2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Towards Synergistic Human-Robot Co-Adaptation via Reciprocal Feedback for Shared Contact Tasks

Deniz Yilmaz[1], Shinya Chiyohara[2], Jun-ichiro Furukawa[2,3], Erhan Oztop[1,4], Hiroshi Imamizu[2,5], Jun Morimoto[2,6], and Barkan Ugurlu[1,2]

***Abstract*— In this work, we propose a human-robot physical interaction scheme designed to facilitate contact-rich manipulation tasks. In the proposed framework, neither the robot nor the human agent can complete the task independently, but a shared cost function aligns their efforts and drives them toward success. The robot agent, governed by a reinforcement learning algorithm, can exert forces and modulate its Cartesian impedance while continuously receiving evaluative feedback in the standard RL training paradigm. Simultaneously, the human agent applies forces via a standard PS4 joystick and receives both vibrotactile and visual feedback reflecting task performance. During training, the learning algorithm receives the superposition of its own and the human's actions, allowing it to implicitly benefit from the human's rapidly adapting strategy. We hypothesize that human agents can adapt more rapidly than RL and, when provided with feedback grounded in real measurements, can make more quantifiable decisions. During this rapid human adaptation phase, the robot concurrently acquires skills from the human, thereby accelerating training and improving overall efficiency. The proposed interaction scheme was evaluated in a realistic simulation environment involving 10 participants. Preliminary results indicate that participants receiving vibrotactile feedback adapted more quickly, enabling the robot to acquire the desired skill in only a few episodes for simple tasks. For more challenging tasks, human-trained RL agents required additional autonomous training, yet still achieved convergence far faster than PPO-only training. This co-adaptive framework combines the complementary strengths of humans and robots, providing a versatile foundation for contact-rich manipulation that may be extended to diverse tasks and robotic platforms.**

I. Introduction

Contact-rich manipulation tasks, such as polishing or sanding, require precise regulation of interaction forces to ensure accuracy, safety, and repeatability [1]. In robotics, these tasks are challenging because the robot must maintain stable contact forces even when surface properties or environmental conditions change. Fully autonomous systems often struggle to adapt to unexpected variations due to limited generalization beyond their training data and slower adaptation to novel conditions [2], while purely human-operated methods can lead to inconsistent results and even operator fatigue [3]. These limitations have motivated the development of hybrid strategies that combine the strengths of both humans and robots.

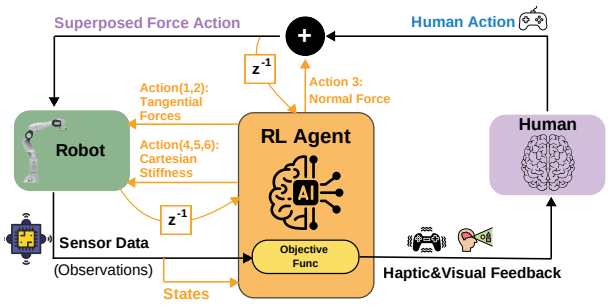

Fig. 1. Proposed framework for human–RL collaborative force regulation, with color-coded blocks for baseline impedance control (green), PPO-based RL agent (orange), and human input via joystick (purple). The unit delay is denoted via z^{-1} to indicate previous actions. Depending on the task, some of the actions may be fixed.

Human-in-the-loop (HITL) learning addresses these challenges by enabling humans and robots to collaboratively acquire skills through ongoing interaction [4]. In such systems, humans accelerate robot learning by providing demonstrations or corrective actions, while the robot's autonomy increases as its control policy improves. HITL has demonstrated strong results in various application areas, including shared control, teleoperation, and assistive robotics [5].

In human–robot collaboration, adaptation is mutual: as the robot's control policy improves, the human refines her/his strategy in response [6]. Reciprocal feedback enables both agents to adjust actions in real time, supporting co-adaptation and improving performance [7]. In physical human–robot interaction (pHRI), this may lead to smoother cooperation and greater efficiency. Yet, existing human-in-the-loop approaches typically provide only visual feedback to the human and treat the human agent's actions as the ground truth [8]. This one-way adaptation often limits the human's ability to refine their strategy in response to the robot's changing behavior, potentially slowing down joint learning and reducing overall task efficiency.

Addressing this gap, the contribution of this paper is a co-adaptive pHRI framework in which neither the human nor the robot can initially complete the task alone, yet together

[1]Faculty of Engineering, Ozyegin University, 34794 Cekmekoy, Istanbul, Türkiye. deniz.yilmaz.25438@ozu.edu.tr, {barkan.ugurlu,erhan.oztop}@ozyegin.edu.tr,
[2]Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International (ATR), 619-0288 Kyoto, Japan. s.chiyohara@atr.jp, furukawa@atr.jp, xmorimo@atr.jp
[3]Faculty of Systems Engineering, Wakayama University, 640-8510 Wakayama, Japan.
[4]OTRI, SISREC, Osaka University, 565-0871 Osaka, Japan. erhan.oztop@otri.osaka-u.ac.jp
[5]Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, 113-0033 Tokyo, Japan. imamizu@l.u-tokyo.ac.jp
[6]Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan.

