frontiers | Frontiers in Neurorobotics

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fnbot.2022.1068706

OPEN ACCESS

EDITED BY
Hang Su,
Fondazione Politecnico di Milano, Italy

REVIEWED BY
Wen Qi,
Politecnico di Milano, Italy
Hui Zhou,
Nanjing University of Science and Technology, China

*CORRESPONDENCE
Rong Song
songrong@mail.sysu.edu.cn

RECEIVED 13 October 2022
ACCEPTED 28 November 2022
PUBLISHED 22 December 2022

CITATION
Yang R, Zheng J and Song R (2022) Continuous mode adaptation for cable-driven rehabilitation robot using reinforcement learning. *Front. Neurorobot.* 16:1068706. doi: 10.3389/fnbot.2022.1068706

Continuous mode adaptation for cable-driven rehabilitation robot using reinforcement learning

Renyu Yang[1,2], Jianlin Zheng[1,2] and Rong Song[1,2]*

[1]Key Laboratory of Sensing Technology and Biomedical Instrument of Guangdong Province, School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China, [2]School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China

Continuous mode adaptation is very important and useful to satisfy the different user rehabilitation needs and improve human–robot interaction (HRI) performance for rehabilitation robots. Hence, we propose a reinforcement-learning-based optimal admittance control (RLOAC) strategy for a cable-driven rehabilitation robot (CDRR), which can realize continuous mode adaptation between passive and active working mode. To obviate the requirement of the knowledge of human and robot dynamics model, a reinforcement learning algorithm was employed to obtain the optimal admittance parameters by minimizing a cost function composed of trajectory error and human voluntary force. Secondly, the contribution weights of the cost function were modulated according to the human voluntary force, which enabled the CDRR to achieve continuous mode adaptation between passive and active working mode. Finally, simulation and experiments were conducted with 10 subjects to investigate the feasibility and effectiveness of the RLOAC strategy. The experimental results indicated that the desired performances could be obtained; further, the tracking error and energy per unit distance of the RLOAC strategy were notably lower than those of the traditional admittance control method. The RLOAC strategy is effective in improving the tracking accuracy and robot compliance. Based on its performance, we believe that the proposed RLOAC strategy has potential for use in rehabilitation robots.

KEYWORDS

admittance control, cable-driven rehabilitation robot, human-robot cooperation, human-robot interaction, optimal control, robot compliance

1 Introduction

Stroke is one of the leading causes of neurological and functional disability. In China, two million people suffer from stroke each year. Rehabilitation robots have attracted tremendous interest among researchers globally, as they can provide high-intensity, repetitive, and interactive rehabilitation training for post-stroke patients and overcome the labor-intensiveness of traditional manual rehabilitation training (Kwakkel et al., 2008). Rehabilitation robots, including various exoskeleton-type rehabilitation robots, such as ARMin (Nef et al., 2007), RUPERT (Huang et al., 2016), and UL-EXO7 (Kim et al., 2013) mimic the role of therapists to provide assistive forces to each joint of the human arm in rehabilitation training. However, these exoskeletons with hulking rigid links and motors attached to the human arm significantly increase the movement inertia, resulting in change in human arm dynamics, which will reduce the transparency of human–robot interaction (HRI) (Mao et al., 2015). To reduce the moving mass of the robot, a novel rehabilitation robot called cable-driven rehabilitation robot (CDRR), wherein the end-effectors are driven by cables instead of hulking rigid links, was developed, which improved the HRI performance owing to its excellent characteristics of low inertia, compliant structure, safety, and transparency (Jin et al., 2018). Mao et al. (2015) developed a cable driven exoskeleton (CAREX) for upper arm rehabilitation, which uses multi-stage cable-driven parallel mechanism to reduce the movement inertia, and the feasibility was verified in patients. Alamdari and Krovi (2015) designed a home-based cable-driven parallel platform robot driven by five cables for upper-limb neuro-rehabilitation in three-dimensional space. Cui et al. (2017) designed a 7-degrees of freedom (DOFs) cable-driven arm exoskeleton can easily assist the upper limbs to realize complex training tasks, involving rotation, translation, and their combination. Chen et al. (2019) designed a cable-driven parallel waist rehabilitation robot and a two-level control algorithm was proposed to assist patients with waist injuries to perform rehabilitation training.

The control strategies applied in rehabilitation robots play a critical role in the rehabilitation effectiveness (Guanziroli et al., 2019). According to the different recovery stages of post-stroke patients, the control strategies mainly include passive and active control (Proietti et al., 2016). Passive control is generally used to drive the patient repetitively move along predefined trajectories to improve the movement ability and reduce muscle atrophy, which is commonly adopted in the early recovery stages for patients with severe impairment (Jamwal et al., 2014). In active control, the rehabilitation robot assists the patient by complying with human motion intentions; it is mainly applied to patients with mild impairment. Koenig and Riener (2016) pointed out that passive control ignores the patient's voluntary engagement, which is one of the essential factors to facilitate neuroplasticity and motor function recovery of post-stroke patients (Warraich and Kleim, 2010), so its effect of stimulating neuroplasticity is limited. Performance-adaptive control strategies for patients with different levels of motor disabilities are necessary to meet user rehabilitation needs and recovery stages (Sainburg and Mutha, 2016). Meuleman et al. (2016) developed a variable admittance control for LOPES II, which can implement both active control to passive control. Wolbrecht et al. (2008) proposed an assist-as-needed (AAN) control strategy to allow robots to provide only essential assistance according to the patient's movement performance.

Obtaining suitable impedance/admittance parameters for the control strategy is essential to improve HRI performance for rehabilitation robots. The bio-inspired method assuming fixed impedance such as the musculoskeletal model (Pfeifer et al., 2012) or measurements of biological joint impedance (Erden and Billard, 2015) was used to estimate the impedance parameters through offline identification. The linear quadratic regulator (LQR) was adopted to obtain the desired admittance parameters through a cost function (Matinfar and Hashtrudi-Zaad, 2016). These methods would be good candidates when accurate models are available and their parameters can be well estimated. It is not practically applicable in rehabilitation training scenarios, because it is difficult to build the human dynamics model due to its features of nonlinearity, complexity, and variability (Driggs-Campbell et al., 2018). In addition, modeling and measurement errors are inevitable. To deal with this problem, the reinforcement learning (RL) algorithm was used to solve the given LQR problem, minimizing a cost function for optimizing the overall human–robot system performance (Modares et al., 2016; Li et al., 2017). RL algorithms have shown unprecedented successes in solving optimal control policy problems such as deep RL, including several policy search methods and deep Q-network (DQN) (Mnih et al., 2015; Silver et al., 2016). Doya (2000) used the knowledge of the system models to learn the optimal control policy and extend to continuous-time systems. To handle unknown dynamics, adaptive dynamic programming (ADP) with special a critic–actor structure has been extensively studied (Vrabie et al., 2009; Jiang and Jiang, 2012; Modares et al., 2015), which has become a promising tool for learning impedance/admittance parameters for the human–robot system. The ADP-based RL (ADPRL) approach was employed to automatically tune 12 impedance parameters and configure a robotic knee with human-in-the-loop (Wen et al., 2019; Gao et al., 2021). To achieve a compliant physical robot–environment interaction, Peng et al. (2022) used the ADPRL approach to obtain the desired admittance parameters based on the cost function composed of interaction force and trajectory tracking without the knowledge of the environmental dynamics. However, a fixed contribution weight of the cost function was adopted in previous studies, which cannot achieve continuous mode adaptation between the passive and active working mode.

Continuous mode adaptation is very important and useful to satisfy the different user rehabilitation needs and improves human–robot interaction (HRI) performance for rehabilitation robots. In this study, we present a novel reinforcement-learning-based optimal admittance control (RLOAC) strategy, which can achieve on-the-fly transitions between the passive and active working mode according to the human voluntary force. Firstly, we employed an RL algorithm to calculate the optimal admittance parameters for adapting to the different needs of patients without prior knowledge of the human dynamics model and formulated a new control strategy, which applied the optimal admittance parameters real time by minimizing the cost function to realize the desired HRI performance. Secondly, to promote patients' voluntary engagement, the contribution weights of the cost function were adjusted according to the human voluntary force.

2 Control strategy design

2.1 RLOAC framework

The RLOAC framework consists of two control loops—inner loop and outer loop—as illustrated in **Figure 1**. The inner-loop is intended for position control, which compensates for the robot nonlinear dynamics and guarantees trajectory tracking accuracy and stability. This module was implemented and reported in our previous work (Yang et al., 2022). The outer loop includes three modules: (1) a virtual training environment module provides visual feedback of the trajectory tracking and obstacle avoidance (TTOA) movement task to the subject and outputs the predefined trajectory P_t, detailed in Section "4.2 Adaptation to human dynamics"; (2) an optimal admittance control method is employed to yield the desired trajectory P_d to obtain the optimal HRI performance according to the human voluntary force F_h; and (3) an RL algorithm is designed to calculate the optimal parameters $\boldsymbol{K}$ online, considering that the human and robot dynamics parameters are difficult to identify in practice. The details of the outer-loop designs are presented below.

