Adaptive Impedance Control Based on Reinforcement Learning in a Human-Robot Collaboration Task with Human Reference Estimation

Article *in* International Journal of Mechanics and Control · June 2020

3 authors:

Min Wu
Technische Universität Kaiserslautern

Yanhao He
Technische Universität Kaiserslautern

Steven Liu
Technische Universität Kaiserslautern

ISSN 1590-8844
International Journal of Mechanics and Control, Vol. XX, No. XX, 20XX

ADAPTIVE IMPEDANCE CONTROL BASED ON REINFORCEMENT LEARNING IN A HUMAN-ROBOT COLLABORATION TASK WITH HUMAN REFERENCE ESTIMATION

Min Wu* Yanhao He* Steven Liu*

* Institute of Control Systems, Department of Electrical and Computer Engineering, University of Kaiserslautern

ABSTRACT

In this work an adaptive impedance control scheme in a human-root collaboration task is designed. Both motion reference and impedance parameters of the robot control are adapted in real-time so that no priori task information is required. Reinforcement learning is used to find an optimal impedance parameter set to minimize a task-orient cost function, without fully knowledge of the system dynamic. The learned parameters are further adjusted by taking human's disagreement into consideration. The human motion reference is estimated from the lineralized contact dynamic model using system identification technology. The proposed method enables robot to generate active contribution in the collaboration and be flexible to variation of the task or enviroment. Experimental results are presented to illustrate the performance.

Keywords: Human Robot Collaboration, Impedance Control, Reinforcement Learning

1 INTRODUCTION

Human robot collaboration has recently received considerable attention in both academic community and technology companies. Its applications are found not only productive in the industrial field, but also profitable and inspiring in the area of daily personal service.[1] In many cases robot and human contact with each other and form a tightly coupled dynamical system to accomplish a task. This is specially characterized as physical human robot interaction (pHRI).[2] Safety and cognition are the two main research fields in terms of pHRI. The safety issue has been intensively studied in the last years and successfully handled by designed pre- and post reactions of collision. The former focuses on obstacle detection and real-time trajectory planning for collision avoidancee. [3, 4] The latter aims to detect collisions based on integrated joint torque sensor as well as the robot dynamic model, then generate reactive motion to avoid injury of human. [5, 6] This technology has been already implemented in several collaborative robots, i.e. ABB Yumi, KUKA iwwa, FRANKA EMIKA panda, etc.

On the other hand, cognition is one skill that many robots are not able to provide jet. Hence, leader-follower scheme is still one of the most widely used strategies in pHRI. Under its concept, the human works as leader and the follower (robots) try to adapt their movements to synchronize with the leader. However, this scheme is not optimal since: 1) extra human effort is required for the coordination of the group motion. 2) the robot only works ”passively” under human guidance and cannot make proactive contribution. 3) sometimes the robot can even make better decisions, especially in the case that the sense of human is interrupted.

One important feature of the cognitive robot is that it can understand human's intension and make proactive contribution in a joint task with human. This concept leads to a shared control strategy [7], where both human and robot ”share” a common task and should contribute control effort together at the same time. As generally known, robots have a flair for precision and repeatability, while human is good at analysis and decision making. Hence, how to allocate and adapt the role of each participant to exploit strength of both human and robot is becoming an interesting research topic. [8].

Contact author: Min Wu [1], Yanhao He [1], Steven Liu [1]
[1]Erwin-Schroedinger Str.12, 67663 Kaiserslautern, Germany
E-mail: mwu@eit.uni-kl.de, yanhao@eit.uni-kl.de, sliu@eit.uni-kl.de

Within the structure of pHRI, contact force is an important interaction modality for coordination/collaboration between human and robot. It should be particularly considered and carefully handled in robot control. Impedance control [9] provides an effective solution since it is able to deal with both motion and force control problem at the same time. Moreover, previous research [10] shows that the human motor control could also be described by an impedance control scheme, in which the impedance parameters are continuously adapted by muscle activities in response to dynamic environments. Based on this assumption, a pHRI scenario could be physically interpreted as a shared impedance control, in which both human and robot are modeled as mechanical impedance and generate force/torque together to influence the dynamic of the payload. Under this concept, the role allocation between participants could be achieved by varying their impedance parameters. For example, participant would dominate the control as long as he/she applies large forces by increasing its impedance. In short, the dynamic role allocation in pHRI could be formulated as an adaptive impedance control problem.

Many previous researches aim to find a proper adaptive mechanism according to different features. In [11] the damping factor is tuned based on the robot end-effector's velocity, i.e. high damping is provided when human performs fine movements at low velocity, while low damping is used for large movements at high velocity. In [12] the control law of robot is adjusted by taking feedback of contact force. In [13] the authors first study the human motor control scheme and then implement in robot to achieve a human-like adaption on the impedance parameters. From our point of view the adaption of the impedance parameters should take both task requirement and human's intention into consideration. Hence, we attend to combine different task-relied and human-relied features together in a cost function and try to solve it as an optimization problem.

Conventional optimal control is usually model-based and not-adaptive. It is not sufficient for pHRI since there exists high uncertainties in human control goals and control mechanisms. Machine learning techniques, especially reinforcement learning (RL), provides promising solution. Under its concept, the control policy of the agent is evaluated by taking feedback from enviroment, which is quantified by a reward function. Then the policy is improved so that the cumulative reward, defined as value function is maximized. In [14], an review of the RL literature and its application to robotics is provided. Recently, a combination of RL and impedance control has been studied in various of research works. In [15], path integral based RL was used to optimize both trajectories and impedance parameters. The key concept is to use dynamic movement primitives to represent parameterized policy, then optimized by RL. In [16] the authors proposed a reinforcement learning based adaptive control in a pHRI task. An Actor-Critic structure with two neural networks is used for estimating human's control policy and updating the robot control. In [17], a cascaded RL based impedance control loop is provided. The task-specific outer-loop aims to learn a optimal impedance model, while the robot-specific inner-loop controller aims to achieve the prescribed impedance behavior. In [18], fuzzy Q-Learning algorithm is used to regulate the desired damping vector by minimizing the jerk of the motion. In all the above mentioned literatures, an accurate model of the agent as well as the enviroment dynamics is not required.

In this work, a novel RL based approach of variable impedance control is presented to enable the robot cognitively interact with human by learning the optimal impedance parameters. Moreover, a human reference estimator is provided so that the robot keeps estimating the desired goal/path of human and adapting its motion reference based on the estimation. Hence, no prior knowledge of the collaboration task is required. The main contributions are listed as follows:

1. Design of an adaptive impedance controller based on the Q- Learning technology. The learning algorithm aims to solve an optimal control problem defined by previous study on the performance indexes of human-robot collaboration. [19]

2. Further adjustment of the learned impedance parameter by taking feedback of the interaction force, which is regarded as the degree of human's acceptance on the robot's control action.

