains were observed, performance eventually plateaued or declined during mid-to-late training. This degradation is likely caused by over-amplifying the inherent noise and uncertainty of ErrP-based signals, which can bias the reward toward conservative corrections rather than task-progressing actions. Notably, Table I shows that $\alpha = 0.5$ produced the lowest success rate and reduced return, yet yielded the smallest collision count and lowest collision-occurrence rate. This indicates that an excessively strong influence of neural feedback suppresses risk-taking behavior and shifts the policy toward overly cautious motion planning. While detrimental to overall task performance, this outcome reveals that high-weighted feedback naturally promotes safety-oriented behaviors, suggesting potential applications in risk-sensitive or human-aligned robotic control.

C. Cross-Subject Robustness of Implicit Neural Feedback

To assess whether the performance gains observed with implicit human feedback generalize across individuals with heterogeneous EEG characteristics [28], we conducted a subject-level robustness analysis using the best-performing feedback weight, $\alpha = 0.3$. For each of the twelve participants in the EEG dataset, the corresponding EEGNet decoder was used to generate neural evaluative signals during policy training, and each experiment was repeated across five random seeds. The resulting episodic return curves, shown in Fig. 2, reveal a consistent pattern across nearly all subjects: RLIHF accelerates learning relative to the sparse-reward baseline, yielding faster early-stage convergence and higher returns in the later stages of training. Notably, this trend persists even for participants whose EEG classifiers exhibit only moderate decoding accuracy, indicating that ErrP-derived feedback—despite its noise and subject dependence—still provides a sufficiently informative signal to guide policy improvement. Taken together, these findings confirm that implicit human feedback reliably accelerates reinforcement learning, even under substantial inter-individual variability in neural signal quality.

IV. CONCLUSIONS

This work demonstrates that implicit neural evaluative signals, specifically error-related potentials decoded from EEG, can reliably accelerate reinforcement learning in high-DoF robotic manipulation tasks. By integrating a centered, probability-based human-feedback reward with sparse environmental reinforcement, the proposed framework provides dense evaluative guidance that substantially speeds up policy convergence while preserving stability in continuous-control settings. Our α-sweep analysis revealed that small-to-moderate feedback weights consistently produced faster return growth, improved trajectory optimality, and reduced collisions, confirming that neural feedback acts as an effective catalyst for learning in otherwise sparse and challenging reward landscapes. Moreover, cross-subject evaluations using twelve independently trained decoders showed that this acceleration effect persists despite substantial inter-individual variability in EEG signal quality. Even subjects with moderate decoder accuracy contributed sufficiently informative evaluative cues to guide policy improvement. These findings collectively indicate that implicit neural feedback not only scales to high-dimensional manipulation but also offers a practical pathway toward efficient, human-aligned skill acquisition.

REFERENCES

[1] P. F. Christiano *et al.*, "Deep reinforcement learning from human preferences," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.

[2] K. Lee, S.-A. Kim, J. Choi, and S.-W. Lee, "Deep reinforcement learning in continuous action spaces: A case study in the game of simulated curling," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 2937–2946.

[3] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022.

[4] M. T. Villasevil *et al.*, "Breadcrumbs to the goal: Goal-conditioned exploration from human-in-the-loop feedback," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2023.

[5] K. Lee, L. Smith, and P. Abbeel, "PEBBLE: Feedback-efficient interactive reinforcement learning via relabeling experience and unsupervised pre-training," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021.

[6] H. Fujisawa, H. Sako, Y. Okada, and S.-W. Lee, "Information capturing camera and developmental issues," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, 1999, pp. 205–208.

[7] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain, "Nighttime face recognition at long distance: Cross-distance and cross-spectral matching," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2012, pp. 708–721.

[8] Y. Cui *et al.*, "The EMPATHIC framework for task learning from implicit human feedback," in *Proc. Conf. Robot Learn. (CoRL)*, 2021.

[9] J. Kim *et al.*, "Abstract representations of associated emotions in the human brain," *J. Neurosci.*, vol. 35, no. 14, pp. 5655–5663, 2015.

[10] S. K. Prabhakar, H. Rajaguru, and S.-W. Lee, "A framework for schizophrenia EEG signal classification with nature inspired optimization algorithms," *IEEE Access*, vol. 8, pp. 39 875–39 897, 2020.

[11] S. K. Ehrlich and G. Cheng, "A feasibility study for validating robot actions using EEG-based error-related potentials," *Int. J. Soc. Robot.*, vol. 11, no. 2, pp. 271–283, 2019.

[12] J.-H. Cho, J.-H. Jeong, and S.-W. Lee, "NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13 279–13 292, 2021.

[13] I. Akinola *et al.*, "Accelerated robot learning via human brain signals," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2020.

[14] D.-G. Lee, H.-I. Suk, S.-K. Park, and S.-W. Lee, "Motion influence map for unusual human activity detection and localization in crowded scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 10, pp. 1612–1623, 2015.

[15] M.-C. Roh, H.-K. Shin, and S.-W. Lee, "View-independent human action recognition with volume motion template on single stereo camera," *Pattern Recognit. Lett.*, vol. 31, no. 7, pp. 639–647, 2010.

[16] D. Xu, M. Agarwal, E. Gupta, F. Fekri, and R. Sivakumar, "Accelerating reinforcement learning using EEG-based implicit human feedback," *Neurocomputing*, vol. 460, pp. 139–153, 2021.

[17] S.-W. Lee, *Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007, Proceedings*. Springer Science & Business Media, 2007, vol. 4642.

[18] S.-W. Lee, J. H. Kim, and F. C. Groen, "Translation-, rotation-and scale-invariant recognition of hand-drawn symbols in schematic diagrams," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 4, no. 01, pp. 1–25, 1990.

[19] Y. Zhu *et al.*, "robosuite: A modular simulation framework and benchmark for robot learning," *arXiv preprint arXiv:2009.12293*, 2020.

[20] S. Kim, H.-B. Shin, and S.-W. Lee, "Aligning humans and robots via reinforcement learning from implicit human feedback," *arXiv preprint arXiv:2507.13171*, 2025.

[21] V. J. Lawhern *et al.*, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018.

[22] S.-W. Lee and H.-H. Song, "A new recurrent neural-network architecture for visual pattern recognition," *IEEE Trans. Neural Netw.*, vol. 8, no. 2, pp. 331–340, 1997.

[23] S.-W. Lee, "Multilayer cluster neural network for totally unconstrained handwritten numeral recognition," *Neural Netw.*, vol. 8, no. 5, pp. 783–792, 1995.

[24] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, and B. Benatallah, "Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 32, no. 1, 2018.

[25] S.-W. Lee and S.-Y. Kim, "Integrated segmentation and recognition of handwritten numerals with cascade neural network," *IEEE Trans. Syst., Man, Cybern., C. Appl. Rev.*, vol. 29, no. 2, pp. 285–290, 1999.

[26] S.-W. Lee and A. Verri, *Pattern recognition with support vector machines: First international workshop, SVM 2002, Niagara Falls, Canada, August 10, 2002. Proceedings*. Springer, 2003, vol. 2388.

[27] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018.

[28] H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S.-W. Lee, "Predicting BCI subject performance using probabilistic spatio-temporal filters," *PLoS One*, vol. 9, no. 2, pp. 1–15, 2014.