2.2 Optimal admittance control

The predefined trajectory $P_t \in \mathbb{R}^n$ set by the therapist, which is outputted directly to the CDRR, can be expressed in the form of a state equation in the Cartesian space.

$$\dot{\overline{P}}_t = \boldsymbol{A}\overline{P}_t + \boldsymbol{B}\ddot{P}_t, \tag{1}$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & I_n \\ 0 & 0 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ I_n \end{bmatrix}, \overline{P}_t = [P_t\ \dot{P}_t]^T \tag{2}$$

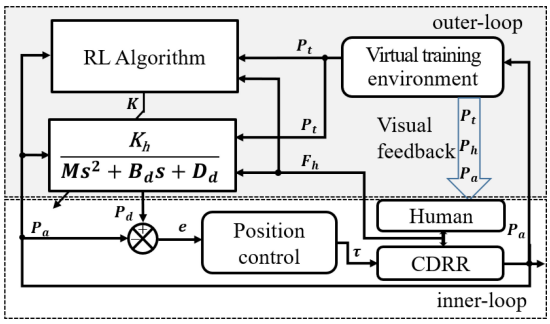

FIGURE 1
The proposed control framework.

The relationship between the human voluntary force F_h and the movement of the end-effector P_d can be described by the following admittance model (Modares et al., 2016; Li et al., 2017):

$$M_d\ddot{P}_d + B_d\dot{P}_d + D_d P_d = K_h F_h + l(P_t) \tag{3}$$

where M_d, B_d,D_d, and K_h are the inertia, damping, stiffness, and proportional gain matrices of the human voluntary force, respectively; $l(P)$ is an auxiliary input term, which will be designed later. By defining the augmented state $\overline{P}_d = [P_d,\dot{P}_d]^T$, (3) is expressed in the form of a state equation as

$$\dot{\overline{P}}_d = A\overline{P}_d + Bu, \tag{4}$$

where A and B are defined as in (2), and $u \in \mathbb{R}^m$. Combining (3) and (4) u is expressed as

$$u = M_d^{-1}\left(-B_d\dot{P}_d - D_d P_d + K_h F_h + l(P_t)\right) \tag{5}$$

The trajectory deformations are defined as $e_d = P_t - P_d$ and $\overline{e}_d = [e_d,\dot{e}_d]^T$. Combining (1) and (4), the trajectory deformation dynamics is expressed as

$$\dot{\overline{e}}_d = A\overline{e}_d + B\left(\ddot{\overline{P}}_t - u\right). \tag{6}$$

Similar to the approach in **Suzuki and Furuta (2012)**, human dynamics is expressed as

$$\dot{F}_h = -T^{-1}F_h + T^{-1}K_d\dot{e}_d + T^{-1}K_p e_d, \tag{7}$$

where K_p, K_d, and T are proportional coefficient of the human brain controller, differential coefficient, and time constant of the neuromuscular system, respectively. Defining the state variate as $X = [e_d,\dot{e}_d F_h]^T$ and then combining (6) and (7), a state equation for the HRI system can be established as

$$\dot{F}_h = -T^{-1}F_h + T^{-1}K_d\dot{e}_d + T^{-1}K_p e_d, \tag{8}$$

$$\overline{A} = \begin{bmatrix} 0 & I_n & 0 \\ 0 & 0 & 0 \\ T^{-1}K_p & T^{-1}K_d & -T^{-1} \end{bmatrix}, \overline{B} = [0 I_n 0], \tag{9}$$

$$\begin{aligned} u = & \ M_d^{-1}\left(-B_d\dot{P}_d - D_dP_d + K_hF_h + l\left(P_t\right)\right) \\ = & \ M_d^{-1}\left(B_d\dot{e}_d + D_de_d + K_hF_h\right) \\ & + M_d^{-1}\left(l\left(P_t\right) - M_d\ddot{P}_t - B_d\dot{P}_t - D_dP_t\right) \\ & \equiv u_e + u_d \end{aligned} \tag{10}$$

The control input u can be divided into two elements, feedback control input u_e and feedforward control input u_d(Modares et al., 2016; Li et al., 2017). We designed the auxiliary input term $l(P_t)$ in (10) as

$$l\left(P_t\right) = M_d\ddot{P}_t + B_d\dot{P}_t + D_dP_t. \tag{11}$$

Then, (10) can be rewritten as

$$\dot{X} = \overline{A}X + \overline{B}u_e. \tag{12}$$

The feedback control input can be rewritten as

$$u_e = KX, K = M_d^{-1}\left[\begin{array}{ccc} B_d & D_d & K_h \end{array}\right], \tag{13}$$

where $K \in \mathbb{R}^{3n \times 3n}$ is the control gain, which contains the admittance parameters. To minimize e_d, $\dot{e_d}$, u_e, and F_h, a cost function is designed as follows:

$$J = \int_0^\infty \left({e_d}^TQ_1e_d + {\dot{e_d}}^TQ_2\dot{e_d} + u_e^TR_1u_e + F_h^TR_2F_h\right)dt, \tag{14}$$

where $Q_1, Q_2, R_1, R_2 \in \mathbb{R}^{n=n}$ are the weighting factors of $e_d, \dot{e_d}, u_e$, and F_h, which allow a trade-off between the tracking error and human voluntary force. Q and $R_2$are defined as follows:

$$Q = \left[\begin{array}{ccc} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & R_2 \end{array}\right], R_2 = diag(r_2, \cdots, r_2), \tag{15}$$

where r_2 is the diagonal element of R_2. R_1 and R_2 determine the relative contributions of shared control between human and robot, respectively, to the cost J. Robotic systems are capable of adaptation of their autonomy level through dynamical adjustment of R_2. A smaller R_2 indicates a higher propensity for robots to lead the shared control task, vice versa. Since the motion capability and intention of the subject can be estimated by her/his voluntary force, R_2 should be adjusted according to human voluntary force to improve HRI performance in terms of robot compliance. A larger human voluntary force indicates a stronger capability and motion intentions to deviate the trajectory from the predefined trajectory. In this case, humans should be assigned the dominant role whereas the robots show greater compliance with the human voluntary actions, which can be achieved by increasing R_2. The reverse is true for a smaller human voluntary force. Thus, by modulating R_2 according to the human voluntary force, robots can realize continuous mode adaptation between passive and active working mode. The weighting element r_2 can be adjusted as follows:

$$r_2 = \begin{cases} r_{min} + \ \gamma\left(F_h, \alpha\right)\left(r_{max} - r_{min}\right), \ if ||F_h||_2 > F_c, \alpha \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \\ r_{min}, \ otherwise \end{cases} \tag{16}$$

where r_{min}, r_{max} are the minimum and maximum values of r_2. F_c is the threshold value of the human voluntary force. $||\cdot||_2$ denotes the 2-norm of a vector. $||F_h||_2 \leq F_c$ implies that F_h contains only sensor noises or involuntary force, which means the user cannot exert a voluntary force and therefore the CDRR should operate in the passive working mode. α is the magnitude of the directional difference between F_h and the optimal control input u_e^*. The condition $\alpha \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ indicates that the direction of F_h agrees with that of u_e^*. The conditions $||F_h|| > F_c$ and$\alpha \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ indicate that the user has some capability to correctly perform the cooperative control tasks; hence, the CDRR should operate in the active working mode. $\gamma\left(F_h, \alpha\right) \in [0, 1]$ is a weight factor, which is used to transit r_2 smoothly between r_{min} and r_{max} and is defined as

$$\gamma\left(F_h, \alpha\right) = tanh\left(\mu \cdot max\{0, ||F_h||_2 - ||F_c||_2\}^2 \cdot max\left(0, cos\alpha\right)\right), \tag{17}$$

where μ is a scale factor, which determines the ramping rate of γ. The weight factor $\gamma\left(F_h, \alpha\right)$ for $\mu = 0.5$, $F_c = 1.5\ N$ is illustrated in **Figure 2**.

Based on the optimal theorem (Kwakernaak and Sivan, 1972), the optimal admittance parameters can be obtained using the LQR algorithm with the exact model parameters of the human control and robot system dynamics. The optimal parameters that minimize the cost function (14) are given by

$$K^* = -R_1^{-1}\overline{B}^TP^*, u_e^* = K^*X, \tag{18}$$

where P^* is the solution to the following algebraic Riccati equation (ARE):

$$\overline{A}^TP + P\overline{A} - P\overline{B}R_1^{-1}\overline{B}^TP + Q = 0. \tag{19}$$

Thus, the optimal admittance parameters and proportional gain of the human voluntary force (M_d, B_d, D_d, and K_h) are determined.