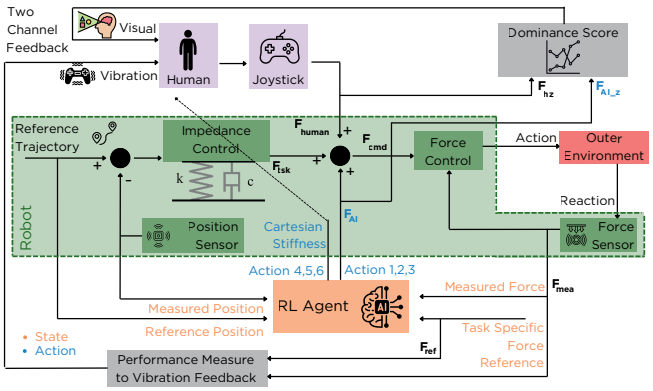

Fig. 2. System architecture showing sensor data flow, human/RL actions, objective functions, and real-time feedback. Action-1: force; Action-2: stiffness. For simplicity, the force control block encompasses joint-level force control, sensing, and task-space force estimation as a single unit. Depending on the task, some of the actions may be fixed.

they achieve rapid, stable, and efficient skill acquisition. The robot learns a force–stiffness policy and receives feedback as a reward for reinforcement learning, while the human simultaneously applies corrective force through a joystick and receives both visual and vibrotactile feedback based on task performance. A distinctive element is the integration of real-time vibrotactile feedback [9], complementing visual cues to support informed and precise human adjustments guided by measured task performance. By superimposing the human agent's previous force action with the RL agent's action, the robot leverages rapid human adaptation to accelerate its own learning. For simpler tasks, comparable performance to fully autonomous training was reached within only 10–15 episodes, while for more challenging tasks, human-trained agents still required further training but converged about $4x$ faster than PPO-only training. This yields an efficient co-adaptation scheme for contact-rich manipulation.

The remainder of this paper is as follows. The proposed method is disclosed in section II. The results are presented and discussed in section III. The paper is concluded in IV.

II. Methods

A. Problem Statement & General Overview

Contact-rich manipulation tasks require high fidelity force control, skillful adaptation to the environment, and robot platforms. Achieving high performance in such tasks demands experts, both to program the robot for the specific hardware and to fine-tune control strategies for the task at hand. This limits scalability and makes it challenging to deploy solutions across diverse human operators and robotic systems. To address this limitation, we propose a human–robot co-adaptation framework in which neither agent can complete the task alone, but both improve through interaction: the human is guided by real-measurement-based vibrotactile and visual feedback, while the robot learns in parallel via reinforcement learning from combined actions. This enables rapid human adaptation, efficient skill transfer, and deployment across varied operators and robotic platforms without expert intervention.

The proposed framework integrates human input and an RL agent within a unified architecture; see Fig. 1. The human provides force input exclusively along the z axis via a joystick, whereas the RL agent generates six actions: three regulating **stiffness** and three regulating **force** along the x, y, and z axes. Depending on the task, some of these actions may be fixed. The final force command is obtained component-wise: $F^{x,y}_{cmd} = F^{x,y}_{tsk} + F^{x,y}_{\text{AI}}$ and $F^{z}_{cmd} = F^{z}_{tsk} + F^{z}_{\text{AI}} + F_{\text{human}}$, showing that the human input contributes only to the z direction. The Cartesian stiffness is exclusively regulated by the RL agent, which also considers its *previous actions* (z^{-1}) to promote consistent behavior. Real-time visual and vibrotactile feedback guide the user during training, and a dominance metric tracks the evolving balance of control between the human and RL, as detailed in Fig. 2.

B. Baseline Controller: Impedance Control

To achieve trajectory tracking while regulating the interaction force, we employed an operational-space impedance controller as the baseline method [10]. In this approach, the task-space torque command τ_{tsk} is computed as below:

$$\tau_{tsk} = J(q)^{\intercal} F_{tsk} \quad (1)$$

$$F_{tsk} = \Lambda(x)\left(\ddot{x}_{ref} + K_p e_x + K_d \dot{e}_x\right) + \eta(x, \dot{x}) \quad (2)$$

where x and q represent task space and joint space variables. The underscript ref denotes reference values. Trajectory tracking error is e_x and computed as: $e_x = x_{ref} - x$. $J(q)$ is the end-effector Jacobian matrix. The diagonal matrices K_p and K_d store Cartesian stiffness and damping coefficients. The z-axis Cartesian stiffness was assigned by the RL agent, while the remaining axes were set to relatively high values to ensure accurate trajectory tracking. The damping coefficients were chosen to achieve critically damped behavior consistent with the selected stiffness values. $\Lambda(x)$ and $\eta(x, \dot{x})$ stand for the inertia matrix and nonlinear terms in operational task space, respectively. They can be computed as follows:

$$\Lambda(x) = \left(J(q)M(q)^{-1}J(q)^{\intercal}\right)^{-1} \quad (3)$$

$$\eta(x, \dot{x}) = \Lambda(x)\left(J(q)M(q)^{-1}C(q, \dot{q}) - \dot{J}(q)\dot{q}\right) \quad (4)$$

