Exploiting Interaction Dynamics for Learning Collaborative Robot Behaviors[*]

Leonel Rozo,[1] **João Silvério** [1] **, Sylvain Calinon** [2,1] **and Darwin G. Caldwell** [1]

[1]Department of Advanced Robotics - Istituto Italiano di Tecnologia (Genoa, Italy)

[2]Idiap Research Institute (Martigny, Switzerland)

Abstract

The new generation of smarter and safer robots aimed at assisting humans in industries and homes demands to innovate the ways in which these machines are designed and controlled. In this context, one of the biggest challenges is to empower these collaborative robots with a wide range of learning and adaptation capabilities so that they can easily assist humans in a vast variety of scenarios, ranging from assembly lines to health-care facilities. In this paper we propose to teach a collaborative robot reactive and proactive behaviors that exploit the interaction dynamics between the robot and the user. We call the proposed approach Adaptive Duration Hidden-Semi Markov Model (ADHSMM) that enables the robot to both react to the user actions and lead the task when needed. ADHSMM is used to retrieve a sequence of states governing a trajectory optimization technique that provides the reference and gain matrices to the robot controller. The proposed framework is tested in handover and transportation tasks using a 7 DOF backdrivable manipulator.

1 Introduction

In contrast to industrial robots, which have been caged to keep humans safe and out of harm's way, collaborative robots are being designed to work alongside humans, assisting them with a variety of tasks. This entails a close – physical – interaction between robots and humans, which may take place in highly unstructured environments, such as offices, homes, and healthcare facilities, among others. In consequence, collaborative robots are required to safely interact with their users, adapt to their needs, understand their actions, and be easily trainable to handle short runs of different tasks. Therefore, learning and adaptation capabilities are crucial so that this new generation of robots can perform naturally.

Programming by demonstration (PbD) is a promising solution to teach robots how to carry out a task from several examples given by a teacher [Billard *et al.*, 2008]. This approach can also be used to naturally learn how to collaborate and interact with humans in a large range of tasks and scenarios [Rozo *et al.*, 2016]. In this paper, PbD is exploited to endow a robot with collaborative behaviors (reactive and proactive) that are learned from kinesthetic demonstrations showing the interaction dynamics between robot and user. Specifically, reactive behaviors refer to actions that are conditioned on the interaction with the user, allowing the robot to adapt to the user movements. Proactive behaviors involve taking the lead of the task by carrying out self-initiated actions that exploit the taught knowledge. In order to provide the robot with these behaviors, we propose to learn a model of the collaborative task with a modified version of the Hidden Semi-Markov Model [Yu, 2010] where the duration probability distribution is adapted online according to the interaction, which permits to shape the temporal dynamics of the task as a function of user actions.

Several works have focused on teaching robots collaborative roles that are purely reactive to the partner actions. Amor *et al.*, [2013] proposed to learn separate models of two persons interacting during a collaborative task, encapsulating the adaption of their behaviors to the movements of the respective partner. One of these models was then transferred to the robot so that it is able to autonomously respond to the behavior of its user. Maeda *et al.*, [2015] proposed to use probabilistic interaction primitives [Paraschos *et al.*, 2013] to learn collaborative movements that need to be coordinated with the user actions by exploiting the correlations between human and robot trajectories. At a higher level task representation, Wilcox *et al.*, [2012] proposed an adaptive algorithm for handling HRC tasks where the temporal behavior is adapted online based on the user preferences. Their method is built on dynamic scheduling of simple temporal problems and formulated as a nonlinear program considering person-specific workflow patterns. In contrast to our learning framework, the aforementioned approaches only provide the robot with reactive behaviors, that is, without proactive behaviors learned during the demonstrations of the task.

Other works have exploited PbD to teach collaborative robots follower and leader roles.[1] Evrard *et al.*, [2009] pro-

---

[1]A leader role is considered a proactive behavior since the robot exploits the task knowledge to take the lead during execution.

posed to use Gaussian mixture models (GMM) and Gaussian mixture regression (GMR) to respectively encode and reproduce these types of behaviors. Medina *et al.*, [2011] endowed a robot with a cognitive system providing segmentation, encoding and clustering of collaborative behavioral primitives, that incrementally updated during reproduction. One of the main differences with respect to [Evrard *et al.*, 2009] is that the robot starts behaving as a follower and its role progressively becomes more proactive as it acquires more knowledge about the task. Li *et al.*, [2015] addressed the role allocation problem through a formulation based on game theory. A continuous role adaptation is achieved by modifying the contribution of the human and the robot in the minimization of a linear quadratic cost, according to the disagreement between them. Similarly, Kulvicius *et al.*, [2013] used dynamic movement primitives in HRC where interaction forces were considered. The learning problem was treated as that of finding an acceleration-based predictive reaction for coupled agents, in response to force signals indicating disagreements due to obstacle avoidance or different paths to follow.