2.3 RL algorithm

The disadvantage of solving the ARE (19) by using the LQR algorithm is that it requires the exact parameters of the human–robot system dynamics, which is difficult to know in practice. Several RL algorithms have been designed to overcome this limitation (Vrabie et al., 2009; Jiang and Jiang, 2012; Modares and Lewis, 2014). In this study, the RL algorithm (Jiang and Jiang, 2012) was employed for online calculation of the optimal admittance parameters for adapting to the needs of different patients under the human–robot system dynamics parameters completely unknown. Based on Theorem 2 in Jiang and Jiang

(2012), the numerical approximation form of the Bellman equation for the aforementioned LQR problem of the system in (12) to solve the ARE (19) is given below.

$$\begin{aligned} & X^T(t+\delta t)P^kX(t+\delta t)-X^T(t)P^kX(t) \\ & =-\int_t^{t+\delta t}X^T(\tau)\left[Q+\left(K^k\right)^TR_1K^k\right]X(\tau)\,d\tau \\ & +2\int_t^{t+\delta t}{u_e}^T(\tau)R_1K^{k+1}X(\tau)\,d\tau+2\int_t^{t+\delta t}\left[K^kX(\tau)\right]^T \\ & R_1K^{k+1}X(\tau)\,d\tau \end{aligned} \tag{20}$$

It is clear that (20) does not rely on the dynamic parameters $\overline{A}$ or $\overline{B}$ in (10). Then, the Kronecker product is used to express (20) as (Jiang and Jiang, 2012)

$$X^T(t+\delta t)P^kX(t+\delta t) = \overline{X}^T(t+\delta t)\overline{P}^k, \tag{21}$$

$$X^T(t)P^kX(t) = \overline{X}^T(t)\overline{P}^k, \tag{22}$$

$$\begin{aligned} & X^T(\tau)\left[Q+\left(K^k\right)^TR_1K^k\right]X(\tau) \\ & =X^T(\tau)\otimes X^T(\tau)\,vec\left(Q+\left(K^k\right)^TR_1K^k\right), \end{aligned} \tag{23}$$

$$\begin{aligned} & {u_e}^T(\tau)R_1K^{k+1}X(\tau) \\ & =\left[u_e^T(\tau)\otimes X^T(\tau)\right](R_1\otimes I_n)vec\left(K^{k+1}\right), \end{aligned} \tag{24}$$

$$\begin{aligned} & \left[K^kX(\tau)\right]^TR_1K^{k+1}X(\tau) \\ & =\left[X^T(\tau)\otimes X^T(\tau)\right]\left[I_n\otimes\left(K^{k+1}\right)^TR_1\right]vec\left(K^{k+1}\right), \end{aligned} \tag{25}$$

Where

$$\begin{aligned} X &= [X_1\cdots X_n], \\ \overline{X} &= \left[X_1^2, X_1X_2, \cdots, X_1X_n, X_2^2, X_2X_3, \cdots, X_{n-1}X_n, X_n^2\right] \\ \overline{P}^k &= \left[P_{11}^k, 2P_{12}^k, \cdots, 2P_{1n}^k, P_{22}^k, 2P_{23}^k, \cdots, 2P_{n-1,n}^k, P_{nn}^k\right] \end{aligned} \tag{26}$$

Combining (21) and (22), the left-hand side of (20) can be written as

$$\begin{aligned} & X^T(t+\delta t)P^kX(t+\delta t)-X^T(t)P^kX(t) \\ & =\left[\overline{X}^T(t+\delta t)-\overline{X}^T(t)\right]\overline{P}^k \end{aligned} \tag{27}$$

By combining (21)–(25), (20) can be rewritten as

$$\begin{aligned} & \left[\overline{X}^T(t+\delta t)-\overline{X}^T(t)\right]\overline{P}^k \\ & =-vec\left(Q+\left(K^k\right)^TR_1K^k\right)\int_t^{t+\delta t}X^T\otimes X^Td\tau \\ & +2(R_1\otimes I_n)\,vec\left(K^{k+1}\right)\int_t^{t+\delta t}u_e^T\otimes X^Td\tau \\ & +2\left[I_n\otimes\left(K^{k+1}\right)^TR\right]vec\left(K^{k+1}\right)\int_t^{t+\delta t}X^T\otimes X^Td\tau \end{aligned} \tag{28}$$

We introduce the following definitions to reduce (27) into a simple form:

$$\begin{aligned} \delta_{XX} &= \overline{X}^T(t+\delta t)-\overline{X}^T(t), \\ I_{XX} &= \int_t^{t+\delta t}X^T\otimes X^Td\tau, \\ I_{Xu} &= \int_t^{t+\delta t}X^T\otimes X^Td\tau, \\ b^k &= -\mathrm{I}_{XX}vec\left(Q+\left(K^k\right)^TR_1K^k\right), \\ \Gamma^k &= \left[\delta_{XX},\ -2\mathrm{I}_{XX}\left(\mathrm{I}_n\otimes\left(K^{k+1}\right)^TR_1\right)-2\mathrm{I}_{Xu}(R_1\otimes\mathrm{I}_n)\right]. \end{aligned} \tag{29}$$

Then, (27) can be simplified as

$$\Gamma^k\begin{bmatrix}\overline{P}^k \\ vec\left(K^{k+1}\right)\end{bmatrix}^k=b \tag{30}$$

Refer to study (Jiang and Jiang, 2012), a least-squares (LS) method is implemented online to obtain the optimal solution P^*. First, set $u_e = K^0+\varphi$ as the initial input. K^0 is the initial value of the control gain. φ is a probing noise. Then, the online data are collected and δ_{XX}, I_{XX}, and I_{Xu} are calculated until the following rank condition is satisfied:

$$rank([I_{XX}, I_{Xu}]) = \frac{3n(3n+1)}{2}+3mn \tag{31}$$

After the rank condition is satisfied, the LS solution is obtained as

$$\begin{bmatrix}\overline{P}^k \\ vec\left(K^{k+1}\right)\end{bmatrix}=[(\Gamma^k)^T\Gamma^k]^{-1}(\Gamma^k)^Tb^k \tag{32}$$

Then, the policy is improved as $u_e = K^{k+1}X$ and the above procedure of LS is repeatedly implemented until $||K^{k+1}-K^k|| < \varepsilon$. Finally, the optimal K^* is obtained. The RL algorithm is shown in **Table 1**.

Remark 1: To satisfy persistently exciting condition, the probing noise φ is added to the control input signal, which is necessary to guarantee nonsingular in LS solving process.

Remark 2: In Modares et al. (2016) and Li et al. (2017), RL algorithm is employed to solve the ARE with partial knowledge of system dynamics. Specifically, $\overline{A}$ is not needed in solving process, but $\overline{B}$ is still required for policy improvement. In contrast, In this study, we referred to Jiang and Jiang (2012) and employed the RL algorithm, which only uses the online information of input and system states, to solve ARE (19) neither relying on $\overline{A}$ nor $\overline{B}$. As can be seen from the definitions of $\overline{A}$ and $\overline{B}$ in (8), one can conclude that completely both human control dynamic parameters in (7) and robotic impedance parameters in (3) are not required in our method. The convergence of the RL algorithm was proofed by Theorem 7 in Jiang and Jiang (2012). Although both this study and previous study (Jiang and Jiang, 2012) employed the RL algorithm to obtain the optimal admittance parameters for improving the HRI performance with completely unknown dynamics parameters, their study does not address HRI issue for robot.

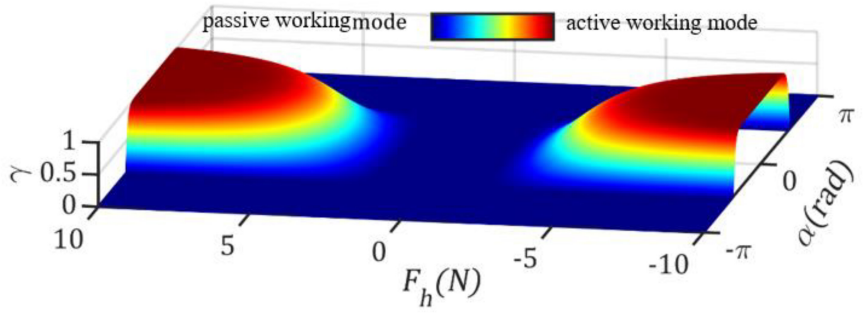

FIGURE 2
Smooth transition of weight factor $\gamma(F_h, \alpha)$ between 0 and 1. $\gamma(F_h, \alpha) = 0$ and $\gamma(F_h, \alpha) = 0$ correspond to passive and active working mode, respectively.