In (3)-(4), $M(q)$ is the joint space inertia matrix. $C(q, \dot{q})$ is the term that stands for Coriolis and centrifugal terms in the joint space. The final torque command, τ_{cmd} is synthesized as follows:

$$\tau_{cmd} = \tau_{tsk} + \tau_g + N_p \tau_{nl} \quad (5)$$

where τ_g represents the gravity compensation torque values. The term τ_{nl} stands for null-space torques and N_p is the dynamically consistent null-space projector:

$$N_p = \mathbb{1} - J(q)^{\intercal}\Lambda(x)J(q)M(q)^{-1} \quad (6)$$

Null-space torques are designed as in the following:

$$\tau_{nl} = -k_1 M(q)^{-1}\,\text{diag}\,\frac{1}{(\tau_i^{max})^2}\tau_{tsk} - k_2\dot{q} + k_3\nabla_{\mathbf{q}}\xi(q) - k_4\nabla_{\mathbf{q}}J_{cnt}(q) - k_5\nabla_{\mathbf{q}}J_{cnv}(q) \quad (7)$$

where we minimize task torques and joint velocities, maximize manipulability, respect joint limits and penalize elbow-down configuration, thereby ensuring safe, efficient, and well-coordinated motion during task execution. In (7), τ_i^{max} is the maximum torque of the i^{th} joint. The gradients $\nabla_{\mathbf{q}}\xi(q)$ is $\nabla_{\mathbf{q}}J_{cnt}(q)$, and $\nabla_{\mathbf{q}}J_{cnv}(q)$ correspond to manipulability, joint limitations, and elbow-up configuration, respectively.

C. AI Agent: Reinforcement Learning Architecture

For autonomous skill acquisition, we chose the Proximal Policy Optimization (PPO) RL algorithm for its robustness in continuous action spaces, stability through clipped policy updates, and competitive sample efficiency compared to other on-policy methods [11]. The learning setup consists of an observation space, an action space, and a reward function.

The observation space was chosen as a 33-dimensional state vector including end-effector pose, 6D twist, estimated joint torques, contact wrench, and target states from the reference trajectory, enabling the agent to have information on both the robot's state and its interaction with the environment.

The action space depends on the task: in the simplest cases it comprises two continuous outputs (a force command in $[0, 300]$ N and a Cartesian stiffness along the z axis in $[50, 150]$ N/m), whereas in the full setting it expands to six dimensions (three forces and three stiffness components along the x, y, and z axes). In standard PPO, the environment transitions as $s_{t+1} \sim P(s_{t+1} \mid s_t, a_t^{\text{AI}})$, where a_t^{AI} is the RL-generated action. In the proposed method, the applied action is obtained as the superposition $\tilde{a}_t = a_t^{\text{human}} + a_t^{\text{AI}}$, and this combined action is fed back to the PPO agent for policy update. PPO evaluates these updates using an *advantage* estimate $\hat{A}_t$, which measures how much better an action is compared to the policy's average at a given state. The human contribution introduces a beneficial bias in this estimate, shifting transitions toward positive rewards earlier in training, reducing variance in $\hat{A}_t$, enhancing the clarity of policy gradient updates, and thereby enabling faster, more stable learning [12].

The reward function, computed at every timestep, consists of five components: $r = \sum_{i=1}^{5} r_i$. Error-based terms use the Huber loss [13], which balances precision and robustness by applying quadratic penalties for small deviations and linear penalties for large ones. For all reward terms, $\mathcal{L}$ denotes the Huber loss and w the corresponding weight.