Our approach is similar to the foregoing works in the sense that it provides the robot with both reactive and proactive behaviors that are learned from demonstrations. However, unlike [Li *et al.*, 2015] and [Kulvicius *et al.*, 2013], the behavior in the approach that we propose is not a function of the partners disagreement, but instead depends on the interaction dynamics of the task, that is, temporal patterns observed during the demonstration phase, and the way the temporal dynamics is shaped by the interaction with the human partner. Our controller shares similarities with the approach presented by [Li *et al.*, 2015], with the difference that the role allocation is not directly affecting the robot control input, but is instead driven by a linear quadratic regulator. Additionally, our time-independent trajectory retrieval approach provides gain matrices that exploit the variability of the task and shape the robot compliance accordingly.

The rest of the paper is organized as follows: Section 2 presents the proposed framework for learning reactive and proactive collaborative behaviors. Section 3 describes the handover and transportation task, which is used to evaluate the proposed algorithm. Finally, conclusions and future routes of research are given in Section 4.

2 Adaptive Duration Hidden Semi-Markov Model (ADHSMM)

A hidden Markov model (HMM) is characterized by an initial state distribution Π_i, a transition probability matrix $a_{i,j}$, and an emission distribution for each state in the model, commonly expressed as a Gaussian distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. In HMM, the self-transition probabilities $a_{i,i}$ only allow a crude implicit modeling of the number of iterations that we can expect to stay in a given state i before moving to another state. Indeed, the probability of staying d consecutive time steps in a state i follows the geometric distribution (see for example [Rabiner, 1989])

$$\mathcal{P}_i(d) = a_{i,i}^{d-1}(1 - a_{i,i}), \qquad (1)$$

decreasing exponentially with time.
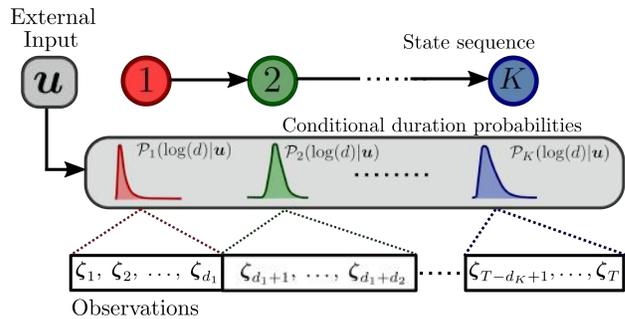
Figure 1: Adaptive duration Hidden-Semi Markov Model.

Variable duration modeling techniques such as the *hidden semi-Markov model* (HSMM) sets the self-transition probabilities $a_{i,i}$ of the HMM to zero, and replaces it with an explicit model (non-parametric or parametric) of the relative time during which one stays in each state, see for example [Yu and Kobayashi, 2006; Zen *et al.*, 2007b].

Since the state duration is always positive, its distribution should preferably be modeled by a function preserving this property. Thus, it is here proposed to use a univariate normal distribution $\mathcal{N}(\mu_i^P, \Sigma_i^P)$ with mean μ_i^P and associated covariance matrix Σ_i^P to model the logarithm of the duration, which is equivalent to the use of a lognormal distribution to fit the duration data. Indeed, if d is lognormally distributed, $\log(d)$ is normally distributed.

In the resulting HSMM, the probability to be in state i at time step t given the partial observation $\boldsymbol{\zeta}_{1:t} = \{\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_t\}$, namely $\alpha_{t,i} \triangleq \mathcal{P}(s_t = i \mid \boldsymbol{\zeta}_{1:t})$, can be recursively computed with (see for example [Rabiner, 1989])

$$\alpha_{t,i} = \sum_{d=1}^{d^{\max}} \sum_{j=1}^{K} \alpha_{t-d,j} \, a_{j,i} \, \mathcal{N}_{d,i}^D \prod_{s=t-d+1}^{t} \mathcal{N}_{s,i},$$

$$h_{t,i} = \frac{\alpha_{t,i}}{\sum_{k=1}^{K} \alpha_{t,k}}, \qquad (2)$$

where $\quad \mathcal{N}_{d,i}^D = \mathcal{N}\big(\log(d) \mid \mu_i^P, \Sigma_i^P\big) \quad$ and $\qquad (3)$

$$\mathcal{N}_{s,i} = \mathcal{N}\big(\boldsymbol{\zeta}_s \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\big).$$