3 System description

To perform our research on an upper-limb rehabilitation robot, a 3-DOF CDRR prototype was developed in our laboratory. As shown in **Figure 3A**, the CDRR constructed to demonstrate and test the proposed control strategy consisted of a cubic mechanical framework, cable transmission mechanism, actuator module, sensors, controller (MicroLabBox, dSPACE, Germany), and personal computer [intel i7-8700 CPU 3.2 G and 32 GB of random access memory (RAM), China] with ControlDesk (dSPACE, Germany) and MATLAB R2019b software. The cable transmission and actuator module consisted of four cables, pulleys, four winches, an end-effector, and four motors (DM1B-045G, Yokogawa, Japan) with servo drivers (UB1DG3, Yokogawa, Japan). The four cables were pre-stretched high stiffness and made of lightweight steel wires. One end of each cable was fastened to the end-effector and the other end was fastened to the winch. The winches were driven by the motors to control the lengths of the cables (**Figure 3B**). The sensors on the CDRR included S-shaped tensile/force sensors (HSTL-BLSM, Beijing Huakong Xingye Technology Company, China) mounted on the mechanical framework to measure the cable tension, a 6-axis F/T sensor (SRI-V-210105-G, Sunrise Instruments, China) attached to the end-effector to measure the human voluntary force between the CDRR and human, and a motion capture system (OptiTrack, NaturalPoint, USA) with four cameras (Flex3, NaturalPoint, USA) used to measure the position of marker placed on the end-effector. The control strategy and data acquisition and recording were implemented on the controller with sampling frequency of 1 kHz. The guidance monitor with a virtual training environment was used to design the exercise game.

TABLE 1 Reinforcement-learning algorithm.

RL Algorithm
1 Select an admissible policy $u_e = K^0 + \varphi$;
2 For $k = 0, 1, 2\cdots$, given K^k, collect online data, calculate δ_{XX}, I_{XX}, and I_{Xu} until the rank condition given by equation (31) is satisfied, and then solve out $\overline{P}^k$, K^{k+1};
3 Improve control policy $u_e = K^{k+1}X$, go to step 2 until $\|\|K^{k+1} - K^k\|\| = \varepsilon$; ;
4 Use $u_e = K^{k+1}X$ as the approximated optimal policy to the system.

4 Simulations studies

4.1 Optimization of admittance parameters through RL algorithm

Simulation studies were conducted to investigate the convergence speed and accuracy of RL algorithm. For comparison, the optimal admittance parameters were obtained using the RL algorithm and LQR algorithm (Matinfar and Hashtrudi-Zaad, 2016), respectively. According to Suzuki and Furuta (2012), we assumed that the human dynamics can be modeled as (7) with $K_p = 779, K_d = 288$, and $T = 0.18$ when applying the LQR method to simplify the simulations. The matrices $\overline{A}$ and $\overline{B}$ in (9) then become

$$\overline{A} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 4327.8 & 0 & 1600 & 0 & -5.556 & 0 \\ 0 & 4327.8 & 0 & 1600 & 0 & -5.556 \end{bmatrix},$$

$$\overline{B} = \begin{bmatrix} 0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0 \end{bmatrix}^T \quad (33)$$

The matrices R_1 and R_2 in the cost function (14) were set as

$$Q = diag\,(5000, 5000, 500, 500, 1, 1)\,, R_2 = I_2. \quad (34)$$

Similar to Matinfar and Hashtrudi-Zaad (2016) and Yang et al. (2021), the optimal admittance parameters obtained directly by

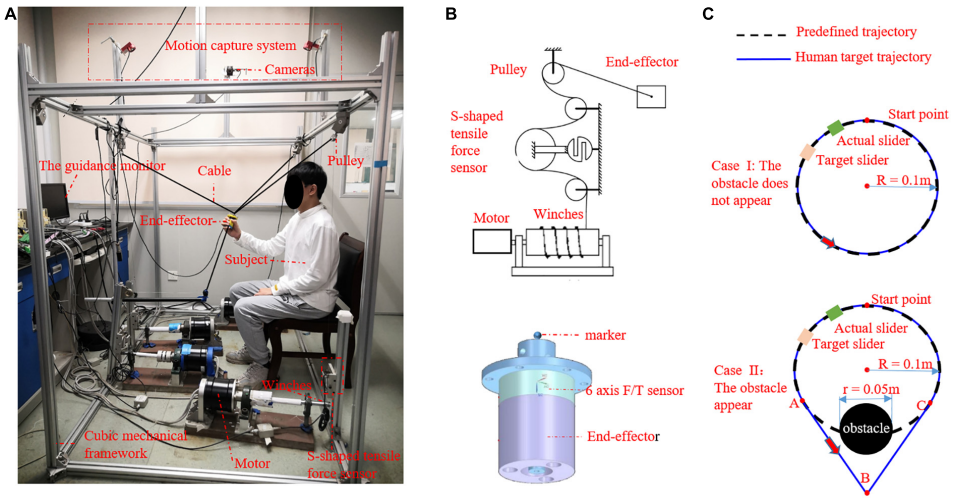

FIGURE 3
(A) Prototype of 3-degrees of freedom (DOFs) cable-driven rehabilitation robot (CDRR), **(B)** one cable transmission mechanism and actuator module, **(C)** the graphical guidance interface.

the LQR algorithm by considering the exact parameters of the human–robot system model (12) were

$$K^* = \begin{bmatrix} 151.459\ 0\ 58.257 & 0\ 0.810\ 0 \\ 0\ 151.459\ 0 & 58.257\ 0\ 0.810 \end{bmatrix}. \quad (35)$$

Generally, it is nearly impossible to obtain the actual parameters of the human–robot system model (12). To avoid requiring these parameters, the RL algorithm was reformulated and fit into the optimal admittance parameters calculated online in Section "3 System description." The initial values of the system parameters were set as

$$K_0 = \begin{bmatrix} 1200\ 1400\ 1400 & 1500\ 60\ 4 \\ 1500\ 1400\ 1500 & 2000\ 70\ 10 \end{bmatrix}, P_0 = 10\mathbf{I}_6,$$

$$X_0 = \begin{bmatrix} 0.1\ 0.1 & 0\ 0 & 0\ 0 \end{bmatrix}^T. \quad (36)$$

To satisfy the requirement of persistent excitation, we chose a probing noise given by

$$\varphi = \sum_{\omega}^{100} 0.001 = (rand - 0.5)\, sin\left(\omega \times (rand - 0.5)\right)/\omega, \quad (37)$$

where *rand* is a random number that varies from 0 to 1. The sampling time was selected as $T = 0.001$ and 100 samples were collected in each iteration. After 18 iterations, the optimal admittance parameters obtained by the RL algorithm were

$$K = \begin{bmatrix} 151.462\ 0.003\ 58.257 & 0\ 0.810\ 0 \\ 0.006\ 151.464\ 0 & 58.256\ 0\ 0.810 \end{bmatrix}. \quad (38)$$

Figure 4A illustrates the evolution of the admittance parameters and Figure 4B show that of the error $||K - K^*||_2$

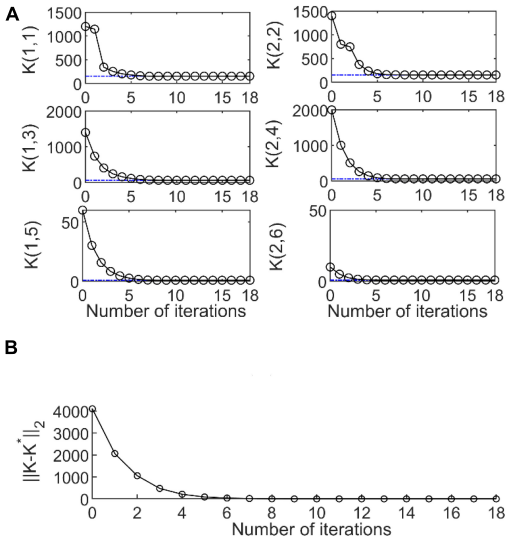

FIGURE 4
(A) Evolution of admittance parameters *K* for the duration of simulation, Blue lines are the optimal admittance parameters K* obtained with the linear quadratic regulator (LQR) algorithm for a specific human control dynamic model. Black lines are the results obtained by reinforcement learning (RL) method under unknown human dynamics; $K(i, j)$ is the element of *K*, *i*, and *j* are index of row and column, respectively. **(B)** Convergence of $||K - K^*||_2$.

between the RL algorithm and LQR method. After five iterations (0.5 s), the convergence errors of the optimal admittance parameters were lower than 0.01. Thus, the RL algorithm has similar accuracy as that of the LQR algorithm and acceptable convergence speed.