3. Design of a human reference estimator based on a simple linear contact stiffness model.[20] The model parameters are identified on-line through batch least-square method.

The rest of this paper is organized as follows: In section 2 the model and structure of the proposed shared impedance control method are overviewed. In section 3 the approach is discussed in detail and the experimental validation with a FRANKA EMIKA robot are shown in section 4. At the end a conclusion and an outlook of future work are drawn in section 5.

2 SYSTEM MODEL AND PROBLEM FORMULATION

2.1 SYSTEM MODEL

Considering a human-robot collaboration task of jointly transporting a rigid object, the object dynamic is described by the following equation:

$$\boldsymbol{M}_o\ddot{\boldsymbol{x}} + \boldsymbol{C}_o\left(\dot{\boldsymbol{x}}, \boldsymbol{x}\right) = \boldsymbol{G}\boldsymbol{u}, \tag{1}$$

where $\boldsymbol{x}$ describes the object's configuration, $\boldsymbol{M}_o$ is the inertia Matrix, $\boldsymbol{C}_o\left(\dot{\boldsymbol{x}}, \boldsymbol{x}\right)$ represents the centrifugal and gravity term, $\boldsymbol{G} = (\boldsymbol{G}_h, \boldsymbol{G}_r)$ represents the grasping matrix,

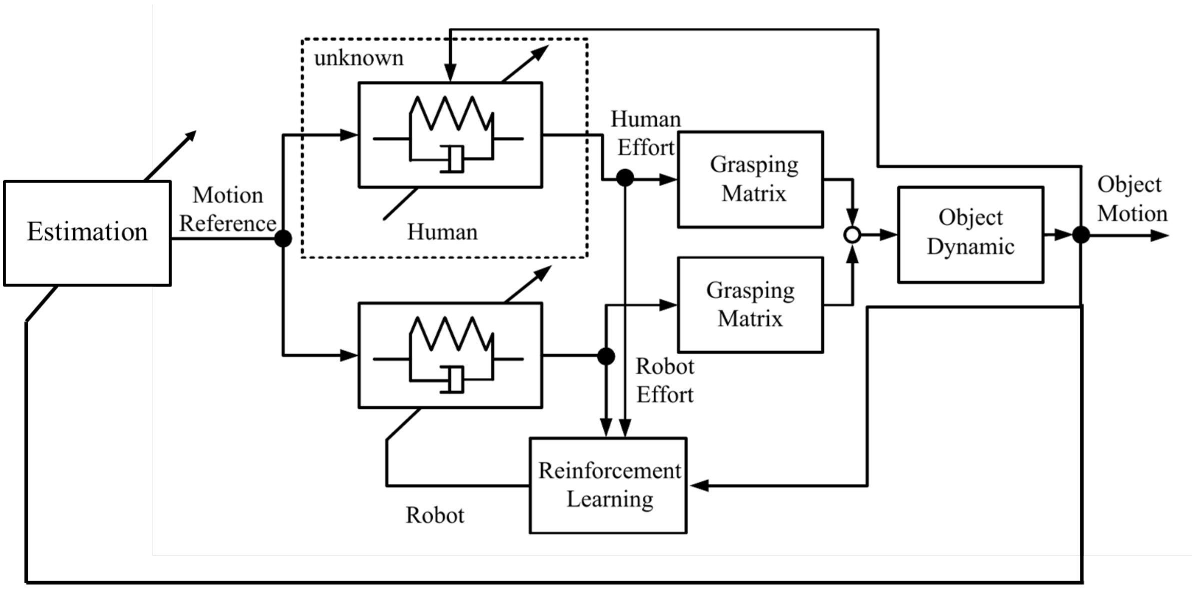


Figure 1 Overview of the shared control system

$\boldsymbol{u} = (\boldsymbol{u}_h, \boldsymbol{u}_r)^T$ contains the generated wrench of both human and robot.
In order to simplify further analysis, following boundary conditions are made:

- Only a translational motion in 2D-plane is considered. The impedance parameters in other degrees of freedom (DOF) are set high and not adapted.
- Grasping of the object is rigid.
- Measurement or estimation of motion and contact force at the robot end-effector is provided.
- The object dynamic is holonomic and the control input has redundancy.

The whole system is modeled as a shared impedance control scheme (see Figure 1), in which human and robot are both regarded as mechanical impedance (contains spring and damper) and jointly generate required wrench to influence the object dynamic. The robot end-effector dynamic is described by the following impedance model in Cartesian space:

$$\boldsymbol{M}_d\ddot{\tilde{\boldsymbol{x}}}_r + \boldsymbol{D}_d\dot{\tilde{\boldsymbol{x}}}_r + \boldsymbol{K}_d\tilde{\boldsymbol{x}}_r = \boldsymbol{f}_h, \quad (2)$$

where $\boldsymbol{M}_d, \boldsymbol{D}_d$ and $\boldsymbol{K}_d$ are the desired mass, damping and stiffness, $\tilde{\boldsymbol{x}}_r$ represents the motion error of the robot control and $\boldsymbol{f}_h$ represents the wrench generated by human.
The human is modeled as a PD- type controller, so that:

$$\boldsymbol{D}_h\dot{\boldsymbol{x}}_h + \boldsymbol{K}_h\tilde{\boldsymbol{x}}_h = \boldsymbol{f}_h, \quad (3)$$

where $\boldsymbol{K}_h$ and $\boldsymbol{D}_h$ are the human control gains and $\tilde{\boldsymbol{x}}_h$ represents the human control error.

2.2 PROBLEM FORMULATION
The control task for the human-robot team is formulated as: generating proper control effort $\boldsymbol{u}$ to achieve an accurate, efficient and smooth collaboration, which aims to 1) maintain the desired object motion, 2) keep the internal force acts on the workpiece small, 3) reduce the total energy consumption. Based on these performance indexes, a global linear quadric optimization problem is defined as follows:

$$\min_{\boldsymbol{u}} \int_0^\infty \left(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u}\right) \mathrm{d}t, \quad (4)$$

where $\boldsymbol{X} = (\tilde{\boldsymbol{x}}_o, \dot{\boldsymbol{x}}_o)^T$, $\boldsymbol{Q}, \boldsymbol{R} \succ 0$ are positive definite weighting matrices. Since the cost function is influenced by the control effort of both human and robot, the collaboration could be regarded as a cooperative optimal control for a multi-agent system.
Furthermore, a local optimization problem for the robot could be formulated as follows:

$$\min_{\boldsymbol{u}_r} \int_0^\infty \left(\boldsymbol{X}_r^T\boldsymbol{Q}_r\boldsymbol{X}_r + \boldsymbol{u}_r^T\boldsymbol{R}_r\boldsymbol{u}_r + \boldsymbol{f}_h^T\boldsymbol{R}_h\boldsymbol{f}_h\right) \mathrm{d}t, \quad (5)$$

where $\boldsymbol{X}_r = (\tilde{\boldsymbol{x}}_r, \dot{\boldsymbol{x}}_r)^T$. Note that by minimizing (5), tracking error, energy consumption and human effort are minimized.
Analytically the solution of this optimal control problem could be determined by solving the Hamilton-Jacobian-Bellman (HJB) equation. However, complete knowledge of the system dynamic is required. Even through it could be approximately described by (1), (2) and (3), the model parameters still remain unknown, especially the human control gains. Moreover, these parameters may even change over time, which means that the HJB

equation needs to be recursively solved on-line. With regard to these issues, a model-free RL algorithm, namely Q-learning is used in this work for solving the optimization problem without full knowledge of the system model. Furthermore, Q- learning is beneficial to handling problems with stochastic transitions.[21] The algorithm iteratively evaluates the control policy by examining the action value function (Q- function) and update the policy along the gradient of Q- function until convergence. Within the structure of impedance control, it yields an optimal allocation of the impedance parameters.

