2017 IEEE-RAS 17th International Conference on
Humanoid Robotics (Humanoids)
Birmingham, UK, November 15-17, 2017

Robots Learning from Robots: A Proof of Concept Study for Co-Manipulation Tasks

Luka Peternel and Arash Ajoudani

***Abstract*— In this paper we study the concept of robots learning from collaboration with skilled robots. The advantage of this concept is that the human involvement is reduced, while the skill can be propagated faster among the robots performing similar collaborative tasks or the ones being executed in hostile environments. The expert robot initially obtains the skill through the observation of, and physical collaboration with the human. We present a novel approach to how a novice robot can learn the specifics of the co-manipulation task from the physical interaction with an expert robot. The method consists of a multi-stage learning process that can gradually learn the appropriate motion and impedance behaviour under given task conditions. The trajectories are encoded with Dynamical Movement Primitives and learnt by Locally Weighted Regression, while their phase is estimated by adaptive oscillators. The learnt trajectories are replicated by a hybrid force/impedance controller. To validate the proposed approach we performed experiments on two robots learning and executing a challenging co-manipulation task.**

I. Introduction

Biological systems provide a remarkable source of inspiration for the development of robot skills necessary to operate in unstructured and unpredictable environments. In particular, the planning and control of robotic manipulation and interaction have been strongly influenced by the human motor performance to achieve a seamless robotic physical interaction behaviour [1], [2].

A well-known approach for the robots to acquire human-like skills is based on the learning from the human demonstration [3]. Due to the generality and flexibility of the learning by demonstration techniques, they are considered as intuitive alternatives to the classical trajectory planning approaches in the robotics research community.

Within the context of robot interaction control, much of the learning by demonstration studies observe and learn high-level task representations, such as manipulator end-effector trajectories or forces, while executing a task [4]. The learnt trajectories are then tracked by a pre-selected robot controller (e.g., position, impedance or force). This step requires involvement of a human expert to ensure the chosen control framework and its parameters can faithfully reproduce the learnt task. To reduce the robot's dependency on human in this aspect, more recent techniques have been proposed that enable the robot to select and develop appropriate low-level control modalities autonomously from multiple human observations [5]–[7].

Authors are with HRI[2] Lab, Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy, Email: luka.peternel@iit.it

This work was supported in part by the H2020 project SOMA: Soft-bodied intelligence for Manipulation (645599). The authors would like to thank Dr. Wansoo Kim for the assistance during the experiments.

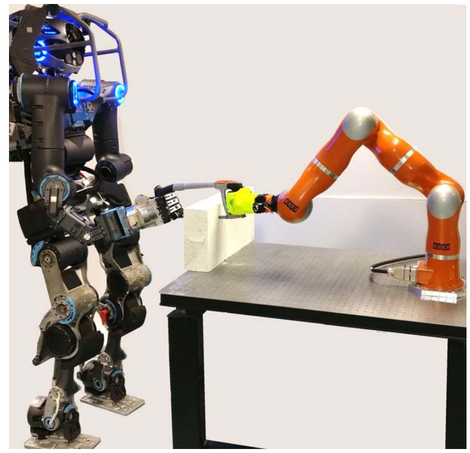

Fig. 1: The concept illustration: robots learning from collaborating expert robots. In such scenarios, novice robots with different morphology or sensing capabilities can obtain the basic knowledge of the task from a cloud or by a direct replication of the control code. Accordingly, to enhance the task execution performance, the control parameters of the novice robot can be tuned based on online observations in the task (Cartesian) space while collaborating with an expert robot. Note that this is a concept illustration and the proof-of-concept experiments in this paper included two KUKA lightweight robots.

Our recent work in this direction presented a method to autonomously devise an appropriate control framework from human demonstrations without a prior knowledge of the demonstrated task [7]. The method analyses the patterns and consistency in the sensory data obtained from repeated observations and determines appropriate control modalities in each Cartesian axis of the robot. Therefore, the high-level learning of the task trajectories can be based on each axis' controller input (e.g. position trajectories for the position/impedance controlled axes, and force trajectories for the force controlled axes).

The above-mentioned methods provide the robot with the capabilities to obtain the skill for autonomous execution of interactive tasks. However, many tasks require involvement of more than one agent for their successful execution. An interdisciplinary application field that can benefit from such robot capabilities is human-robot collaboration, in which the task execution is shared among the human and the robot partner(s) [6], [8]–[11]. In this scenario, one of the most important aspects of the robot is the ability to adapt to the human behaviour. The robot can use feedback from different sensory systems to detect the human behaviour and intention and directly control the collaborative actions through an interaction controller [12]–[16]. The control policy can also be learnt from the collaborating human [**?**], [10], [17], [18]. In addition, gradual adaptation [17], [18] or reinforcement