4.2 Adaptation to human dynamics

The TTOA movement task was designed which applied to across passive and active working mode for different subjects and displayed in a graphical guidance interface. As shown in **Figure 3C**, a predefined trajectory P_t (black dotted line) represented a suitable basic movement task created for patients without voluntary movement ability, which was defined as

$$P_t = [0.1cos(0.1\pi t+0.5\pi), 0.65+0.1sin(0.1\pi t+0.5\pi)]. \quad (39)$$

P_a, which was the actual position of the end-effector, was displayed in real time with a green slider. P_h (blue line) was the human target path, which remained unknown to the CDRR but was displayed in real time with an orange slider in the graphical interface and can be seen by the subject. During this task, the subject was instructed to look at the green slider and orange slider in the graphical interface and control the end-effector by using her/his hand and let the green slider track the orange slider with the best performance. In order to engage and challenge patients with less severe impairments, an obstacle with a diameter of 0.05 m and center at the coordinates of $O_2\begin{pmatrix}0\\0.5500\end{pmatrix}$ may appear on the path of P_t when the orange slider reaches the point A $\begin{pmatrix}-0.0707\\0.5793\end{pmatrix}$, and disappear when the orange slider arrives at point C $\begin{pmatrix}0.0707\\0.5793\end{pmatrix}$ (show as Case II on the bottom row in **Figure 3C**). The human target path was described follows

$$P_h = \begin{cases} P_t & t_0 \le t < t_1 \\ A+(t-t_1)(B-A)/(t_2-t_1) & t_1 \le t < t_2 \\ B+(t-t_2)(C-B)/(t_3-t_2) & t_2 \le t < t_3 \\ P_t t_3 \le t < t_4 \end{cases}, \quad (40)$$

where A $\begin{pmatrix}-0.0707\\0.5793\end{pmatrix}$, B $\begin{pmatrix}0\\0.4500\end{pmatrix}$, and C $\begin{pmatrix}0.0707\\0.5793\end{pmatrix}$ were the joined points; t_0 = 0 s, t_1 = 7.5 s, t_2 = 10.0 s, t_3 = 12.5 s, and t_4 = 20.0 s. When the target slider reached point A at t_1 = 7.5 s, the subject needed to adjust its path and plan a new bypath. The arc AC moved partly into triangle ABC in P_h to bypass the obstacle.

In this simulation, the feasibility of the proposed RLOAC strategy was verified through simulation of the TTOA movement task. The RLOAC strategy was implemented by the method presented in Section "2 Control strategy design," and the initial parameters were set as in the above simulation example. The parameters $\mu = 0.5$, $F_c = 1.5N$, $r_{\min} = 1$, and $r_{\max} = 300$ were adopted. The human dynamics model (7) was used to simulate the human voluntary force. To verify whether the proposed method can adapt itself to patients with different capabilities, three types of disturbance forces were added to the human voluntary force to simulate the movements of three types of patients with high, moderate, and low levels of capabilities (Suzuki and Furuta, 2012). Similar to Suzuki and Furuta (2012), the disturbance forces were designed as shown in **Table 2**. The simulation results are presented in **Figure 5**. **Figure 5A** illustrates the simulation results of the human voluntary forces exerted by patients with high, moderate, and low levels of capabilities. **Figure 5B** shows the trajectory tracking results under these three simulation conditions. All trajectory tracking errors were small under these three simulation conditions. Thus, the RLOAC strategy is suitable for patients with different capabilities.

To compare the performance of the proposed RLOAC strategy with those of other methods without online optimization of the admittance parameters, further simulation was conducted by utilizing the following traditional admittance control (TAC) to perform the aforementioned task (Culmer et al., 2010).

$$P_d = P_t+F_h/(K_{TAC}+C_{TAC}S), \quad (41)$$

where the stiffness matric $K_{TAC} = diag(125, 125)$ and damping matric $C_{TAC} = diag(49.4, 49.4)$. Both RLOAC and TAC method used the same inner loop controller. The detailed design can be referred to **Yang et al. (2022)** We evaluated the HRI performance in terms of the tracking accuracy and robot compliance. The absolute tracking error was used to evaluate the tracking accuracy, which was defined as follows:

$$||Error(t)||_2 = ||\,||P_a(t)-P_h(t)||_2. \quad (42)$$

The energy per unit distance (EPUD) was adopted to evaluate the robot compliance (Lee et al., 2018; Zhou et al., 2021), which was defined as follows:

$$EPUD(t) = |F_h(t)\cdot\triangle d(t)|/|\triangle d(t)| \quad (43)$$

where $\triangle d = P_a-P_t$ is the trajectory deviation made by the subject from P_t to P_a. A smaller EPUD(t) value indicates higher robot compliance with the human motion intentions (**Lee et al., 2018**; **Zhou et al., 2021**).

The simulation results are presented in **Figure 6**. The tracking accuracy of the proposed RLOAC is higher than that of the TAC, especially when bypassing an unpredictable obstacle, as shown in **Figure 6A**. Moreover, the value of $EPUD(t)$ needed to bypass the obstacle was notably smaller with the proposed RLOAC when compared with the TAC, as shown in **Figure 6B**. This comparison of the simulation results indicates that the CDRR with the proposed RLOAC achieved higher accuracy and compliance with the human motion intention.

TABLE 2 The design of three typed of disturbance force.

Levels of capabilities	Amplitude of disturbance force
High	0 N
Moderate	Random (−5 N, 5 N)
Low	Random (−10 N, 10 N)

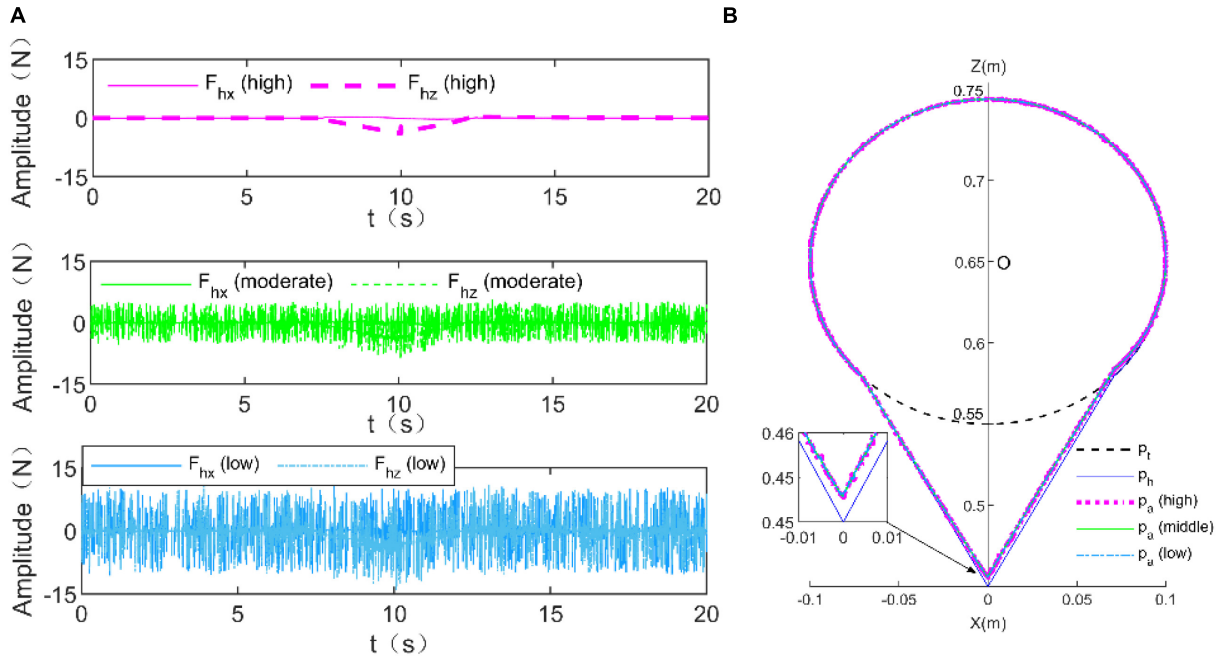

FIGURE 5
Simulation results for trajectory tracking performance under the voluntary force generated by the human dynamics models with three different levels of capabilities. **(A)** Human voluntary forces for high (violet), moderate (green), and low (blue) levels of capabilities. **(B)** Trajectory tracking performance.

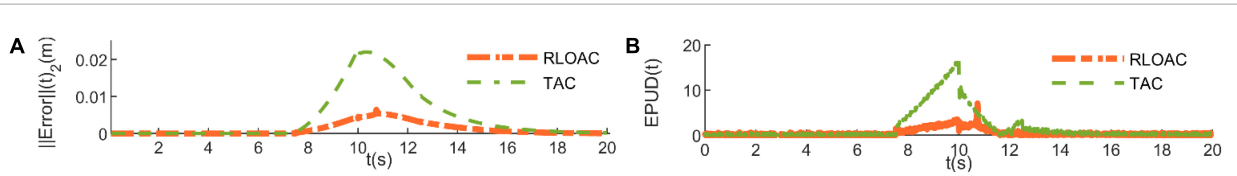

FIGURE 6
Comparison of simulation results of reinforcement-learning-based optimal admittance control (RLOAC) (orange line) and traditional admittance control (TAC) (green line). **(A)** Tracking accuracy. **(B)** Robot compliance.

5 Experimental studies

Because human control behavior and motor learning are complex and have variable characteristics, which cannot be described in the above simulations, we further investigated the validity of the proposed method through experimentation with human subjects on the 3-DOF CDRR constructed by us and illustrated in Figure 3.