As mentioned before, the optimal problem in (4) is solved by human and robot cooperatively. Besides, their dynamics are strongly coupled by motion constraints. Hence, the control action of the robot will cause changes in human control as well. In several previous researches, people attend to identify human's impedance parameter [13], cost function [22] and control effort [16] on-line in order to have more knowledge on how they could be influenced by robot control, and consider this counter-effect into optimization problem as well. However, due to high uncertainty in human model, it can not be guaranteed that the learned parameters completely match to human's behavior. In our approach, instead of estimating human control mechanisms, a negotiation strategy is proposed by directly taking feedback of the contact force as a measurement of human disagreement. In pHRI, it is commonly agreed that the interaction force plays an important role in human intention recognition. In the negotiation phase, the impedance parameters that determined by RL are further modified to suppress the interaction force.

So far the proposed method only enables adaption of the impedance parameters. On the other hand, the desired path of the robot keeps constant. A significant drawback is that if human attends to change the desired goal, the robot will always try to go back to its original path so that the human needs to make a correction all the time. In order to reduce the human effort, the robot should be able to estimate the desired path of human. As mentioned before, since human and robot are modeled as mechanical impedances and coupled by the object dynamic, a simple linear model could be used to describe the contact dynamic. Based on this model, the human reference position is estimated by batch least-square method.

3 APPROACH

In this section first the RL algorithm for learning the optimal impedance parameter is introduced. Then the estimation of the human reference path is presented. Finally, these two phases are summarized in one algorithm.

3.1 VARIABLE IMPEDANCE CONTROL BASED ON Q-LEARNING

Firstly, under the concept of RL, a cost function (negative reward) is defined as follows:

$$r\left(\tilde{\boldsymbol{x}}_r, \dot{\boldsymbol{x}}_r, \boldsymbol{u}\right) = \tilde{\boldsymbol{x}}_r^T \boldsymbol{Q}_1 \tilde{\boldsymbol{x}}_r + \dot{\boldsymbol{x}}_r^T \boldsymbol{Q}_2 \dot{\boldsymbol{x}}_r + \boldsymbol{u}_r^T \boldsymbol{R}_r \boldsymbol{u}_r, \tag{6}$$

where $\tilde{\boldsymbol{x}}_r = \boldsymbol{x}_r - \boldsymbol{x}_d$, are diagonal positive definite weighting matrices. Note that in comparison to (5), the human force $\boldsymbol{f}_h$ is removed. The main reason is that according to the definition, reward function relies on system state and control action. However, we argue that the human force cannot be regarded as a state variable of the robot dynamic. Its dependence on other robot states is not completely clear. Hence, in this work the human force is not considered in the learning phase but handled separately in another way, which be introduced later.

Correspondingly, the value function $V\left(\boldsymbol{X}_r\right)$, which represents the conditional expected value of the future cost when starting in state at time t by applying control input $\boldsymbol{u}_r$, is written as:

$$V\left(\boldsymbol{X}_r\right) = \int_t^{\infty} r\left(\boldsymbol{X}_r, \boldsymbol{u}_r\right) \mathrm{d}\tau. \tag{7}$$

In [21], it is classified as an integral RL problem, which could be transformed in to a standard RL problem by discretizing the following continuous-time Bellman equation:

$$0 = r\left(\boldsymbol{X}_r\right) + \dot{V}^{\boldsymbol{u}_r}. \tag{8}$$

Using Euler method, a discrete-time equation is obtained:

$$\begin{aligned} &0 = r\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) + \frac{V(k+1) - V(k)}{\Delta t} \\ &\Leftrightarrow V(k) = r(k) \cdot \Delta t + V(k+1), \end{aligned} \tag{9}$$

which could be used for value update.

According to [14], V is specifed as a state-value function and usually used when the environment model is known. For unknown enviroment model, as in our case, an action-value function (Q- function) is preferred, which is defined as:

$$\begin{aligned} Q\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) = {}& r\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) \cdot \Delta t \\ &+ Q\left(\boldsymbol{X}_r(k+1), \boldsymbol{h}\left(\boldsymbol{X}_r(k+1)\right)\right), \end{aligned} \tag{10}$$

where $\boldsymbol{u}_r = \boldsymbol{h}\left(\boldsymbol{X}_r\right)$ defines the control policy.

Under the concept of impedance control, the robot end-effector dynamic could be approximately represented by a linear discrete-time state-space model, such as:

$$\boldsymbol{X}_r(k+1) = \boldsymbol{A}_d \boldsymbol{X}_r(k) + \boldsymbol{B}_d \boldsymbol{u}_r(k), \tag{11}$$

where

$$\boldsymbol{A}_d \approx \boldsymbol{I} + \boldsymbol{A}\Delta t = \boldsymbol{I} + \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I} \\ -\boldsymbol{M}_d^{-1}\boldsymbol{K}_d & -\boldsymbol{M}_d^{-1}\boldsymbol{D}_d \end{pmatrix} \Delta t, \tag{12}$$

$$\boldsymbol{B}_d \approx \left(\boldsymbol{I} + \frac{\Delta t}{2}\boldsymbol{A}\right) \Delta t \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{M}_d^{-1}. \end{pmatrix} \tag{13}$$

Based on the superposition principle of linear systems, the Q- function can be written as:

$$\begin{aligned}
&Q\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) \\
&= \left(\boldsymbol{X}_r^T(k)\boldsymbol{Q}_r\boldsymbol{X}_r(k) + \boldsymbol{u}_r^T(k)\boldsymbol{R}_r\boldsymbol{u}_r(k)\right)\cdot \Delta t \\
&+ \left(\boldsymbol{A}_d\boldsymbol{X}_r(k) + \boldsymbol{B}_d\boldsymbol{u}_r(k)\right)^T \boldsymbol{P}\left(\boldsymbol{A}_d\boldsymbol{X}_r(k) + \boldsymbol{B}_d\boldsymbol{u}_r(k)\right) \\
&= \begin{pmatrix}\boldsymbol{X}_r(k)\\ \boldsymbol{u}_r(k)\end{pmatrix}^T \underbrace{\begin{pmatrix}\boldsymbol{S}_{xx} & \boldsymbol{S}_{xu}\\ \boldsymbol{S}_{xu}^T & \boldsymbol{S}_{uu}\end{pmatrix}}_{\boldsymbol{S}} \begin{pmatrix}\boldsymbol{X}_r(k)\\ \boldsymbol{u}_r(k)\end{pmatrix}
\end{aligned} \tag{14}$$