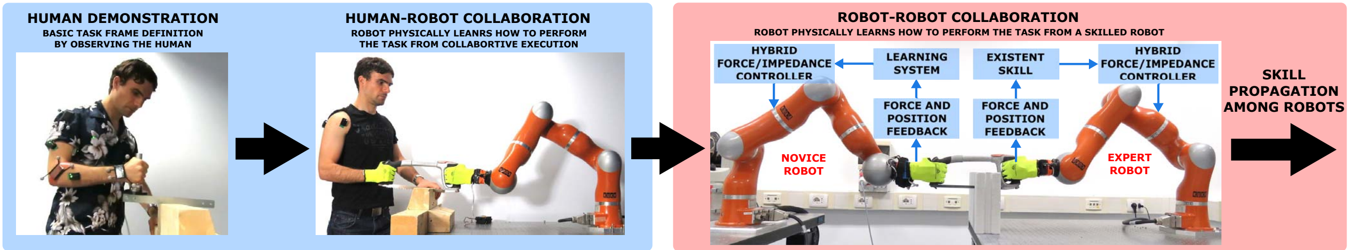

Fig. 2: The entire skill transfer chain contains three segments. The blue area shows the skill acquisition for an expert robot, while the red area shows the proposed method of novice robot learning from the expert robot. Note that the red area is the study of this paper. In the first segment the expert robot obtains a basic knowledge of the task by observing the human, which allows it to define the task frame control framework. In the second segment the expert robot uses that basic knowledge to collaborate with the human and to learn the task specifics under the given conditions through physical interaction. In the final segment the expert robot performs the task with another (novice) robot so that it can learn the skill to execute the task. The skill can be propagated in the same manner among other robots, while the human involvement (blue area) may not be necessary any more.

learning [19] can be employed to correct the learnt interactive skill.

Although human-robot collaboration is essential for effective execution of various tasks, there may be constrains on human involvement (e.g., hazardous or hostile working environment, etc.) or constraints on time in which the skill should be propagated among the numerous robots. In this case, non-proficient robots should be capable to learn from collaborating skilled robots and therefore propagate the existing robotic skill among themselves in a faster manner.

Some aspects of the skill, such as basic control framework [7], can be transferred to the novice robot[1] without physically performing the task (e.g., downloaded from the cloud). However, such specific skill can be difficult to generalise for different tasks and conditions. This is similar to human learning, where only limited degree of skill can be obtained through the observation of another human, while the full skill can only be acquired through physically performing the given task under given conditions. In the same manner, it is better for the novice robot to learn the specific skill while actually performing the task so that the skill can be optimised to the task and environmental conditions.

Reinforcement learning techniques [19], [20] can be used by the novice robot to obtain the refined skill on its own through trial-and-error. This is usually achieved by maximising some notion of cumulative reward. However, the definition of a reward function in more complex co-manipulation tasks is problematic and may lead to a biased physical interaction behaviour [21].

To address some of the above-mentioned issues, in this work, we explore a novel concept of novice robot learning from a collaborating expert robot by demonstration (concept is illustrated in Fig. 1). The expert robot obtains the skill from the human based on the methods we previously proposed: basic task framework is derived from multiple human observations [7] and high-level trajectories are obtained through physical human-robot task execution [18]. The step beyond the state-of-the-art is the propagation of robotic skill from the expert robot to the novice robot to replicate a similar human-robot collaboration performance under the new conditions in a robot-robot setting. The robot physical interaction behaviour is governed by a hybrid Cartesian force/impedance controller. This consideration is to keep the observation and tuning of the parameters in the Cartesian space so that the skill transfer can be potentially propagated among robots with different morphology or sensing capabilities. The novice robot learns task specifics under given conditions directly from physical interaction with the expert robot through an on-line multi-stage learning process that can learn and reproduce reference trajectories and impedance behaviour.

We study and validate the concept of robot learning from expert robots with experiments on a collaborative sawing task. This task offers challenging conditions in terms of both motion and stiffness coordination between the two partners, in addition to the rough interaction with the unstructured environment [9], [16]. The implementation of a hybrid force-impedance controller in this work requires a good torque control capacity, therefore two torque-controlled KUKA lightweight robots are used in the experiments that share a similar characteristics (see Fig. 2, right side). Nevertheless, as mentioned above, since the observation and tuning of the task and control parameters are performed in the task (Cartesian) space, different robots with torque control modality and adequate number of degrees of freedom (depending on the task constraints) can be used instead.

II. Method

The entire process of skill transfer is shown in Fig. 2. The expert robot gained the very basic knowledge of the task by observing the human performing the given task [7]. This basic knowledge includes the definition and configuration of hybrid force/impedance controller that controls the physical interaction behaviour. The expert robot then uses the defined task frame control framework to collaborate with the human and learns how to perform the co-manipulation task (trajectories, impedance behaviour, etc.) under given conditions through physical interaction [18]. The skill transfer from the expert robot to the novice robot is done by the proposed novel method that consists of three consequent stages: reference motion learning stage, impedance behaviour learning stage and collaboration on equal terms stage. The block scheme of the proposed robot-robot learning approach is shown in the red section of Fig. 2.