5.1 Experimental setup

Ten healthy subjects four males and six females, age [mean (M) 28 years with standard deviation (SD) of 4.92], height (M 1.67 m with SD 4.77), and weight (M 57.05 kg with SD 7.45) with no history of neurological impairment were recruited for the experiment. All subjects provided informed consent before participating in the experiment. They were instructed to grasp the end-effector of the CDRR and perform the TTOA movement task (detailed in Section "4.2 Adaptation to human dynamics"), as shown in Figure 3C. To better show the condition of switching between passive and active working modes, the TTOA movement task was set as follows:

(1) The obstacle may appear randomly with a probability of 50%. Depending on the non-appearance or appearance of the obstacle, the task scenarios is called Case II as shown on the top row in Figure 3C or Case II on the bottom row in Figure 3C, respectively. The equation of P_h is different in these two cases. Specifically, P_t and P_h overlap in Case I, which are both expressed as (39). P_t andP_h only partial overlap in Case II. P_t is expressed as (39), while P_h is expressed as (40).

(2) The task of Case I and Case II were conducted periodically by performing one cycle per period.

Initially, the experimenter demonstrated the TTOA movement task to the subjects and ensured that each subject understood the task. Then, the subjects were allowed to practice 20 unrecorded cycles/period. After this preliminary experiment, to ensure a fair comparison, the same experimental protocol was conducted for two trials by each subject: once with TAC and once using the proposed RLOAC in a random order unknown to the subject. For each control strategy (TAC or RLOAC), each subject executed 10 cycles/period, including five cycles/period for Case I and five for Case II, and the data were recorded for analysis. Further, Case I and Case II appeared in a random order unknown to the subject.

5.2 Data analysis

We performed a quantitative evaluation of the HRI performance based on the following measures:

(1) Mean absolute tracking error (MATE), defined as

$$MATE = \frac{1}{s_j}\sum_{i=1}^{s_j} ||P_a(t_i) - P_h(t_i)||_2, \tag{44}$$

where $P_a(t_i)$ and $P_h(t_i)$ represented the actual position and the human target position of the end-effector at the *ith* sampling instant, respectively; and $s_j\ (j = 1, 2, 3)$ were the total number of samples for each subject during the entire experiment, the active working mode, and the passive working mode, respectively.

(2) The EPUD for each subject, defined as

$$EPUD = \sum_{i=1}^{s_j} \left|F_h(t_i) \cdot \triangle d(t_i)\right| / \sum_{i}^{s_j} |\triangle d(t)(t_i)| \tag{45}$$

where $F_h(t_i)$ and $\triangle d(t_i)$ represented the human voluntary force and the trajectory deviation at the *ith* sampling instant, respectively.

We employed paired-samples *t*-tests with a significance level of $\alpha = 0.05$ to test differences in MATE and EPUD of the 10 subjects between the two control strategies (TAC and RLOAC) (Losey and O'Malley, 2018). All statistical analyses were performed using SPSS 19 (SPSS, Inc., Chicago, IL, USA).

6 Experimental results

Our findings are presented in Figures 7–9. Figure 7 depicts the results from the subject 2 when using the TAC strategy, whereas Figure 8 depicts the results of the subject 2 when using the proposed RLOAC strategy. Specifically, Figures 7A, 8A illustrate the tracking trajectories of 10 cycles for the TAC and the proposed RLOAC strategy, respectively. The first and second rows of Figures 7B, 8B illustrate the tracking errors and the EPUD(*t*), respectively. The deep blue scatterplot shows the results for active working mode, and the red scatterplot shows the results for passive working mode. The shadows in each subplot indicate the time durations for which the obstacle appeared.

It is clear from Figure 8 that with the proposed RLOAC strategy, the CDRR can assist the subject cooperatively by complying with her/his motion intention to complete the movement task with desirable performances in terms of the accuracy and compliance. As shown in the first row of Figures 7B, 8B, the tracking error, $||Error(t)||_2$, when using the RLOAC strategy was small and acceptable, and it was mostly notably lower than that of the TAC strategy. As seen in the second row of Figures 7B, 8B, the compliance indicated by EPUD(t) for the RLOAC strategy was below 13 in each case, which was notably better than that of the TAC strategy. The red sections in of Figures 7B, 8B show that the active working mode time for the proposed RLOAC strategy was notably longer than that for the TAC strategy for performing the same movement task. As seen in the triangular part of the tracking trajectories, by using the RLOAC strategy, the CDRR can comply with the subjects' voluntary actions and bypass unpredictable obstacles with greater accuracy and compliance than that possible with the TAC strategy. When using the RLOAC strategy, the tracking error and vibrations of the end-effector can be decreased through voluntary control by the subject, as it ensures better compliance and smooth switching between passive and active working mode. Even without obstacles in the path, the active working mode can be adopted to decrease the tracking error and vibrations. On the contrary, with the TAC strategy, the active working mode is rarely adopted in the absence of obstacles. Thus, the results of the representative subject show that the proposed RLOAC strategy has better accuracy and compliance, and can promote active working mode in comparison with the TAC strategy.

Figure 9 depicts the results for all 10 subjects in the form of mean ± std to statistically detect the differences between the two control strategies. As shown in Figure 9, the MATEs for the entire experiment of 10 cycles/period, the active working mode, and the passive working mode were (0.0160 ± 0.0021, 0.0168 ± 0.0043, and 0.0152 ± 0.0039) m and (0.0189 ± 0.0027, 0.0228 ± 0.0044, and 0.0170 ± 0.0091) m with the RLOAC and TAC strategies, respectively. By comparing the trials of RLOAC and TAC, there were statistically significant decreases in the MATE for the entire experiment ($p = 0.015$) and during the active working mode ($p = 0.001$). These differences were not statistically significant during passive working mode ($p = 0.693$). A similar pattern was discovered for the EPUDs. The detailed EPUDs of the RLOAC trials for the entire experiment, the active working mode, and the passive working mode were

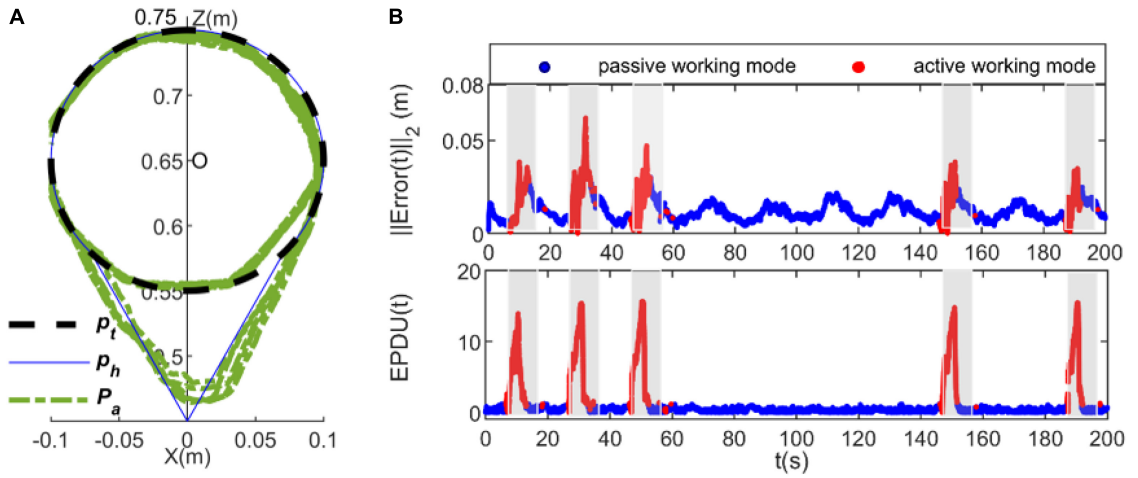

FIGURE 7
Experimental results of the proposed traditional admittance control (TAC) strategy measured from subject 2. **(A)** Tracking trajectories, **(B)** tracking error and EPUD(t). The shadows in each subplot indicate the time duration of appearance of the unpredicted obstacle and the intervention required from the participant to bypass it. The deep blue scatterplot shows the results of the passive working mode, and the red scatterplot shows the results of the active working mode.

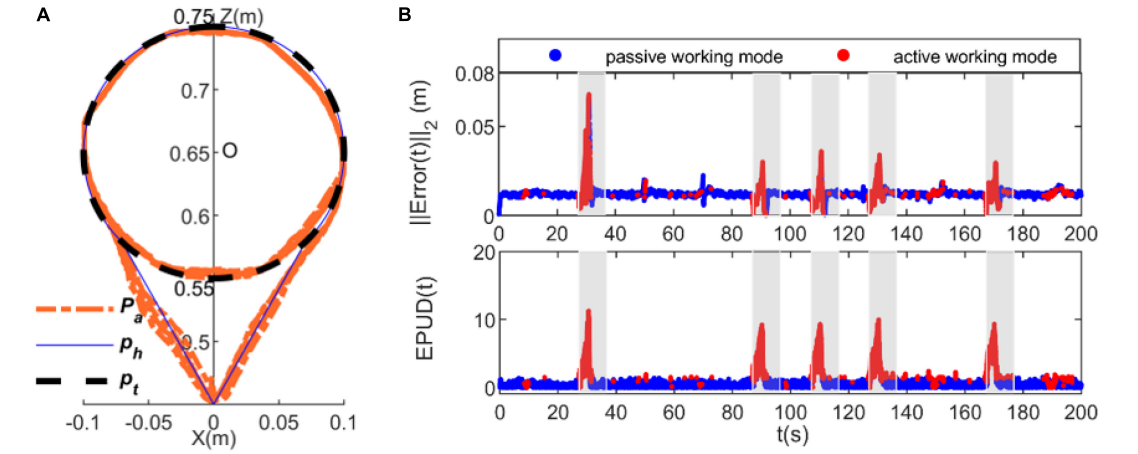

FIGURE 8
Experimental results of subject 2 when using the proposed RLOAC. **(A)** Tracking trajectories, **(B)** tracking error and EPUD(t).