The aim of Q- learning is to find the minimum of Q-function. According to (14), it is equivalent to optimize the $\boldsymbol{S}$ matrix so that an optimal co-relation between all the state- and input variables is achieved.
Since an analytical solution of the optimal $\boldsymbol{S}$ is hard to achieve, the value-function-approximation (VPA) technology is used. (14) could be reformulated as follows:

$$Q\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) = \boldsymbol{\mathcal{S}}^T\boldsymbol{\varphi}\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right), \tag{15}$$

where

$$\boldsymbol{\varphi}\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) = \begin{pmatrix}\boldsymbol{X}_r(k)\otimes\boldsymbol{X}_r(k)\\ 2\boldsymbol{X}_r(k)\otimes\boldsymbol{u}_r(k)\\ \boldsymbol{u}_r(k)\otimes\boldsymbol{u}_r(k)\end{pmatrix}, \tag{16}$$

$$\boldsymbol{\mathcal{S}} = \begin{pmatrix}\boldsymbol{S}_{xx} & \boldsymbol{S}_{xu} & \boldsymbol{S}_{uu}\end{pmatrix}^T. \tag{17}$$

Using VPA, the Q-learning problem is solved by minimizing the temporal deference (TD) error of the Q- function, defined as

$$\begin{aligned}
e(k) = &- Q\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) + r\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) \\
&+ Q\left(\boldsymbol{X}_r(k+1), \boldsymbol{h}(\boldsymbol{X}_r(k+1))\right).
\end{aligned} \tag{18}$$

Since Q is parameterized by $\boldsymbol{\mathcal{S}}$, the above equation is equivalent to:

$$\begin{aligned}
e(k) = &- \boldsymbol{\mathcal{S}}^T\boldsymbol{\varphi}\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) + r\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) \\
&+ \boldsymbol{\mathcal{S}}^T\boldsymbol{\varphi}\left(\boldsymbol{X}_r(k+1), \boldsymbol{u}_r(k+1)\right).
\end{aligned} \tag{19}$$

Then an optimization problem of minimizing the TD error is defined as:

$$\min_{\boldsymbol{\mathcal{S}}} \frac{1}{2}e^2(k) \tag{20}$$

The gradient decent method can be used for solving this optimization problem. The gradient of the cost function in respect to parameter set $\boldsymbol{\mathcal{S}}$ is calculated as:

$$\begin{aligned}
\frac{\partial \frac{1}{2}e^2(k)}{\partial \boldsymbol{\mathcal{S}}} &= e(k)\frac{\partial e(k)}{\partial \boldsymbol{\mathcal{S}}} \\
&= \boldsymbol{\Phi}(k)\left(\boldsymbol{\mathcal{S}}^T\boldsymbol{\Phi}(k) - r(k)\right),
\end{aligned} \tag{21}$$

where

$$\boldsymbol{\Phi}(k) = \boldsymbol{\varphi}\left(\boldsymbol{X}_r(k), \boldsymbol{u}_r(k)\right) - \boldsymbol{\varphi}\left(\boldsymbol{X}_r(k+1), \boldsymbol{u}_r(k+1)\right). \tag{22}$$

The parameter set $\boldsymbol{\mathcal{S}}$ is then updated until convergence, following:

$$\boldsymbol{\mathcal{S}}^+ \leftarrow \boldsymbol{\mathcal{S}} - \boldsymbol{l}\otimes\left(\boldsymbol{\Phi}(k)\left(\boldsymbol{\mathcal{S}}^T\boldsymbol{\Phi}(k) - r(k)\right)\right), \tag{23}$$

where the vector $\boldsymbol{l}$ represents the step size and its elements are all positive.
Note that in this work, it is assumed that motions in different DOFs are independent from each other, so that the parameter matrices $\boldsymbol{S}_{xx}, \boldsymbol{S}_{xu}$ and $\boldsymbol{S}_{uu}$ are all diagonal, which simplifies the leaning process.
As long as the minimum of Q value is reached, the optimal control input $\boldsymbol{u}_r^\star$ is determined by solving:

$$\frac{\partial Q^\star\left(\boldsymbol{X}_r, \boldsymbol{u}_r\right)}{\partial \boldsymbol{u}_r^\star} \equiv \boldsymbol{0} \tag{24}$$

It yields:

$$\begin{aligned}
\boldsymbol{u}_r^\star &= -\boldsymbol{S}_{uu}^{\star-1}\boldsymbol{S}_{xu}^{\star T}\boldsymbol{X}_r \\
&= -\boldsymbol{S}_{uu}^{\star-1}\boldsymbol{S}_{xu}^{\star T}\begin{pmatrix}\boldsymbol{x}_r - \boldsymbol{x}_d\\ \dot{\boldsymbol{x}}_r\end{pmatrix}
\end{aligned} \tag{25}$$

which has the same form as a mechanical spring-damper system. Hence the RL phase could be concluded as: to learn an optimal virtual impedance parameter set $\{\boldsymbol{K}_d^\star, \boldsymbol{D}_d^\star\}$ (both are sub-matrices of $\boldsymbol{S}_{uu}^{\star-1}\boldsymbol{S}_{xu}^\star$), which generates an elastic force to 1) cancel the motion error, 2) achieve sufficient damping, 3) minimize the whole control energy.
Stability: according to the basis theory of impedance control [9], the robot dynamic is stable only if $\boldsymbol{K}_d^\star$ and $\boldsymbol{D}_d^\star$ are positive definite and bounded. To satisfy this condition, a more tightened constraint is defined in the learning phase, i.e. $\boldsymbol{S}_{xx}, \boldsymbol{S}_{xu}$ and $\boldsymbol{S}_{uu}$ are all positive definite and bounded.
As mentioned before, the impedance parameters are determined by solving the local optimization problem of the robot. Hence, there is no guarantee that they can completely satisfy the human's expectation. As discussed in previous study [12], interaction force contributes significantly to the sensory communication in pHRI and could be used as measurement of "disagreement". If user keeps applying large force, it is assumed that he/she indents to make a correction. Based on that, a further adjustment of the impedance parameter, named as "negotiation", is performed. The aim is to decrease the impedance parameters as the contact force increases. The reason is that, from both physical and psychological point of view, acceptable forces are crucial at first glance in pHRI.[19] Hence the robot should not generate large resistant force on human. Under this concept, for each relevant DOF, a discount factor α_i is

defined as follows:

$$\alpha_i = \begin{cases} 1 - \frac{\|f_i\|}{f_{t,i}}, & \text{if } \|f_i\| < f_{t,i} \\ 0. & \text{else} \end{cases} \tag{26}$$

where f_i represents the interaction force on i-th DOF, and the force threshold $f_{t,i} > 0$ is a design parameter. Then the robot stiffness is modified as:

$$\hat{K}_i = (K_i^\star)^{\alpha_i}. \tag{27}$$

The new damping is computed by:

$$\hat{D}_i = D_i^\star \sqrt{\frac{\hat{K}_i}{K_i}}, \tag{28}$$

to maintain the learned optimal damping factor of the equivalent end-effector dynamic.
At last, depending on the robot dynamic equation, taking $\boldsymbol{M}_d = \boldsymbol{I}$, the joint control input is determined by:

$$\begin{aligned} \boldsymbol{\tau} = &\boldsymbol{M}(\boldsymbol{q})\boldsymbol{J}^{-1}(\boldsymbol{q}) \left(-\hat{\boldsymbol{K}}\tilde{\boldsymbol{x}}_r - \hat{\boldsymbol{D}}\dot{\boldsymbol{x}}_r - \dot{\boldsymbol{J}}(\dot{\boldsymbol{q}}, \boldsymbol{q})\dot{\boldsymbol{q}} \right) \\ &+ \boldsymbol{C}(\dot{\boldsymbol{q}}, \boldsymbol{q})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}), \end{aligned} \tag{29}$$

where $\boldsymbol{\tau}$ is the vector of input joint torques, $\boldsymbol{q}$ is the vector of joint angles, $\boldsymbol{M}$ is the inertia matrix, $\boldsymbol{C}$ is the matrix describing centrifugal and Coriolis terms, $\boldsymbol{g}$ is the vector of gravitational moments, $\boldsymbol{J}$ is the Jacobian matrix that describes the velocity kinematic.

3.2 HUMAN REFERENCE ESTIMATION

In order to make the robot flexible to the change of human's intention, a human reference estimator is proposed. In general, human reference estimation is difficult, since there exists large uncertainties in the human motor control system, especially in the case of free motion. However, in pHRI, since human and robot usually need to maintain a task specified group formation, their dynamics are tightly coupled by kinematic constraints. The uncertainty then becomes smaller and the human motion is more predictable. In this work, we aim to export the human reference, namely the desired path, from the coupled dynamic of human and robot. As presented in [20], a simple second-order linear system (see Figure 2) could be used to represent the contact dynamic between human and robot. The coupled dynamic is described as follows:

$$\boldsymbol{K}_h(\boldsymbol{x}_{h,d} - \boldsymbol{x}) - \boldsymbol{D}_h\dot{\boldsymbol{x}} = -\boldsymbol{M}_d\ddot{\boldsymbol{x}} - \boldsymbol{D}_d\dot{\boldsymbol{x}} + \boldsymbol{K}_d(\boldsymbol{x}_d - \boldsymbol{x}) \tag{30}$$

Then the human reference position $\boldsymbol{x}_{h,d}$ becomes:

$$\begin{aligned} &\boldsymbol{x}_{h,d} = \\ &\boldsymbol{K}_h^{-1}\left(-\boldsymbol{M}_d\ddot{\boldsymbol{x}} + (\boldsymbol{D}_h - \boldsymbol{D}_d)\dot{\boldsymbol{x}} + (\boldsymbol{K}_h - \boldsymbol{K}_d)\boldsymbol{x} + \boldsymbol{K}_d\boldsymbol{x}_d\right), \end{aligned} \tag{31}$$

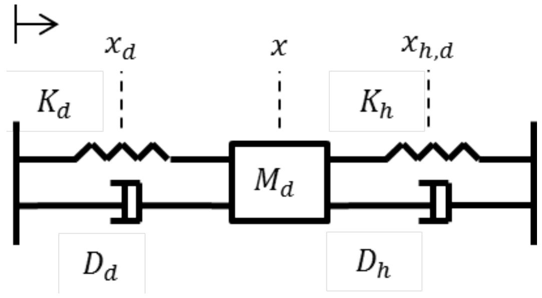


Figure 2 Second-order system formed by human and impedance controlled robot

which can be formulated as a linear function:

$$\boldsymbol{x}_{h,d} = \boldsymbol{\Psi}\boldsymbol{\beta}, \tag{32}$$

where

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{x} & \dot{\boldsymbol{x}} & \ddot{\boldsymbol{x}} \end{pmatrix} \tag{33}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{K}_h^{-1}\boldsymbol{K}_d\boldsymbol{x}_d \\ \boldsymbol{K}_h^{-1}(\boldsymbol{D}_h - \boldsymbol{D}_d) \\ \boldsymbol{K}_h^{-1}(\boldsymbol{K}_h - \boldsymbol{K}_d) \\ -\boldsymbol{M}_d \end{pmatrix} \tag{34}$$

The parameter set $\boldsymbol{\beta}$ could be identified by using least-square method, such that:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\boldsymbol{x}_{h,d}. \tag{35}$$

Note that (35) only requires measurement of motion data and no force information is needed. However, one problem is that in practice the human reference $\boldsymbol{x}_{h,d}$ is a latent variable and cannot be directly measured. To deal with this issue, we make an assumption that the future position of human at the next sampling time equals to the current reference position. i.e. $\boldsymbol{x}_{h,d}(k) = \boldsymbol{x}_h(k+1)$, which means the human motor control is always able to compensate the motion error. Apparently it may not be completely true in reality, the error should not be extremely large, especially within a small time range. Based on the identified parameter set $\hat{\boldsymbol{\beta}}$ and the new coming measurement $\boldsymbol{\Psi}^\star$, the estimated human reference is computed as:

$$\hat{\boldsymbol{x}}_{h,d} = \boldsymbol{\Psi}^\star\hat{\boldsymbol{\beta}}. \tag{36}$$

3.3 ALGORITHM

As in all the proposed adaptive impedance control for the robot is summarized in Algorithm 1.