[1]We refer to the robot that posses the skill for task execution as "expert" and to the robot that does not posses the skill as "novice".

A. Hybrid Force/Impedance Controller

Using the hybrid force/impedance controller, the robot can simultaneously control both contact force and motion of the robot end-effector in different axes of the Cartesian space. This setup is especially useful in tasks involving complex interactions with an unstructured environment [5], [16], [22].

The Cartesian behaviour of each robot was controlled through the interaction force

$$\boldsymbol{F}_{int} = \boldsymbol{F}_{for} + \boldsymbol{F}_{imp}, \tag{1}$$

where $\boldsymbol{F}_{int}$ is the interaction force/torque acting from the robot on the environment, $\boldsymbol{F}_{for}$ interaction force component related to the force production task (e.g., keeping the tool in contact with the environment) and $\boldsymbol{F}_{imp}$ is the interaction force component related to motion control through the impedance. We controlled the $\boldsymbol{F}_{for}$ by a PI controller using the force feedback by

$$\boldsymbol{F}_{for} = \boldsymbol{K}_P^F \boldsymbol{e}_F + \boldsymbol{K}_I^F \int \boldsymbol{e}_F \mathrm{d}t, \tag{2}$$

$$\boldsymbol{e}_F = \boldsymbol{F}_d - \boldsymbol{F}_a, \tag{3}$$

where $\boldsymbol{e}_F$ is the error between the desired force $\boldsymbol{F}_d$ and actual force $\boldsymbol{F}_a$. If some desired force is to be controlled in a certain axis, the diagonal matrices $\boldsymbol{K}_P^F$ and $\boldsymbol{K}_I^F$ should contain the desired gains of PI controller for that axis. For axes with specified impedance control the respective diagonal elements of $\boldsymbol{K}^F$ should be set to zero.

We controlled the impedance by

$$\boldsymbol{F}_{imp} = \boldsymbol{K}\left(\boldsymbol{x}_d - \boldsymbol{x}_a\right) + \boldsymbol{D}\left(\dot{\boldsymbol{x}}_d - \dot{\boldsymbol{x}}_a\right), \tag{4}$$

where $\boldsymbol{x}_d$ is the reference and $\boldsymbol{x}_a$ is the actual pose of the robot end-effector and $\boldsymbol{K}$ and $\boldsymbol{D}$ are robot Cartesian stiffness and damping matrices.

We controlled the desired interaction force $\boldsymbol{F}_{int}$ in Cartesian space at the robot joint torque level

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) + \boldsymbol{J}_r^T \boldsymbol{F}_{int} = \boldsymbol{\tau}, \tag{5}$$

where $\boldsymbol{\tau}$ is vector of robot joint torques, $\boldsymbol{q}$ is vector of joint angles, $\boldsymbol{J}_r$ is robot arm Jacobian matrix, $\boldsymbol{M}$ is mass matrix, $\boldsymbol{C}$ is Coriolis and centrifugal vector and $\boldsymbol{g}$ is gravity vector. While the desired stiffness matrix was learnt, the damping matrix was calculated based on the current stiffness matrix [23].

B. Skill Acquisition for Expert Robot

The expert robot acquired the basic task control framework by the method we recently proposed in [7]. Using the observations of human task performance, the robot determined what variables (position, force, stiffness) are important in each axis of the task frame and derived the basic desired control strategy. For example, the force controller was delegated to the axis in which the contact with the environment should be maintained. On the other hand, the impedance controller was delegated to the axis in which the position should be controlled.

The reference trajectories specific to the co-manipulative task under given conditions were learnt through physical collaborations with the human [18]. The motion trajectories were directly recorded by the robot sensory system, while the stiffness behaviour was observed through the human muscle activity interface [16]. Obtained reference trajectories and stiffness behaviour were then used by the hybrid force/impedance controller to perform the task in a collaborative manner with another agent.

The stiffness behaviour that the expert robot learnt from the human encapsulates the leader/follower role allocation in different phases of the task execution. For example in collaborative sawing, when the first agent is stiff to pull the saw, the second agent is compliant to follow the lead. When the saw reaches the end, the roles are reversed and the second agent becomes stiff to pull the saw back to its own end, while the first agent becomes compliant and follows the partner's action.