The first term penalizes the deviation between the reference and measured contact forces along the z-axis:

$$r_1 = w_1 \cdot \mathcal{L}_{\delta_1}(F_{\text{zref}} - F_{\text{z}}) \tag{8}$$

The second term penalizes the deviation between the reference and measured tool position along the z-axis:

$$r_2 = w_2 \cdot \mathcal{L}_{\delta_2}(z_{\text{ref}} - z_{\text{mea}}) \tag{9}$$

The third term penalizes rapid fluctuations in the contact force to promote smoother interactions and reduce instability:

$$r_3 = w_3 \cdot \mathcal{L}_{\delta_3}(\dot{F}_z) \tag{10}$$

The fourth term penalizes velocity along the z-axis to ensure smooth motion:

$$r_4 = w_4 \cdot \mathcal{L}_{\delta_4}(\dot{z}) \tag{11}$$

Finally, the fifth term encourages stable interaction by rewarding contact forces within a target range, modeled as a Gaussian:

$$r_5 = w_5 \cdot \mathcal{N}(F_{\text{mea}} \mid \mu, \sigma^2) \tag{12}$$

where μ is the midpoint and σ controls the range of the desired force interval. The weights from w_1 to w_5 were set based on the task goals and system behavior, giving more importance to force tracking and stability, and keeping other penalties moderate to allow sufficient exploration.

D. Two-Channel Feedback to Human

To support effective human–robot collaboration and facilitate user adaptation during training, the system provides two feedback channels: vibrotactile feedback through joystick vibration and a real-time visual display.

1) Channel-1: Vibrotactile Feedback: This is delivered through joystick vibration, with intensity inversely proportional to the force error magnitude. As the error decreases, the vibration becomes stronger, giving an intuitive sense of how far they are from the target force. To map the force error into a continuous vibration signal, a double-layered feedback function was designed.

The double-layered Gaussian feedback function uses the force error as input and combines two Gaussian-shaped curves: a wide component for gradual feedback across the full range and a narrow component for sharper feedback near the target. The final feedback signal is the weighted sum of these curves:

$$\text{Feedback}(e) = a \cdot \mathcal{N}(e \mid \mu, \sigma_w^2) + b \cdot \mathcal{N}(e \mid \mu, \sigma_n^2) \tag{13}$$

where $\mathcal{N}(e \mid \mu, \sigma^2)$ is the Gaussian function evaluated at the force error e with mean $\mu = 0$. Weights $a = 0.75$ and $b = 1.0$ control the contribution of the wide and narrow curves, and standard deviations $\sigma_w = 12$ and $\sigma_n = 6$ set their widths.

2) Channel-2: Visual Feedback: To assess the relative control contributions of the human and robot agents during task execution, we compute a dominance score from the root-mean-square (RMS) magnitudes of their applied force signals, as defined by:

$$C = \frac{\text{RMS}(F_{\text{AI}})}{\text{RMS}(F_{\text{AI}}) + \text{RMS}(F_{\text{human}})} \tag{14}$$

Here, $\text{RMS}(F)$ denotes the root-mean-square magnitude of the force signal, and C represents the dominance score. The score ranges from 0 (full human control) to 1 (full robot control) and provides a continuous measure of which agent contributes more to the total force applied to the environment. Monitoring this score across episodes reveals the gradual shift of control to the robot during learning.

The visual feedback included a real-time simulation of the robot and its environment, allowing users to observe

task execution and assess system response. In the simulation, the contact point with the surface is indicated by a red sphere. Human action input and the real-time confidence score were displayed separately to indicate the current shared control state. This simulation environment is implemented using RaiSim [14], a cross-platform multi-body physics engine providing accurate contact dynamics and real-time visualization.

This framework enables the human to adapt quickly by relying on intuitive, real-time feedback on task performance. Meanwhile, the robot continuously observes human behavior and updates its policy via RL. As training progresses, the robot increasingly contributes to task execution and eventually becomes capable of performing the task autonomously. This transition is captured by the dominance score, which gradually increases and approaches one as the robot has full control, under the assumption that human inputs remain within reasonable bounds and are generally aligned with the task goals.

E. Simulation Experiment Protocol

We conducted a simulation-based human–robot interaction experiment with 10 volunteer participants (6 male, 4 female), aged 23–36. The experimental protocol was approved by the institutional ethics committee of Ozyegin University, and all participants provided written informed consent prior to participation.

Before the experiment, participants received a training session where the task objectives, the joystick-based control interface, and the system's behavior were explained. Participants were also shown demonstrations of the feedback mechanisms. A short pre-experiment questionnaire was then administered to assess prior experience, including previous use of joysticks or interaction with robots.

Participants performed two tasks in the 2-action setting, both aimed at helping the robot reach a predefined contact force with human assistance. They were able to apply vertical force (F_{human}) via the right analog stick (0–100 N range). The experiment took place entirely in a simulated environment, allowing participants to observe the robot and its interaction with the environment in real time.

In Task-1, the robot remained fixed with a constant reference force. In Task-2, the robot followed a varying trajectory along the y-axis, and participants assisted by applying force along the z-axis to maintain the target contact force. Task-1 was limited to a maximum of 10 episodes,Task-2 to 15 episodes and tas, where each episode refers to a single trial of the task. These limits were based on preliminary tests showing they were sufficient for the robot to learn the task with human guidance. Although each task had a predefined episode limit, participants were allowed to end early if they believed the robot could perform the task autonomously.