(2.5931 ± 0.5740, 4.4704 ± 0.7217, and 0.5805 ± 0.1470), whereas in the TAC trials, the corresponding EPUDs were (4.0754 ± 0.4845, 6.4994 ± 1.4368, and 0.5569 ± 0.1137). The results of the RLOAC trials were significantly smaller than those of the TAC trials for the entire experiment (p = 0.001) and during the active working mode (p = 0.006), whereas the EPUDs of these two control strategies during the passive working mode had no significant differences (p = 0.560). Thus, the statistical quantification analysis proved that the proposed RLOAC strategy had desirable accuracy and compliance, which were statistically notably better than those of the TAC strategy in the comparison experiment.

7 Discussion and conclusion

In this study, an RLOAC strategy is proposed for a CDRR that can achieve continuous mode adaptation between the passive and active working modes. Experiment with 10 subjects were conducted on a self-designed CDRR, and the results demonstrated the effectiveness of the proposed control strategy. It is demonstrated that the proposed approach can potentially be applied in CDRR.

The RLOAC strategy improved the HRI performance in terms of tracking accuracy and robot compliance. The tracking error (Modares et al., 2016; Li et al., 2017, 2018) and EPUD (Lee et al., 2018; Zhou et al., 2021) are common performance

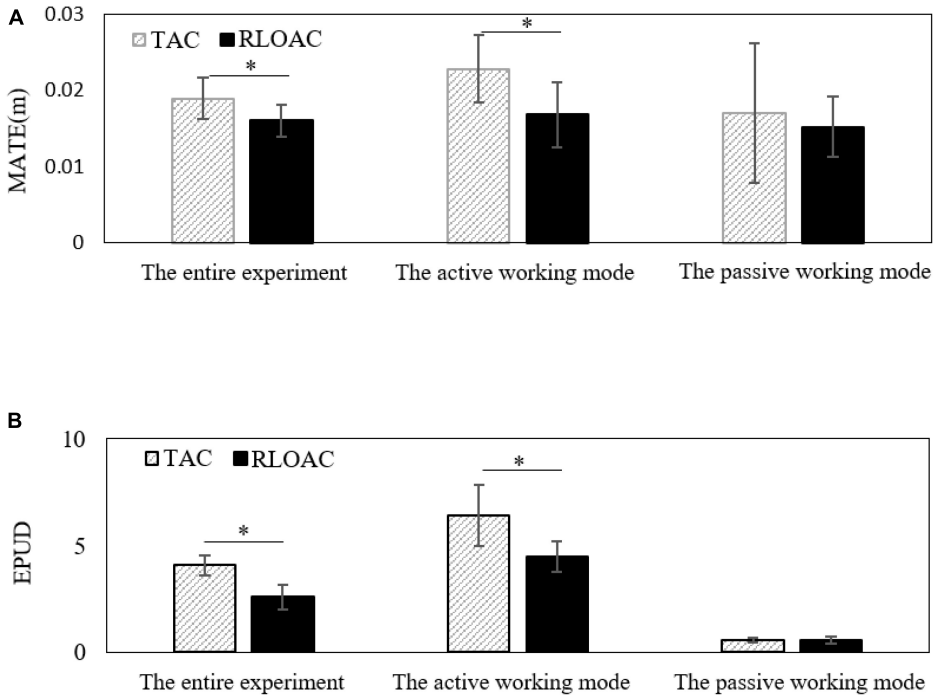

FIGURE 9
Comparison of human–robot interaction (HRI) performance in term of the mean absolute tracking error (MATE) **(A)** and energy per unit distance (EPUD) **(B)** between the proposed reinforcement-learning-based optimal admittance control (RLOAC) and traditional admittance control (TAC) in the entire experiment, the active working mode, and the passive working mode, respectively. Each plot shows the mean value for the subjects when TAC was used (grid) and when the proposed RLOAC was used (black). The error bars indicate the standard deviation, and the asterisk "*" indicates $p < 0.05$.

indexes of the HRI. A smaller tracking error indicates that the subjects can control the CDRR's motion more accurately (Modares et al., 2016; Li et al., 2018), and a smaller EPUD indicates higher robot compliance (Lee et al., 2018; Zhou et al., 2021). There was decrease in the means of absolute tracking error and EPUD during exercise, because the CDRR with RLOAC can obtain the suitable admittance parameters to optimize HRI performance. Significant differences between the RLOAC and TAC strategy were found in the performance metrics during both active working mode and in the overall experiment, because the contributions to the control task between subject and robot can be adjusted as necessary for rapid adaptation to the changes in human voluntary actions and task requirements. Thus, the CDRR with RLOAC exhibited high levels of compliance with human motion intentions and self-adaptive optimization to human dynamics. The controller type did not have a statistically significant effect in the passive working mode, because the tracking error was mainly determined by the inner loop position controller in this working mode. The increase in time spent in active working mode indicates that the RLOAC strategy can promote voluntary engagement during exercise. Because the subjects participated in the control loop, and their voluntary force were utilized to perceive their motion intentions (Li et al., 2018). Continuous mode adaptation according to subjects' voluntary force facilitated subjects driving the robot at their will made them feel in control during exercise, which may increase their motivation and confidence to use the affected limb (Proietti et al., 2016).

Comparing the RLOAC strategy with the traditional control strategies highlights its advantages. The well-recognized TAC strategy, widely applied in rehabilitation robots, was chosen as a comparison method because both the RLOAC and TAC yield the desired trajectories based on the human input forces using an admittance model to obtain robot compliance. The fixed admittance parameters were adopted in the TAC strategy, which meant that it could not adapt to the variability of human dynamics. In contrast, using reinforcement learning, the RLOAC can obtain suitable admittance parameters to optimize HRI performance. In contrast to most optimization algorithms, the RLOAC strategy can adjust admittance parameters online without the knowledge of human and robot dynamics models. Although adaptive impedance control has been applied to optimize interaction performance, as pointed out in Riener et al. (2005), Culmer et al. (2010), Proietti et al. (2016), and Zhou et al. (2021), admittance control is more stable than impedance

control. Therefore, the RLOAC strategy is more suitable for rehabilitation robots due to using admittance control. Use of RL algorithm to obtain the optimal admittance parameters and optimal HRI performance by minimizing a cost function has been suggested in previous studies (Modares et al., 2016; Li et al., 2017). However, in their studies, a partial knowledge of the system dynamics was still required. In contrast, the RL algorithm in this study was improved and used to address the HRI issue considering completely unknown human and robot dynamics parameters. Moreover, continuous and real-time mode adaptation was realized by dynamically adjusted contribution weight of the cost function according to the human voluntary force.

The limitations present in this study can be given as follows. We assumed that the human voluntary force can be directly measured by a 6-axis F/T sensor. In fact, the measured force was the interaction force between the human and the end-effector, which is composed of both voluntary and involuntary components. The applicability and clinical effectiveness of the proposed control strategy was not verified in post-stroke patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RY and RS contributed to the conception of the control algorithm. RY designed and performed the simulations and experiments and wrote the first draft of the manuscript. All authors contributed to the manuscript revision and approved the submitted version.

Funding

This research was supported by the National Key Research and Development Program of China (Grant No. 2022YFE0201900), the Shenzhen Science and Technology Research Program (No. SGDX20210823103405040), the Guangdong Science and Technology Plan Project (No. 2020B1212060077), and the Natural Science Foundation of Guangdong Province (No. 2020A1515010735).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alamdari, A., and Krovi, V. (2015). "Modeling and control of a novel home-based cable-driven parallel platform robot: PACER," in *Proceedings of the 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Hamburg.

Chen, Q., Zi, B., Sun, Z., Li, Y., and Xu, Q. (2019). Design and development of a new cable-driven parallel robot for waist rehabilitation. *IEEE ASME Trans. Mechatron.* 24, 1497–1507.