C. Novice Robot Learning from Expert Robot

In the *first stage* of the robot-robot learning process, the novice robot started as a pure follower and learned the motion through the physical interaction. The expert robot physically guided the novice robot by producing the reference motion of the tool in all phases of the task (e.g., motion of the saw in the sawing axis), while the novice robot only stabilised the saw on it own end. To achieve this, the novice robot was fully compliant, while the expert robot was fully stiff. Both robots produced the desired force to first establish a contact between the saw and the object, and then maintain the force in the cutting direction. Assuming the rigid coupling between the two robots through a tool, the motion produced by the expert robot affects the motion of the novice robot. While the expert robot produced the desired motion, the novice robot learned this motion and encoded it as its own end-effector reference trajectory.

The trajectories were encoded by Dynamical Movement Primitives (DMPs) [24]. We choose DMPs because they have several advantageous properties, such as: can represent both point-to-point and rhythmic movements learnable, are not explicitly dependant on time, are robust to perturbations, can include coupling terms to realise closed-loop reactive behaviours, and can easily be modulated on-line to change various movement characteristic [25]. The pattern of the desired motion f_d was approximated by

$$f_d = \frac{\ddot{x}_d}{\Omega^2} - \alpha\left(\beta\left(-x_d\right) - \frac{\dot{x}_d}{\Omega}\right). \tag{6}$$

where x_d, $\dot{x}_d$ and $\ddot{x}_d$ are the robot motion that we wish to learn and its derivatives, Ω is the execution frequency and $\alpha = 8$ and $\beta = 2$ are positive constants. The estimated desired pattern f_d was used to build a nonlinear shape function of DMP

$$\dot{z} = \Omega\left(\alpha\left(\beta\left(-y\right) - z\right) + f\right), \tag{7}$$

$$\dot{y} = \Omega z, \tag{8}$$

where y is the encoded trajectory, ϕ is the phase of the task, and nonlinear shape function f is defined as

$$f(\phi)=\frac{\sum_{i=1}^{N}\psi_i(\phi)w_i}{\sum_{i=1}^{n}\psi_i(\phi)}, \tag{9}$$

where weights w determine the shape of trajectory and uniformly distributed Gaussian kernels $\psi_i(\phi)$. Kernels were defined as

$$\psi_i(\phi)=e^{h(\cos(\phi-c_i)-1)}. \tag{10}$$

where h determines the width, c_i centres and N number of Gaussian kernels (we selected $N=25$).

We use Locally weighted Regression to learn the encoded trajectories because it enable the fast updates of the existing models and is therefore suitable for online learning [26]. Each weight w_i of kernel ψ_i was updated using recursive least squares method [27]

$$w_i(t+1) \quad = \quad w_i(t)+\psi_i P_i(t+1) r e_r(t), \tag{11}$$

$$e_r(t)=f_d(t)-w_i(t)r, \tag{12}$$

$$P_i(t+1)=\frac{1}{\lambda}\left(P_i(t)-\frac{P_i(t)^2r^2}{\frac{\lambda}{\psi_i}+P_i(t)r^2}\right), \tag{13}$$

where λ is forgetting factor that controls the forgetting rate of older data and was set to 0.995 for our experiments. Initial setting of parameters was $r=1$, $w_i(0)=0$ and $P_i(0)=1$ for $i=1,2,...,N$.

The aim of the *second stage* of the learning process was for the novice robot to learn the desired impedance behaviour that encodes the leader/follower role allocation in different phases of the task. Unlike in [16], [18], where this information was directly passed from the human to the collaborating robot through the muscle activity interface, in the proposed approach the novice robot learns the stiffness behaviour by observing the physical interaction.

After the first stage, the expert robot stopped being fully stiff and started to produce the desired stiffness behaviour throughout different phases of the collaborative task (e.g., half of period stiff and half of period compliant as in human-robot collaborative sawing task). The proposed impedance learning approach prescribes the novice robot to observe the actual motion of the tool with respect to the reference motion trajectory learnt in the previous (first) stage. If the desired motion trajectory is followed in a certain phase of the task, the novice robot assumes that it is the expert robot's turn to execute the collaborative task and therefore remains compliant. On the other hand, if the desired motion trajectory is not followed, the novice robot assumes that it is its moment to take over the execution of the task in that phase and therefore the impedance is increased. The stiffness part of the impedance learning is defined by

$$K_{d*}(\phi)=\begin{cases}k_{high} & \text{if } |e(\phi)|\geq e_{th},\\ k_{low} & \text{if } |e(\phi)|< e_{th},\end{cases} \tag{14}$$

$$e(\phi)=x_{d*}(\phi)-x_a(\phi), \tag{15}$$

where K_{d*} is the learnt desired stiffness behaviour[2] in a certain axis, e is the error between the desired end-effector position x_{d*} learnt in the first stage and x_a is actual position measured by the sensors, k_{high} and k_{low} define the stiffening behaviour, and ϕ is the phase of the task. We controlled the phase of the task by adaptive oscillators [28]. The threshold e_{th} determines when the novice robot should be stiff or compliant. The learnt stiffness $K_{d*}(\phi)$ from (14) controls the appropriate element in the stiffness matrix $\boldsymbol{K}$ of the impedance control law (4).