After completing Task-2, participants filled out a second questionnaire evaluating task difficulty, sense of control, and quality of interaction with the robot. They were also invited to provide open-ended feedback on the task structure, feedback modalities, and potential improvements.

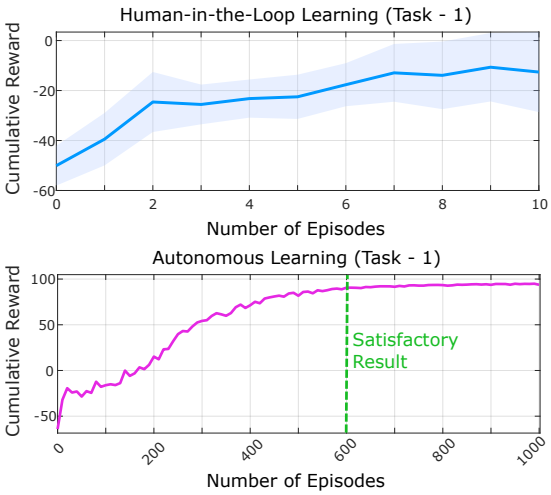

Fig. 3. Reward progression for Task-1, comparing the proposed method with autonomous PPO learning.

In addition to these two tasks, we introduced a more challenging scenario. Task-3 required the robot to follow a circular trajectory in a six-dimensional action space, which included Cartesian stiffness along the x and y axes as well as additional force components. This task was conducted with the best and worst performing participants from the earlier experiments, following the same procedure.

III. Results and Discussion

To evaluate the framework, we defined a success criterion for models learned via human guidance: a participant was considered successful if the mean absolute error between measured and reference contact force stayed within $\pm 12\%$ of the reference throughout a test episode. This threshold reflects realistic tolerances in physical human–robot interaction, accounting for human variability and acceptable robotic control deviation. Success rates quantified human teaching performance across both tasks.

For proof of concept, Task-1 involves static force regulation, while Task-2 adds end-effector motion along the y-axis, introducing greater interaction dynamics. Task-3 represents a substantially more challenging scenario, requiring force regulation along a circular trajectory in an expanded six-dimensional action space. The pre-defined human episode limit was determined empirically based on pre-test results, considering the dimensionality of the action space and the inherent properties of the task. This progression of tasks enables analysis of how human guidance and shared control perform under increasingly demanding conditions.

A. Task 1: Learning a Static Contact Force

Fig. 3 compares learning curves of the proposed method and autonomous PPO training under two-channel feedback for Task-1. In our case, participants had a maximum of 10 episodes, which is sufficient to guide the robot to the target behavior, whereas PPO-only training, with the chosen hyperparameters, required nearly 600 episodes for similar performance. The lower reward in the proposed method

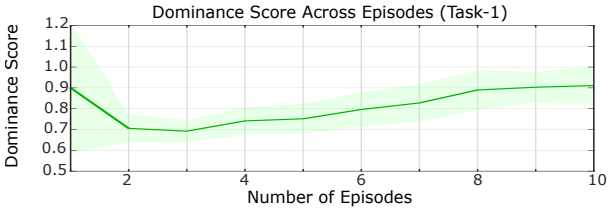

Fig. 4. Progression of dominance score over training episodes in Task-1, illustrating changes in control contribution between human and robot.

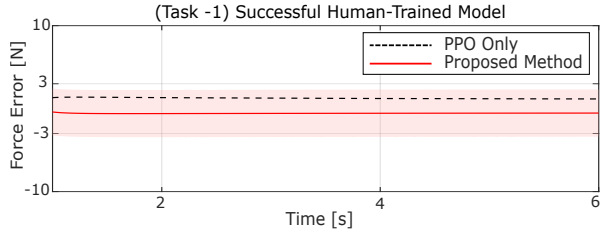

Fig. 5. Force error over time during a representative task execution in Task-1, comparing PPO-only (black dashed line) and the proposed method.

stems from stability-related penalties in the reward function rather than inferior performance, as human inputs avoid these penalties while the PPO agent initially incurs them before stabilizing. Hence, it gradually accumulates reward over more episodes through the Gaussian force-tracking term; see eq. (12). These results show how human guidance accelerates training compared to autonomous learning.

Fig. 4 shows the change in dominance scores, where 1 denotes full robot control and 0 full human control. Across participants, the score generally increased over episodes, reflecting a gradual shift of control to the robot as the policy improved. Early episodes exhibited greater variation due to differences in teaching strategies and applied force. As training progressed, scores converged at higher values, indicating reduced human input. This trend matches the episode–reward results, confirming that guidance decreased as the autonomous policy converged.