4 EXPERIMENTAL VALIDATION

4.1 HARDWARE

As a proof of concept, the proposed adaptive impedance control method was tested in a human-robot collaboration

Algorithm 1 Proposed adaptive impedance control

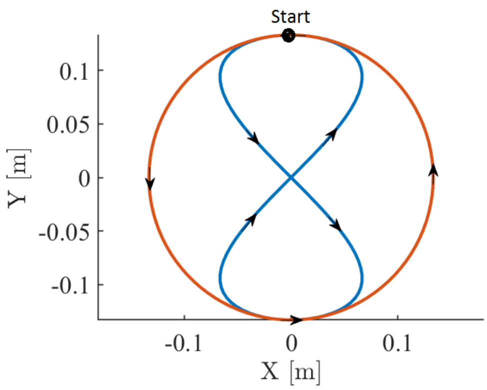
Input: measurement of motion $\boldsymbol{X}_k$, interaction force $\boldsymbol{f}_k$ at current time k
Output: robot torque control input $\boldsymbol{\tau}_{k+1}$ for the next step $k+1$
Reference generation
if 10 measurements collected **then**
  Calculate parameter set $\hat{\boldsymbol{\beta}}$ using (35)
end if
Calculate estimated human reference $\hat{\boldsymbol{x}}_{h,d}$ using (36)
Calculate robot reference $\boldsymbol{x}_d$ based on $\hat{\boldsymbol{x}}_{h,d}$
Q- learning
Calculate $r(k)$ and $\boldsymbol{\Phi}(k)$
while $\|\boldsymbol{\mathcal{S}}^+ - \boldsymbol{\mathcal{S}}\| \geq \varepsilon$ **do**
  $\boldsymbol{\mathcal{S}}^+ \leftarrow \boldsymbol{\mathcal{S}} - \boldsymbol{l} \otimes \left(\boldsymbol{\Phi}(k)\left(\boldsymbol{\mathcal{S}}^T\boldsymbol{\Phi}(k) - r(k)\right)\right)$
end while
Calculate robot impedance parameters based on $\boldsymbol{\mathcal{S}}$
Adjust the impedance parameters using (27), (28)
Calculate the control input $\boldsymbol{\tau}_{k+1}$ in (29)

task with a FRANKA EMIKA robot, which has 7-DOFs and a two-finger gripper (see Figure 3(a)). The Cartesian position and velocity of the end-effector was determined based on the measured joint motion as well as the forward- and velocity kinematics. A direct measurement of the acceleration was not available. Usually it could be calculated by differentiating the velocity signal, which, however, could cause large noises and errors. In this work, a Kalman-filter with *constant acceleration model* [23] was implemented to achieve better estimation of the acceleration. The interaction force was estimated using joint torque measurement and the robot dynamic model. All the above mentioned physical parameters were recorded and calculated with SI base units. The robot was controlled by an external PC through EtherCAT network with a commutation rate of 1 kHz . The proposed learning algorithm and the low-level robot joint control were executed in two separate threads with sampling frequency of 100 Hz and 1 kHz respectively.

4.2 EXPERIMENTAL SETUP

In the experiment, the human operator held the end-effector and followed a 2D reference path (Figure 3(b)) along the pre-specified direction, which is marked by the black arrows. Firstly the human operator was asked to track the inner ”eight” curve (blue). Then in order to further test the flexibility of the proposed control algorithm, the human was asked to switch to the outer circle path (red), when the marked starting point was reached for the second time. Functions of the both paths are shown as follows:

(a)



(b)

Figure 3 a) picture of the experimental setup with a FRANKA EMIKA robot, b) reference path

$$\text{outer path:} \begin{cases} x = 0.133\sin(2\pi i), \\ y = 0.133\cos(2\pi i), \end{cases} \tag{37}$$

$$\text{inner path:} \begin{cases} x = 0.133\sin(2\pi i)\cos(2\pi i), \\ y = 0.133\sin(2\pi i), \end{cases} \tag{38}$$

where $i \in [0, 1]$.

Note that the reference path was not given to the robot, except the starting point. As long as the human tried to move the end-effector, the robot started to estimate the human reference position based on least squares method (see (36)) and used it as its own reference for the inner adaptive impedance control loop. The parameter set $\hat{\boldsymbol{\beta}}$ in (35) would be updated once 10 measurements of the set $\boldsymbol{\Psi}$ were gathered.

For the inner adaptive impedance control loop, the initial parameter sets of the learning algorithm are given as follows:

$$\boldsymbol{Q}_1 = \boldsymbol{Q}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \boldsymbol{R} = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.02 \end{pmatrix}$$

$$\boldsymbol{S}_{xx} = \boldsymbol{I}^{4\times4}, \boldsymbol{S}_{uu} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix},$$

$$\boldsymbol{S}_{xu} = \begin{pmatrix} 10 & 0 & 2.8 & 0 \\ 0 & 10 & 0 & 2.8 \end{pmatrix}$$

Totally 8 participants joined in the experimental study. For comparison three different control methods were tested. Besides the proposed method, a conventional impedance controller with low stiffness and damping was used. In this case the robot behaved compliantly so that the human would dominate the collaboration. Then an adaptive impedance controller with human reference estimation was tested. Note that compare to the proposed method, only the position reference was adapted in this setup, while the stiffness and damping were kept constant and high. The proposed method, as presented in Section 3, contained both reference and impedance adaption. The constant impedance parameter sets used in the other two setups are given as follows:

$$\boldsymbol{K}_{d,low} = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \boldsymbol{D}_{d,low} = \begin{pmatrix} 6.32 & 0 \\ 0 & 6.32 \end{pmatrix},$$

$$\boldsymbol{K}_{d,high} = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}, \boldsymbol{D}_{d,high} = \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix}.$$

All participants repeated the task there times for each control setup. Total error of the path following, total execution time and total force produced by human operator were chosen as performance index for evaluation. Moreover, user experience was also considered. After performing all the measurements, the participants were asked the following two questions:

Q1 *Which control setup provides the best comfortableness?*

Q2 *Which control setup provides the best assistance?*

4.3 RESULTS

Fig. 4 shows the average and variance of all the above mentioned performance indexes. In Fig. 4(a), the integrated tracking error was approximately determined by calculating the area of the reference- and actual path. It can be seen that impedance control with human reference estimation (for both constant and variable impedance parameters) significantly reduced the tracking error. This result shows the importance of making proactive contribution in performance improvement for human-robot collaboration, which, from another point of view, shows that the human-dominated role allocation strategy is not optimal. By taking a further look in Fig. 5(a), especially when the human operator performed the outer circular path, large error and vibration could be observed. In the other two setups, since the robot was able to generate some assistive force based on the estimation of human's motion trend, the tracking performance became much better (Fig. 5(c) and Fig. 5(b)). However, the tracking performances of both setups are close to each other, i.e. the impedance adaption did not contribute much to the error reduction.

By comparing the execution time of one single demonstration (Fig. 4(b)), one can see that the proposed method provided the highest time efficiency.