In the *third stage* of the learning process, the novice robot became an expert robot and used the learnt reference motion and stiffness trajectories to collaboratively perform the task with the expert robot on equal terms. The task execution skill can then be further propagated to other novice robots in the same manner.

III. Experiments

In the experiments we demonstrated the proposed approach by performing a collaborative task involving material (brick and wood) sawing using a two-person saw. Both robots held the saw by the Pisa/IIT Softhand [29]. Please refer to the supplementary multimedia file for a video of experiments. This task requires a good coordination of motion and impedance between the two agents. Since perfect matching between motions of the two agents is practically unfeasible, they should ideally not be stiff at the same time to prevent opposing each other. For example, when the first agent is pulling the saw, the second should remain compliant and wait until the saw reaches the edge. At that time, the roles are switched and the second agent should become stiff to pull the saw, while the first should remain compliant not to oppose the effort. Therefore, during the first stage of learning the novice robot is completely compliant, while the expert robot is completely stiff, and in the second and third stage, the stiffening/complying is reciprocally exchanged.

The sawing motion axis was in x-axis for both robots, which was intersecting the line in which the material should be cut apart. The stiffness parameters in this axis were set to $k_{high}=1100$ N/m and $k_{low}=0$ N/m. The y-axis was aligned along the beam and was kept compliant (0 N/m) as the saw was stabilised by the environment in that direction. The cutting was performed in z-axis and both robots controlled the contact force in this axis. The reference force for PI controller (3) of both robots was set to $F_d=-5$ N.

The transitions between learning stages were determined by time. The duration of the first stage was set to 10 seconds. The duration of the second stage was also set to 10 seconds. The third stage essentially had no time limitation, as the robots started to collaborate on equal terms after the learning was completed. The threshold parameter of the stiffness behaviour learning system was set to $e_{th}=0.02$ m.

The results of the first experiment are shown in Fig. 4. In this experiment the sawing was performed with execution

[2]Low-pass filter can be used to smoothen the switching steps between the high and low stiffness values. Since DMP encoding acts similarly to low-pass filter, we did not use any additional filtering.