Fig. 5 compares the test performance of PPO-only and the proposed method for Task-1. While both approaches achieved low mean force error, the proposed method reached slightly better performance with far fewer training episodes. In PPO-only training, the stiffness and force actions were $62.04\ \mathrm{N/m}$ and $-182.98\ \mathrm{N}$, respectively. With the proposed method, the average values across 10 datasets were $103.10 \pm 1.61\ \mathrm{N/m}$ and $-262.67 \pm 2.18\ \mathrm{N}$. While the proposed method converged rapidly to high performance, its stiffness modulation remained suboptimal, resulting in a relatively stiffer behavior that required greater force input. Nevertheless, variation across participants was modest, as shown by the shaded areas, indicating most participants could guide the robot to stable control.

B. Task 2: Learning a Contact Force Along a Varying Trajectory

In this task, participants helped the robot maintain target contact force while the end-effector followed a varying y-axis trajectory. Unlike Task-1, it added dynamic complexity, requiring continuous input adaptation. Using the success criterion defined earlier, only six participants met the requirement; their results are analyzed below.

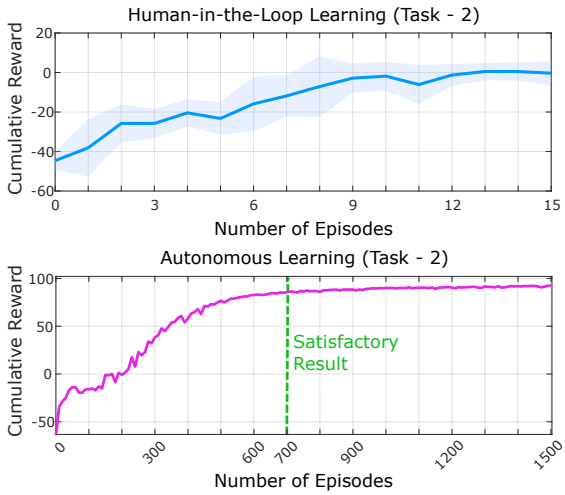

Fig. 6. Reward progression for Task-2, comparing the proposed method with autonomous PPO learning.

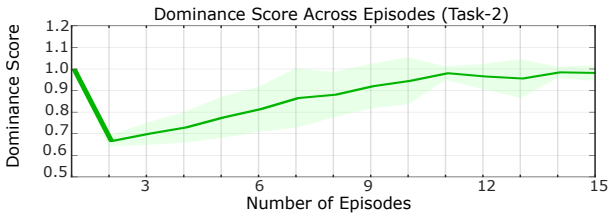

Fig. 7. Progression of dominance score over training episodes in Task-2, illustrating changes in control contribution between human and robot.

Fig. 6 shows episode–reward curves for Task-2. Learning was slower than in Task-1 due to the added complexity of maintaining target force during end-effector motion, yet participants still guided the robot to good performance much faster than PPO-only training. Similar to Task-1, dominance scores increased over episodes, with a slower rise at first due to the added challenge, as shown in Fig. 7. Variation between participants was higher early on, reflecting differing adaptation speeds. By the end of training, most reached high scores, indicating the robot had largely taken over control.

Fig. 8 compares the test performance of PPO-only and the proposed method for Task-2. This figure shows that the applied force in both approaches closely followed the target. PPO-only achieved lower average error, reflecting longer training, whereas the proposed method reached comparable levels within only 15 episodes. In the proposed method, the shaded areas indicate moderate variability, suggesting that most participants were able to guide the robot to stable performance even with dynamic complexity. In PPO-only training, stiffness and force actions were $72.04\ \mathrm{N/m}$ and $-195.68\ \mathrm{N}$. With the proposed method, averages across 6 datasets were $100.77 \pm 1.01\ \mathrm{N/m}$ and $-268.10 \pm 1.30\ \mathrm{N}$.

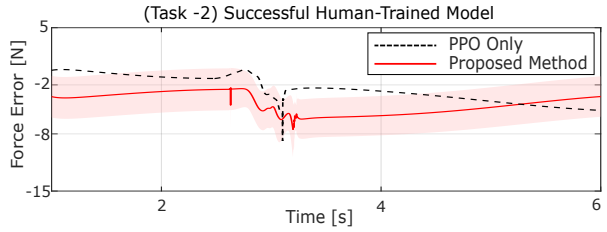

Fig. 8. Force error over time during a representative task execution in Task-2, comparing PPO-only (black dashed line) and the proposed method.

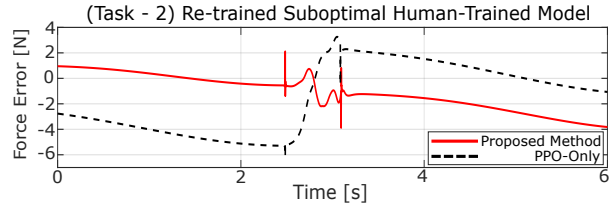

Fig. 9. Force error over time during a representative Task-2 execution, comparing PPO-only (black dashed line) and the proposed fine-tuned model (red).