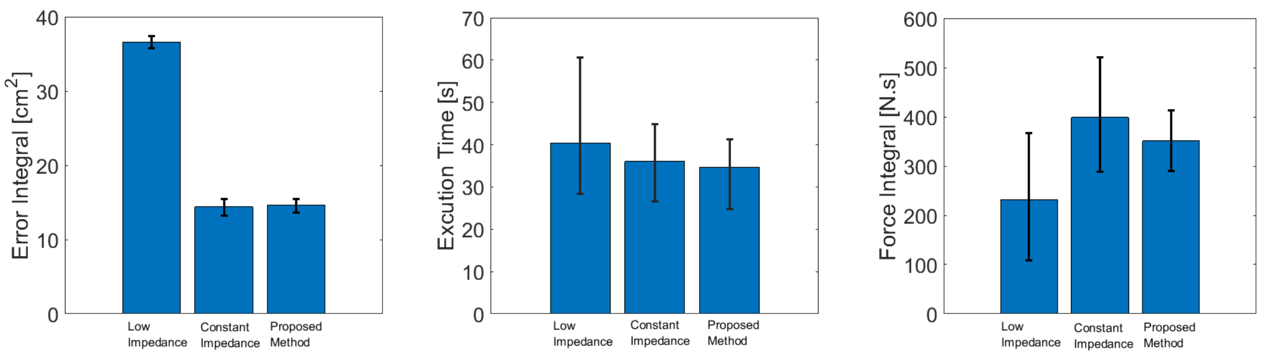
The integral of the interaction force over time could be regarded as measurement of total human effort. As shown in Fig. 4(c), the standard impedance control with low stiffness and damping required the lowest human effort, since the robot behaved compliantly and was easy to be guided. Comparing the other two setups, the proposed impedance control with adaptive parameter required less human force in total.

Fig. 6 shows the plots of force and position in x-direction over time from one single demonstration. With the standard impedance control, even though the total value is smallest, a large peak can be seen when the human started to follow the outer circular path (25 s - 30 s). In the other two setups with human reference estimation, this peak value was significantly reduced, due to the active contribution made by the robot. The time history of force was similar to position, i.e. the peaks occurred when human intended to change the moving directions. This could be further improved enhancing accuracy of the human reference estimation. If the robot correctly recognizes the human's intention and behaves as human expected, no motion correction by the human is necessary so that the human effort could be even lower.

Table I - Statistic of the user experience. Setup 1: standard impedance control with low stiffness and damping; Setup 2: impedance control with human reference estimation and constant parameters; Setup 3: proposed method, impedance control with human reference estimation and variable parameters.

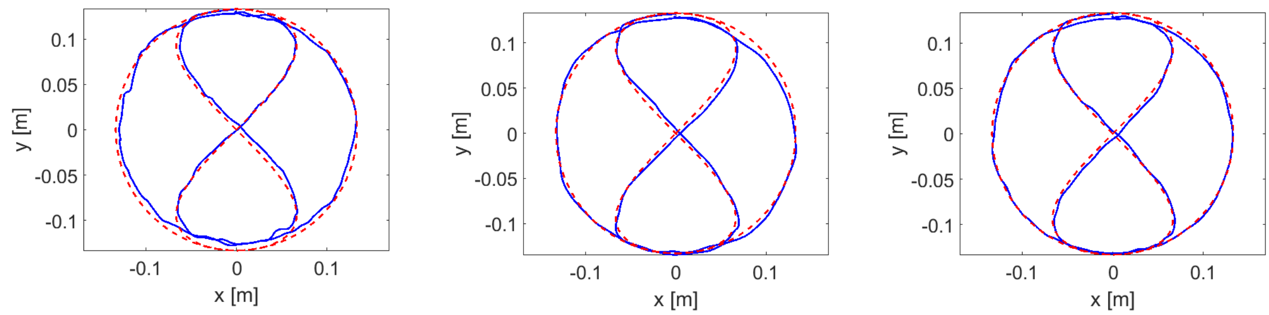
	Comfortableness	Assistance
Setup 1	1	1
Setup 2	2	3
Setup 3	5	4

Additionally, a survey of user experience about all three setups is summarized in Tab. I. 5 of 8 participants chose the proposed method (setup 3) as ”the most comfortable one”, since the robot was ”light” and could provide ”active contribution” as well. Regarding to ”the best assistance”, 4 of 8 chose the proposed method, since they thought the robot ”jointly moved with human” and ”ran smoothly”. On the other hand, 3 of 8 chose the setup with constant impedance parameter set. The participants explained that due to the impedance adaption, sometimes the robot became ”unstable”



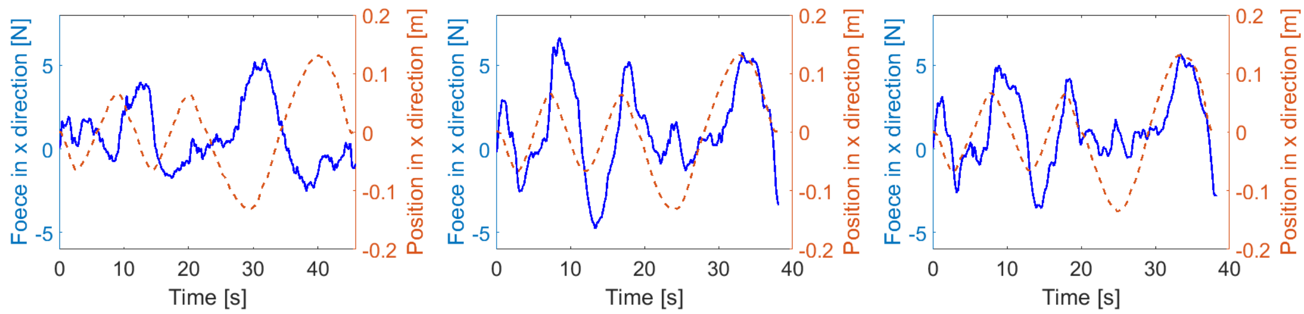
(a) Integrated path tracking error (b) Execution time of one demonstration (c) Intergrated force produced by human

Figure 4 Multiple comparison of the performance indexes between all three setups: 1) standard impedance control with low stiffness and damping; 2) impedance control with human reference estimation and constant parameters; 3) proposed method: impedance control with human reference estimation and variable parameters.



(a) Standard impedance control (b) Impedance control with human reference estimation (c) Impedance control with human reference estimation and adaptive parameters

Figure 5 Path tracking performance of all three setups. The dashed- and solid line represent the reference- and real path respectively.



(a) Standard impedance control (b) Impedance control with human reference estimation (c) Impedance control with human reference estimation and adaptive parameters

Figure 6 Time History of the interaction force and position in x-direction from one single demonstration. The dashed- and solid line represent the position- and force respectively.

and "hard to control". Figure 7 shows how the stiffness dynamically changed over time with the proposed method. This was, however, not preferred by some users. A smooth adaption of the impedance parameter, which also relied on the learning rate of the algorithm, should be considered in the future work.