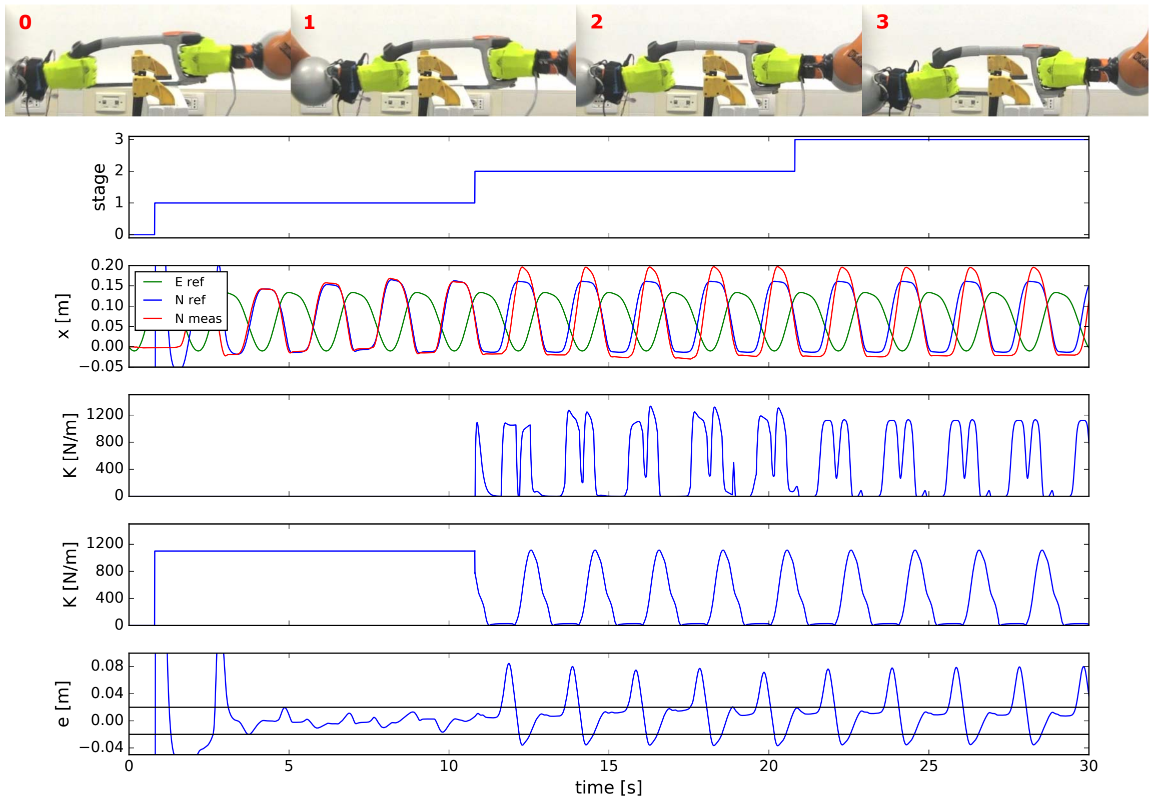

Fig. 3: Results of robot-robot collaboration experiment on two-person sawing task. The sequence of photos in top shows the sawing progress throughout different stages. The first graph shows learning process stages as described in Section II-C. The second graph shows the sawing motion in x-axis. The green line represents the reference motion of the expert robot, the blue line represents the reference motion of the novice robot and the red line represents the actual motion. The third graph shows the stiffness of the novice robot. The fourth graph shows stiffness of the expert robot. The fifth graph shows the error between the reference and actual motion used in the stiffness learning process. The thresholds e_{th} are drawn by black lines.

frequency of 0.5 Hz. The first graph shows the transitions between the learning stages. The second graph shows the motion of both expert and novice robot in x-axis. We can see that in the first stage, only the expert robot produced the motion of the saw by being fully stiff throughout the entire period of the task (see the fourth graph). Meanwhile, the novice robot was learning the desired motion from the motion induced by the expert robot through the rigid coupling (i.e. through the saw). To enable this, the novice robot was fully compliant (see the third graph).

After the sawing motion was learnt, the process entered the second stage where the novice robot was learning the stiffness behaviour. In this stage, the robots used the learning strategy defined by (14) and (15). The expert robot now produced the phase-dependent stiffening pattern that was learnt from the collaboration with the human. The novice robot observed the error between the learnt reference motion in the previous (first) stage and the measured actual motion of the tool (see the fifth graph). If the tool was moving according to the learnt reference trajectory, the novice robot assumed it had to be compliant in that phase as the expert robot was in the lead. In the opposite case, when the tool was not moving according to the reference trajectory, the novice robot assumed it was its turn to take the lead and therefore increased the stiffness to maintain the reference motion.

In the final stage, both robots collaborated on equal terms as the novice robot became proficient in performing the given co-manipulative task. We performed five full learning trials and calculated the work each robot was producing in sawing axis during each period by $W_x = \int K_x(x_d - x_a)dx$. The average ratio (across all trials) between the work produced by the expert robot and the work produced by the novice robot ($\frac{W_E}{W_N}$) in the third stage was 0.66, where a value close to one implies an equal contribution. This indicates that after the learning (in the third stage) the load was shared between the two collaborating robots compared to first stage when only the expert robot produced the work (novice robot was fully compliant). If more equal contribution is desired one can tune the parameter e_{th} that influences the contribution of the novice robot.

We can also see that the learnt stiffness profile of novice robot includes a slight overlap with the stiffness profile of the expert robot. Such overlapping could be avoided if additional constraint was used in the stiffness learning strategy defined by (14) and (15). In this case, information about the direction of motion can be used from the velocity to determine whether the actual motion is lagging behind the reference or overtaking the reference, and then only stiffen when the actual motion is lagging. This condition is defined

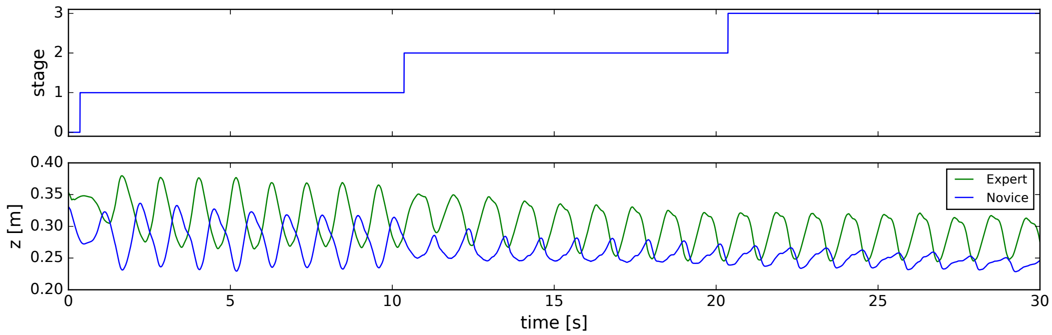

Fig. 4: Results of robot-robot collaboration experiment on two-person sawing task with higher execution frequency. The first graph shows the learning stages. The second graph shows the motion in the cutting direction (z-axis) for expert robot (green line) and novice robot (blue line).

as

$$K_{d*}(\phi) = \begin{cases} k_{high} & \text{if } |e(\phi)| \geq e_{th} \ \& \ e(\phi) \cdot \dot{x}_{ref}(\phi) > 0, \\ k_{low} & \text{otherwise,} \end{cases} \tag{16}$$

where $\dot{x}_{ref}$ is the reference velocity of the tool. However, such additional constraint was not necessary as the task was successfully learnt and collaboratively performed by the general approach defined by (14) and (15).