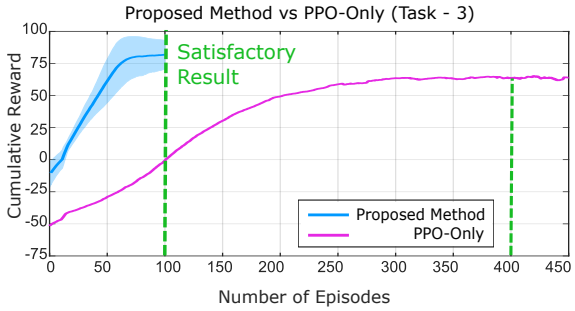

Fig. 10. Reward progression for Task-3, comparing PPO-only with the proposed human-initialized fine-tuning.

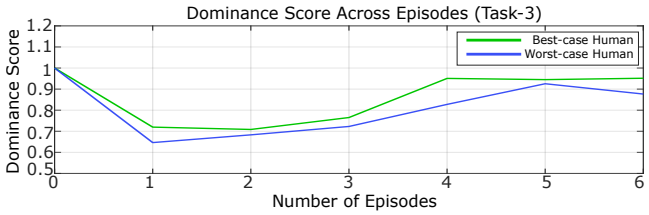

Fig. 11. Progression of dominance score over training episodes in Task-3, illustrating changes in control contribution between human and robot for best(green line) and worst(blue line) human cases.

The method converged rapidly but produced stiffer behavior requiring greater force input. Participant variation was modest, indicating most could guide the robot to stable control.

C. Re-training Suboptimal Human-Trained Models

To improve suboptimal human-trained models in Task-2, we selected the worst performer among the four participants who failed the success criterion and continued training that model autonomously for 150 additional episodes. The resulting policy not only recovered but achieved lower test force error than the PPO-only baseline, as shown in Fig. 9. These results suggest a hybrid strategy in which human guidance rapidly shapes core behavior, even if the initial performance is weak. Subsequent autonomous training can then refine the policy to surpass PPO-only accuracy, requiring far fewer total episodes—e.g., 165 versus 700 in this case.

D. Task 3: Learning a Contact Force Along a Circular Trajectory

In this task, the robot followed a circular trajectory in an expanded six-dimensional action space, including Cartesian stiffness in the x and y axes and additional tangential force componentss. Pre-tests confirmed that the richer action set facilitated faster PPO adaptation, with convergence achieved in roughly 400 episodes instead of 600–700 in the earlier tasks. Consequently, human guidance was required for fewer episodes, and the human episode limit was reduced.

Fig. 10 shows the episode–reward curves for Task-3. With human guidance, the policy reached a satisfactory performance level within about 100 fine-tuning episodes, as indicated by the first vertical line, following the initial human training of 6 episodes. In contrast, PPO-only training required nearly 400 episodes to reach a comparable reward, marked by the second vertical line. Although both approaches eventually converged to similar cumulative reward, incorporating human input substantially accelerated the learning process. Among the human cases, the best participant reached convergence earlier than the worst, highlighting the effect of demonstration quality on adaptation speed.

Fig. 11 shows the dominance score progression for the best-case and worst-case human participants in Task-3. In both cases, the score initially decreased before gradually increasing as control shifted from the human to the robot. The best-case participant maintained higher dominance values throughout, reflecting smoother guidance and earlier convergence, while the worst-case participant exhibited a slower rise. These results highlight that participant quality influenced the speed of adaptation, although both eventually reached comparable levels of robot dominance.

Fig. 12 compares the force-tracking performance of the fine-tuned proposed method and PPO-only in Task-3. Both maintained errors within a similar range, though PPO-only exhibited slightly lower mean error in the steady state. Nevertheless, despite the six-dimensional action space where the human agents could only intervene through force input along the vertical axis, the proposed method enabled much faster training, reaching satisfactory performance in far fewer episodes. This highlights that even limited human input was sufficient to accelerate learning, final tracking accuracy of the two approaches was comparable.

E. Participant Perception and Adaptation Analysis

Post-experiment questionnaires were analyzed using grouped Likert-scale data, summarizing participant feedback, adaptation, and sense of agency. A total of 70% of participants agreed that vibrotactile feedback was intuitive, 90% found it helpful for improving performance, and 80% reported increased engagement when it was used. These

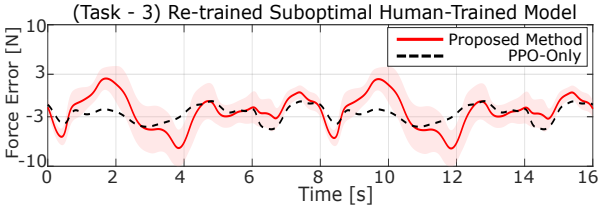

Fig. 12. Force error over time during a representative task execution in Task-3, comparing PPO-only (black dashed line) and the proposed method.

findings indicate that vibrotactile feedback provided more understandable guidance during force regulation, enhancing user involvement and facilitating faster adaptation.