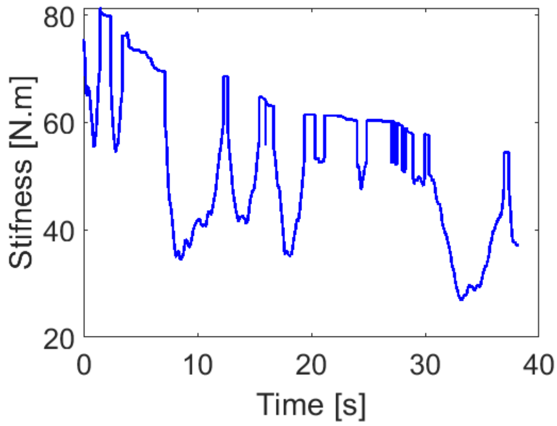


Figure 7 Adaption of the robot stiffness in x-direction performed by the proposed method from one single demonstration

5 CONCLUSION

This paper presented an adaptive impedance control strategy for physical human-robot collaboration. We started our analysis by modeling both human and robot as mechanical impedances, which were coupled by the object's dynamic. Then the collaboration task was formulated as a cooperative optimization problem. Due to unknown system dynamic and human control behavior, analytical solution was impossible to achieve. Hence, a model-free reinforcement learning algorithm was investigated. The proposed control method contained both reference and impedance adaption. Under our concept, no priori information of the task was given to robot. The robot followed human's guidance and estimated its reference path based on the contact dynamic and linear regression method. The control parameter adaption was performed by solving a local optimization problem through Q-learning technology. The interaction force was regarded as a measurement of human disagreement so that the learned impedance parameters were further adjusted .

The proposed method was validated by experiment with 8 participants. As comparison two extra setups were tested, namely 1) a conventional impedance control with low impedance parameters, 2) impedance control with human reference estimation and constant impedance parameters. Total tracking error, execution time and interaction force were chosen as performance indexes. The results showed that the proposed method achieved the best overall performance. Integrating human reference estimation in the robot control played significant role in reducing both tracking error and execution time. Moreover, with on-line impedance adaption the total human effort could be reduced. According to the user experience, the proposed method provided the highest comfortableness. However, the dynamic changing of the impedance was not preferred by all the users, which raises the question *when* and *how often* the robot should adapt its assistive behavior.

Future work aims to enhance the performance of human reference estimation. We believe that a better estimation/prediction of the human motion will further reduce the human effort, and on the robot's side, the number of exploration during learning as well. Moreover, the interaction behavior between human and robot will be further studied, not only limited in motion level, but also in strategy and cognition level.

ACKNOWLEDGEMENT

We would like to thank apl. Prof. Dr.-Ing. Daniel Görges for his helpful advice and comments.

REFERENCES

[1] Goodrich Michael A. and Schultz A.C., Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*, Vol. 1, No. 3, pp. 203-275, 2007.

[2] Ajoudani A., Zanchettin A.M., Ivaldi S., Albu-Schäffer A., Kosuge K. and Khatib O., Progress and prospects of the human-robot collaboration. *Autonomous Robots*, Vol. 42, No. 5, pp. 957-975, 2018.

[3] Mainprice J., Sisbot E.A., Siméon T. and Alami R., Planning safe and legible hand-over motions for human-robot interaction. *IARP/IEEE-RAS/EURON workshop on technical challenges for dependable robots in human environments*. Toulouse, France. 2010.

[4] Pérez-D'Arpino C. and Shah J.A., Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification. *IEEE international conference on robotics and automation*, Seattle, WA, USA, pp. 6175-6182, 2015.

[5] Luca A.D., Albu-Schäffer A., Haddadin S. and Hirzinger G. Collision detection and safe reaction with the DLR-III lightweight manipulator arm. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, pp. 1623-1630, 2006.

[6] Haddadin S., *Towards safe robots: approaching Asimov's 1st law*. Springer, 2011.

[7] Mörtl A., Lawitzky M., Kucukyilmaz A., Sezgin M., Basdogan C. and Hirche S. The role of roles: Physical cooperation between humans and robots. *The International Journal of Robotics Research*, Vol. 31, No. 13, pp. 1656–1674, 2012.

[8] Jarrassé N., Sanguineti V. and Burdet E., Slaves no longer: review on role assignment for human–robot joint motor action. *Adaptive Behavior*, Vol. 22, No. 1, pp. 70–82, 2014.

[9] Hogan N., Impedance Control: An approach to manipulation: Part II-implementation. *ASME International*

Journal of Dynamic Systems, Measurement, and Control, Vol. 107, No. 1, pp. 8-16, 1985.

[10] Svinin M., Goncharenko I., Lee H. and Yamamoto M., Modeling of human-like reaching movements in the manipulation of parallel flexible objects. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Daejeon, Korea, pp. 4830-4835, 2016.

[11] Ficuciello F., Villani L. and Siciliano B., Variable impedance control of redundant manipulators for intuitive human-robot physical interaction. *IEEE Transactions on Robotics*, Vol. 31, No. 4, pp. 850-863, 2015.

[12] Kucukyilmaz A., Sezgin T.M. and Basdogan C, Conveying intentions through haptics in human-computer collaboration. *IEEE World Haptics Conference*, Istanbul, Turkey, pp. 421-426, 2011.

[13] Yang C., Ganesh G., Haddadin S., Parusel S., Albu-Schaeffer A., and Burdet E., Human-like adaptation of force and impedance in stable and unstable interactions. *IEEE Transactions on Robotics*, Vol. 27, No. 5, pp. 918-930, 2011.

[14] Khan S.G., Herrmann G., Lewis F.L., Pipe T. and Melhuish C., Reinforcement learning and optimal adaptive control: An overview and implementation examples. *Annual Reviews in Control*, Vol. 36, No. 1, pp. 42-59, 2012.

[15] Buchli J., Stulp F., Theodorou E., and Schaal S., Learning variable impedance control. *The International Journal of Robotics Research*, Vol. 30, No. 7, pp. 820–833, 2011.

[16] Li Y., Tee K.P., Yan R., Chan W.L. and Wu Y., A framework of human–robot coordination based on game theory and policy iteration. *IEEE Transactions on Robotics*, Vol. 32, No. 6, pp. 1408-1418, 2016.

[17] Modares H., Ranatunga I., Lewis F.L. and Popa D.O., Optimized assistive human–robot interaction using reinforcement learning. *IEEE Transactions on Cybernetics*, Vol. 46, No. 3, pp. 655-667, 2016.

[18] Dimeas F. and Aspragathos N., Reinforcement learning of variable admittance control for human-robot co-manipulation, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Hamburg, Germany, pp. 1011-1016, 2015.

[19] Schmidtler J. and Bengler K., Fast or accurate? - Performance measurements for physical human-robot collaborations. *Procedia Manufacturing*, Vol. 3, pp. 1387-1394, 2015.

[20] Erickson D., Weber M., and Sharf I., Contact stiffness and damping estimation for robotic systems. *The International Journal of Robotics Research*, Vol. 22, No. 1, pp. 41–57, 2003.

[21] Lewis F., Vrabie D. and Syrmos, V., *Optimal control*. 3rd edition, John Wiley & Sons, 2012.

[22] Köpf F., Inga J., Rothfuß S., Flad M. and Hohmann S., Inverse reinforcement learning for identification in Linear-Quadratic dynamic games. *20th IFAC World Congress*, Toulouse, France, Vol. 50, No. 1, pp. 14902-14908, 2017.

[23] Li X.R. and Jilkov V.P., Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39, No. 4, pp. 1333-1364, 2003.