The results of the second experiment are shown in Fig. 4. In this experiment we used a different frequency of task execution (0.9 Hz). Please refer to the supplementary multimedia file for a video of the experiments. In this case, we want to show the difference between the task performances in different stages (see the first graph) of the learning process. The second graph shows the motion of the saw in the direction of the cutting (z-axis). The motion fluctuation in z-axis was decreased for both robots and the cutting was improved when the novice robot obtained the skill (see second and third stages). This indicates an improved collaborative task performance.

IV. Discussion

We presented and experimentally validated the method where a novice robot can learn co-manipulation skills from an expert robot. The results showed that the task performance was improved when the appropriate behaviour was learnt compared to the initial learning stage. The proposed approach can complement other robot learning methods, such as: direct transfer of skill / pre-programmed behaviour between robots, separate human demonstration and individual reinforcement learning. Direct transfer of the skill / pre-programmed behaviour is fast and easy but it can be susceptible when dealing with different robots and environments. In the proposed method the novice robot learns the task from the expert robot by physically performing it, therefore the skill can be optimised for specific conditions. The same can be achieved by human demonstration and reinforcement learning. The advantage of human demonstration is that it can be more effective due to the human supervision, however it requires human involvement, which can be unfeasible in some cases (e.g., hazardous environment, etc.).

One of the advantages of the proposed robot-robot learning method is that it does not require human involvement after the expert robot acquires the skill. This advantage is shared with reinforcement learning, where novice robot can acquire the new skill on its own after the human defines the cost function. However, the reinforcement learning may be difficult in collaborative scenarios where the expert robot simultaneously influences the task. In such cases the exact contribution of the novice robot's action on the task performance may be unclear and the novice robot learning may converge to a biased or an undesirable policy.

The proposed method is limited to the co-manipulation tasks where the expert and novice robots are physically coupled (e.g., through the tool, etc.), as the expert robot teaches the novice robot through physical interaction. Since the expert robot teaches the novice robot the physical interaction task in Cartesian space, the joint space behaviour is not directly learnt and is determined by the local hybrid controller. Therefore, in the current stage, the method is not capable of teaching subtasks in the joint space, such as: obstacle avoidance, etc.

In the present framework the motion learning and the impedance learning are done in two separate stages. In future we will attempt to couple the two stages into a single stage. In this direction, probabilistic skill encoding [10], [11], [30] can be potentially utilised.

In the current analysis the metric of the learnt skill quality was based on the task-specific parameters: motion fluctuation and the cutting speed in the direction of cutting (z-axis). The future research direction will be to find a more general quality metric in terms of task performance. Such general metric can then potentially be used to monitor and control the learning progress in a more adaptive manner.

In this preliminary study we investigated a challenging task of collaborative sawing that requires good coordination between the involved agents and complex physical interaction where force and impedance must be controlled in different axes at the same time. In future, we will test the approach on other tasks, such as collaborative bolt turning, which also involves the physical coupling of agents through the tool [16]. In addition, other robots will be used, such as WALK-MAN humanoid as shown in Fig. 1.

REFERENCES

[1] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[2] A. Ajoudani, *Transferring Human Impedance Regulation Skills to Robots*. Springer, 2016.

[3] A. G. Billard, S. Calinon, and R. Dillmann, "Learning from humans," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer International Publishing, 2016, pp. 1995–2014.

[4] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *ICML*, vol. 97, 1997, pp. 12–20.

[5] A. Ureche, K. Umezawa, Y. Nakamura, and A. Billard, "Task parameterization using continuous constraints extracted from human demonstrations," *Robotics, IEEE Transactions on*, vol. 31, no. 6, pp. 1458–1471, Dec 2015.

[6] L. Rozo, D. Bruno, S. Calinon, and D. G. Caldwell, "Learning optimal controllers in human-robot cooperative transportation tasks with position and force constraints," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ Intl. Conf. on*, 2015.

[7] L. Peternel, L. Rozo, D. Caldwell, and A. Ajoudani, "A method for derivation of robot task-frame control authority from repeated sensory observations," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 719–726, Apr 2017.

[8] P. Evrard, E. Gribovskaya, S. Calinon, A. Billard, and A. Kheddar, "Teaching physical collaborative tasks: object-lifting case study with a humanoid," in *IEEE-RAS Intl. Conf. on Humanoid Robots*, 2009, pp. 399–404.