When human adaptation was analyzed, 90% of participants felt more comfortable with the task over time, and all learned to use the joystick more effectively. Furthermore, 90% better understood the robot's reactions and adapted their behavior based on vibration feedback. These results suggest that rapid human adaptation not only improved immediate task performance but also smoothed the transfer of control to the robot during learning.

Sense of agency (SoA) in shared control is the user's perception of controlling the system's actions and resulting outcomes [15]. On average, 67% of participants agreed with SoA- and trust-related statements, indicating a generally positive perception of control and trust in the robot. Agreement rates were 90% for satisfaction with its responses, 80% for collaboration being intuitive, and 70% for feeling their input contributed to success. Only 20% of participants reported feeling in control for most of the time. This suggests that, as the robot's policy improved, participants progressively handed over more direct control to the robot. Instead, they shifted their focus toward a supervisory role, overseeing the interaction and intervening when necessary. Along with the shift in dominance toward the robot, this shows that SoA changed from direct control to a supervisory role, helped by increased trust and shared task goals.

IV. Conclusion

In this work, we proposed a shared-control framework that enabled rapid human–robot co-adaptation for contact tasks. Experiments across three scenarios demonstrated that human guidance, supported by vibrotactile and visual feedback, allowed the robot to acquire skills within 10–15 episodes for simple force-regulation tasks, whereas PPO-only training required several hundred episodes. In more challenging tasks with an expanded six-dimensional action space, human involvement accelerated learning but an additional 100 autonomous episodes were still necessary to achieve better performance. Even then, the learning process remained about 4× faster than PPO-only training. The gradual shift in dominance from human to robot reflected a smooth transfer of control, maintaining human oversight while benefiting from the robot's long-term adaptation capacity. Overall, the findings indicate both the efficiency and scalability of the proposed framework, and demonstrate its potential as a strategy for accelerating reinforcement learning based human–robot co-adaptation.

The current validation was limited to simulation with single-axis human input, chosen to reduce cognitive complexity and emphasize the dominant control dimension. This simplification constrains the generality of the approach for multi-DoF and real-world contact scenarios, which will be addressed in future work through physical experiments and extended shared-control implementations.

Acknowledgement

This work was supported by the Japan Trust International Research Cooperation Program of the National Institute of Information and Communications Technology (NICT), and by JSPS KAKENHI, Grant Number JP23K24925. The authors thank Ibrahim Burak Ozkaynak.

References

[1] L. Peternel and A. Ajoudani, "After a decade of teleimpedance: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, p. 401–416, Apr. 2023.

[2] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[3] C. Brosque, E. Galbally, O. Khatib, and M. Fischer, "Human-robot collaboration in construction: Opportunities and challenges," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 2020, pp. 1–8.

[4] B. Luo, Z. Wu, F. Zhou, and B.-C. Wang, "Human-in-the-loop reinforcement learning in continuous-action space," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2023.

[5] L. Peternel, E. Oztop, and J. Babic, "A shared control method for online human-in-the-loop robot learning based on locally weighted regression," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2016, p. 3900–3906.

[6] S. Nikolaidis, Y. X. Zhu, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in shared autonomy," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. ACM, Mar. 2017, p. 294–302.

[7] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review," *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, Nov. 2012.

[8] X. Wu, L. Xiao, S. Yixuan, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," *Future Generation Computer Systems*, vol. 135, 05 2022.

[9] N. Thomas, G. Ung, C. McGarvey, and J. D. Brown, "Comparison of vibrotactile and joint-torque feedback in a myoelectric upper-limb prosthesis," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, Jun. 2019.

[10] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 1, p. 43–53, Feb. 1987.

[11] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *ArXiv*, vol. abs/1707.06347, 2017.

[12] D. Yilmaz, B. Ugurlu, and E. Oztop, "Human-in-the-loop training leads to faster skill acquisition and adaptation in reinforcement learning-based robot control," in *2024 IEEE 18th International Conference on Advanced Motion Control (AMC)*. IEEE, 2024, p. 1–6.

[13] P. J. Huber, *Robust Estimation of a Location Parameter*. Springer New York, 1992, p. 492–518.

[14] J. Hwangbo, J. Lee, and M. Hutter, "Per-contact iteration method for solving contact dynamics," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, p. 895–902, 2018.

[15] T. Tanaka and H. Imamizu, "Sense of agency for a new motor skill emerges via the formation of a structural internal model," *Communications Psychology*, vol. 3, no. 1, Apr. 2025.