Cui, X., Chen, W. H., Jin, X., and Agrawal, S. K. (2017). Design of a 7-DOF cable-driven arm exoskeleton (CAREX-7) and a Controller for dexterous motion training or assistance. *IEEE ASME Trans. Mechatron.* 22, 161–172. doi: 10.1109/tmech.2016.2618888

Culmer, P. R., Jackson, A. E., Makower, S., Richardson, R., Cozens, J. A., Levesley, M. C., et al. (2010). A control strategy for upper limb robotic rehabilitation with a dual robot system. *IEEE ASME Trans. Mechatron.* 15, 575–585. doi: 10.1109/tmech.2009.2030796

Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural Comput.* 12, 219–245. doi: 10.1162/089976600300015961

Driggs-Campbell, K., Dong, R., and Bajcsy, R. (2018). Robust, informative human-in-the-loop predictions via empirical reachable sets. *IEEE Trans. Intell. Veh.* 3, 300–309. doi: 10.1109/TIV.2018.2843125

Erden, M. S., and Billard, A. (2015). End-point impedance measurements across dominant and nondominant hands and robotic assistance with directional damping. *IEEE Trans. Cybern.* 45, 1146–1157. doi: 10.1109/tcyb.2014.2346021

Gao, X., Si, J., Wen, Y., Li, M., and Huang, H. (2021). Reinforcement learning control of robotic knee with human-in-the-loop by flexible policy iteration. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 5873–5887. doi: 10.1109/TNNLS.2021.3071727

Guanziroli, E., Cazzaniga, M., Colombo, L., Basilico, S., Legnani, G., and Molteni, F. (2019). Assistive powered exoskeleton for complete spinal cord injury:

Correlations between walking ability and exoskeleton control. *Eur. J. Phys. Rehabil. Med.* 55, 209–216. doi: 10.23736/s1973-9087.18.05308-x

Huang, J., Tu, X. K., and He, J. P. (2016). Design and evaluation of the RUPERT wearable upper extremity exoskeleton robot for clinical and in-home therapies. *IEEE Trans. Syst. Man Cybern. Syst.* 46, 926–935. doi: 10.1109/tsmc.2015.2497205

Jamwal, P. K., Xie, S. Q., Hussain, S., and Parsons, J. G. (2014). An adaptive wearable parallel robot for the treatment of ankle injuries. *IEEE ASME Trans. Mechatron.* 19, 64–75. doi: 10.1109/tmech.2012.2219065

Jiang, Y., and Jiang, Z. P. (2012). Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica* 48, 2699–2704. doi: 10.1016/j.automatica.2012.06.096

Jin, X., Prado, A., and Agrawal, S. K. (2018). Retraining of human gait – are lightweight cable-driven leg exoskeleton designs effective? *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 847–855. doi: 10.1109/tnsre.2018.2815656

Kim, H., Miller, L. M., Fedulow, I., Simkins, M., Abrams, G. M., Byl, N., et al. (2013). Kinematic data analysis for post-stroke patients following bilateral versus unilateral rehabilitation with an upper limb wearable robotic system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 153–164. doi: 10.1109/tnsre.2012.2207462

Koenig, A. C., and Riener, R. (2016). "The human in the loop," in *Neurorehabilitation technology*, eds D. J. Reinkensmeyer and V. Dietz (Cham: Springer International Publishing), 161–181.

Kwakernaak, H., and Sivan, R. (1972). *Linear optimal control systems*. New York, NY: Wiley-interscience.

Kwakkel, G., Kollen, B. J., and Krebs, H. I. (2008). Effects of robot-assisted therapy on upper limb recovery after stroke: A systematic review. *Neurorehabil. Neural Repair* 22, 111–121. doi: 10.1177/1545968307305457

Lee, K. H., Baek, S. G., Lee, H. J., Choi, H. R., Moon, H., and Koo, J. C. (2018). Enhanced transparency for physical human-robot interaction using human hand impedance compensation. *IEEE ASME Trans. Mechatron.* 23, 2662–2670. doi: 10.1109/tmech.2018.2875690

Li, Z. J., Huang, B., Ye, Z. F., Deng, M. D., and Yang, C. G. (2018). Physical human-robot interaction of a robo is exoskeleton by admittance control. *IEEE Trans. Industr. Electron.* 65, 9614–9624. doi: 10.1109/tie.2018.2821649

Li, Z. J., Liu, J. Q., Huang, Z. C., Peng, Y., Pu, H. Y., and Ding, L. (2017). Adaptive impedance control of human-robot cooperation using reinforcement learning. *IEEE Trans. Industr. Electron.* 64, 8013–8022. doi: 10.1109/tie.2017.2694391

Losey, D. P., and O'Malley, M. K. (2018). Trajectory deformations from physical human–robot interaction. *IEEE Trans. Robot.* 34, 126–138. doi: 10.1109/tro.2017.2765335

Mao, Y., Jin, X., Dutta, G. G., Scholz, J. P., and Agrawal, S. K. (2015). Human movement training with a cable driven ARm EXoskeleton (CAREX). *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 84–92. doi: 10.1109/tnsre.2014.2329018

Matinfar, M., and Hashtrudi-Zaad, K. (2016). Optimization-based robot compliance control: Geometric and linear quadratic approaches. *Int. J. Robot. Res.* 24, 645–656. doi: 10.1177/0278364905056347

Meuleman, J., van Asseldonk, E., van Oort, G., Rietman, H., and van der Kooij, H. (2016). LOPES II-design and evaluation of an admittance controlled gait training robot with shadow-leg approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 352–363. doi: 10.1109/tnsre.2015.2511448

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236

Modares, H., and Lewis, F. L. (2014). Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning. *IEEE Trans. Automat. Contr.* 59, 3051–3056. doi: 10.1109/tac.2014.2317301

Modares, H., Lewis, F. L., and Jiang, Z. P. (2015). H-infinity tracking control of completely unknown continuous-time systems via off-policy reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 2550–2562. doi: 10.1109/tnnls.2015.2441749

Modares, H., Ranatunga, I., Lewis, F. L., and Popa, D. O. (2016). Optimized assistive human-robot interaction using reinforcement learning. *IEEE Trans. Cybern.* 46, 655–667. doi: 10.1109/tcyb.2015.2412554

Nef, T., Mihelj, M., and Riener, R. (2007). ARMin: A robot for patient-cooperative arm therapy. *Med. Biol. Eng. Comput.* 45, 887–900. doi: 10.1007/s11517-007-0226-6

Peng, G., Chen, C. L. P., and Yang, C. (2022). Neural networks enhanced optimal admittance control of robot-environment interaction using reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 4551–4561. doi: 10.1109/TNNLS.2021.3057958

Pfeifer, S., Vallery, H., Hardegger, M., Riener, R., and Perreault, E. J. (2012). Model-based estimation of knee stiffness. *IEEE Trans. Biomed. Eng.* 59, 2604–2612. doi: 10.1109/TBME.2012.2207895

Proietti, T., Crocher, V., Roby-Brami, A., and Jarrasse, N. (2016). Upper-limb robotic exoskeletons for neurorehabilitation: A review on control strategies. *IEEE Rev. Biomed. Eng.* 9, 4–14. doi: 10.1109/rbme.2016.2552201

Riener, R., Lunenburger, L., Jezernik, S., Anderschitz, M., Colombo, G., and Dietz, V. (2005). Patient-cooperative strategies for robot-aided treadmill training: First experimental results. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 380–394. doi: 10.1109/tnsre.2005.848628

Sainburg, R. L., and Mutha, P. K. (2016). "Movement neuroscience foundations of neurorehabilitation," in *Neurorehabilitation technology*, eds D. J. Reinkensmeyer and V. Dietz (Cham: Springer International Publishing), 19–38.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Suzuki, S., and Furuta, K. (2012). Adaptive impedance control to enhance human skill on a haptic interface system. *J. Contr. Sci. Eng.* 2012, 365067–365077. doi: 10.1155/2012/365067

Vrabie, D., Pastravanu, O., Abu-Khalaf, M., and Lewis, F. L. (2009). Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica* 45, 477–484. doi: 10.1016/j.automatica.2008.08.017

Warraich, Z., and Kleim, J. A. (2010). Neural plasticity: The biological substrate for neurorehabilitation. *PM R* 2, S208–S219. doi: 10.1016/j.pmrj.2010.10.016

Wen, Y., Si, J., Brandt, A., Gao, X., and Huang, H. (2019). Online Reinforcement learning control for the personalization of a robotic knee prosthesis. *IEEE Trans. Cybern.*50, 2346–2356. doi: 10.1109/TCYB.2019.2890974

Wolbrecht, E. T., Chan, V., Reinkensmeyer, D. J., and Bobrow, J. E. (2008). Optimizing compliant, model-based robotic assistance to promote neurorehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 16, 286–297. doi: 10.1109/TNSRE.2008.918389

Yang, R., Li, Z., Lyu, Y., and Song, R. (2022). Fast finite-time tracking control for a 3-DOF cable-driven parallel robot by adding a power integrator. *Mechatronics* 84:102782. doi: 10.1016/j.mechatronics.2022.102782

Yang, R., Zhou, J., and Song, R. (2021). "adaptive admittance control based on linear quadratic regulation optimization technique for a lower limb rehabilitation robot," in *in Proceedings of the 6th IEEE international conference on advanced robotics and mechatronics (ICARM)*, Changqing.

Zhou, J., Li, Z. J., Li, X. M., Wang, X. Y., and Song, R. (2021). Human-robot cooperation control based on trajectory deformation algorithm for a lower limb rehabilitation robot. *IEEE ASME Trans. Mechatron.* 26, 3128–3138. doi: 10.1109/tmech.2021.3053562