[9] L. Peternel, T. Petrič, E. Oztop, and J. Babič, "Teaching robots to cooperate with humans in dynamic manipulation tasks based on multi-modal human-in-the-loop approach," *Autonomous robots*, vol. 36, no. 1-2, pp. 123–136, Jan 2014.

[10] H. Ben Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *Robotics and Automation (ICRA), 2014 IEEE Intl. Conf. on*, 2014, pp. 2831–2837.

[11] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, vol. 41, no. 3, pp. 593–612, Mar 2017.

[12] R. Ikeura and H. Inooka, "Variable impedance control of a robot for cooperation with a human," in *Robotics and Automation (ICRA), 1995 IEEE Intl. Conf. on*, vol. 3, May 1995, pp. 3097–3102.

[13] K. Kosuge and N. Kazamura, "Control of a robot handling an object in cooperation with a human," in *Robot and Human Communication, 6th IEEE Intl. Workshop on*, Sep 1997, pp. 142–147.

[14] D. Agravante, A. Cherubini, A. Bussy, P. Gergondet, and A. Kheddar, "Collaborative human-humanoid carrying using vision and haptic sensing," in *Robotics and Automation (ICRA), 2014 IEEE Intl. Conf. on*, May 2014, pp. 607–612.

[15] P. Donner and M. Buss, "Cooperative swinging of complex pendulum-like objects: Experimental evaluation," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 744–753, June 2016.

[16] L. Peternel, N. Tsagarakis, and A. Ajoudani, "A human-robot co-manipulation approach based on human sensorimotor information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 7, pp. 811–822, July 2017.

[17] D. Lee and C. Ott, "Incremental kinesthetic teaching of motion primitives using the motion refinement tube," *Autonomous Robots*, vol. 31, no. 2, pp. 115–131, 2011.

[18] L. Peternel, N. Tsagarakis, D. Caldwell, and A. Ajoudani, "Adaptation of robot physical behaviour to human fatigue in human-robot co-manipulation," in *IEEE-RAS Intl. Conf. on Humanoid Robots*, November 2016, pp. 489–494.

[19] I. Palunko, P. Donner, M. Buss, and S. Hirche, "Cooperative suspended object manipulation using reinforcement learning and energy-based control," in *Intelligent Robots and Systems (IROS), 2014 IEEE/RSJ Intl. Conf. on*, Sept 2014, pp. 885–891.

[20] J. Kober, A. Wilhelm, E. Oztop, and J. Peters, "Reinforcement learning to adjust parametrized motor primitives to new situations," *Autonomous Robots*, vol. 33, no. 4, pp. 361–379, Apr. 2012.

[21] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Int. Res.*, vol. 4, no. 1, pp. 237–285, May 1996.

[22] C. Schindlbeck and S. Haddadin, "Unified passivity-based cartesian force/impedance control for rigid and flexible joint robots via task-energy tanks," in *Robotics and Automation (ICRA), 2015 IEEE Intl. Conf. on*, May 2015, pp. 440–447.

[23] A. Albu-Schäffer, C. Ott, U. Frese, and G. Hirzinger, "Cartesian impedance control of redundant robots: recent results with the DLR-light-weight-arms," in *Robotics and Automation (ICRA), 2003 IEEE Intl. Conf. on*, vol. 3, Sept 2003, pp. 3704–3709.

[24] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Learning attractor landscapes for learning motor primitives," in *in Advances in Neural Information Processing Systems*. MIT Press, 2003, pp. 1523–1530.

[25] L. Peternel, T. Noda, T. Petrič, A. Ude, J. Morimoto, and J. Babič, "Adaptive control of exoskeleton robots for periodic assistive behaviours based on EMG feedback minimisation," *PLoS ONE*, vol. 11, no. 2, p. e0148942, Feb 2016.

[26] D. Nguyen-Tuong and J. Peters, "Model learning for robot control: a survey." *Cognitive Science*, vol. 12, no. 4, pp. 319–40, 2011.

[27] S. Schaal and C. G. Atkeson, "Constructive incremental learning from only local information," *Neural Comput.*, vol. 10, no. 8, pp. 2047–2084, Nov. 1998.

[28] T. Petrič, A. Gams, A. J. Ijspeert, and L. Žlajpah, "On-line frequency adaptation and movement imitation for rhythmic robotic tasks," *Int. J. Rob. Res.*, vol. 30, no. 14, pp. 1775–1788, Dec. 2011.

[29] A. Ajoudani, S. Godfrey, M. Bianchi, M. Catalano, G. Grioli, N. Tsagarakis, and A. Bicchi, "Exploring teleimpedance and tactile feedback for intuitive control of the pisa/iit softhand," *Haptics, IEEE Transactions on*, vol. 7, no. 2, pp. 203–215, April 2014.

[30] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Probabilistic movement primitives," in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2616–2624.