For $t < d^{\max}$, the initialization is given by

$$\alpha_{1,i} = \Pi_i \, \mathcal{N}_{1,i}^D \, \mathcal{N}_{1,i},$$

$$\alpha_{2,i} = \Pi_i \, \mathcal{N}_{2,i}^D \prod_{s=1}^{2} \mathcal{N}_{s,i} + \sum_{j=1}^{K} \alpha_{1,j} \, a_{j,i} \, \mathcal{N}_{1,i}^D \mathcal{N}_{2,i},$$

$$\alpha_{3,i} = \Pi_i \, \mathcal{N}_{3,i}^D \prod_{s=1}^{3} \mathcal{N}_{s,i} + \sum_{d=1}^{2} \sum_{j=1}^{K} \alpha_{3-d,j} \, a_{j,i} \, \mathcal{N}_{d,i}^D \prod_{s=4-d}^{3} \mathcal{N}_{s,i},$$

etc., which corresponds to the update rule

$$\alpha_{t,i} = \Pi_i \, \mathcal{N}_{t,i}^D \prod_{s=1}^{t} \mathcal{N}_{s,i} + \sum_{d=1}^{t-1} \sum_{j=1}^{K} \alpha_{t-d,j} \, a_{j,i} \, \mathcal{N}_{d,i}^D \prod_{s=t-d+1}^{t} \mathcal{N}_{s,i}. \qquad (4)$$

Note that the above iterations can be reformulated for efficient computation, see [Yu and Kobayashi, 2006; Yu, 2010].

2.1 Conditional estimation of duration probability

The explicit duration formulation of HSMM assumes that the duration probability $\mathcal{P}_i(d)$ exclusively depends on how long the system stays in state i. Yamagishi and Kobayashi, [2005] noted that such assumption can have drawbacks in some applications such as in speech synthesis where various speaking styles and/or emotions could influence the duration model. In such case, it looks relevant to consider adaptive duration probability. In [Yamagishi and Kobayashi, 2005; Nose *et al.*, 2007] the authors proposed to express the mean μ_i^D of the duration probability $\mathcal{P}_i(d)$ as an affine function of a style vector, whose parameters were estimated by a maximum likelihood linear regression (MLLR) method [Leggetter and Woodland, 1995]. This approach also showed to improve human walking motion synthesis [Yamazaki *et al.*, 2005].

We propose an adaptive duration hidden semi-Markov model (ADHSMM) in which the duration in every state depends on an external input $\boldsymbol{u}$. Unlike [Yamagishi and Kobayashi, 2005; Nose *et al.*, 2007], we express the duration probability as $\mathcal{P}_i(\log(d) \,|\, \boldsymbol{u})$, obtained from a Gaussian mixture model of K^D components encoding the joint distribution $\mathcal{P}_i(\boldsymbol{u}, \log(d))$ for each state i of the HSMM. We thus obtain a GMM for each state, with parameters

$$
\pi_{i,j}^{\mathcal{D}}, \quad \boldsymbol{\mu}_{i,j} = \begin{bmatrix} \boldsymbol{\mu}_{i,j}^U \\ \boldsymbol{\mu}_{i,j}^D \end{bmatrix}, \quad \boldsymbol{\Sigma}_{i,j} = \begin{bmatrix} \boldsymbol{\Sigma}_{i,j}^U & \boldsymbol{\Sigma}_{i,j}^{UD} \\ \boldsymbol{\Sigma}_{i,j}^{DU} & \boldsymbol{\Sigma}_{i,j}^D \end{bmatrix} \tag{5}
$$
$$
\forall i \in \{1, \dots, K\}, \quad j \in \{1, \dots, K^D\}.
$$

In contrast to [Yamagishi and Kobayashi, 2005; Nose *et al.*, 2007] that only consider an affine relationship between μ_i^D and the input vector, our approach permits to encode more complex nonlinear relationships between the duration of the state and the external parameter.

We also propose to define a maximum duration $d_i^{\max}$ for each state i that depends on the duration probability distribution $\mathcal{P}_i(\log(d) \,|\, \boldsymbol{u})$. Indeed, the maximum allowed duration $d^{\max}$ does not necessarily need to be the same for each state i, see for example [Mitchell *et al.*, 1995]. In the experiments, we used $d_i^{\max} = \exp\left(\mu_i^D + 2\,\Sigma_i^{D\,\frac{1}{2}}\right)$, which means that $\sim 95\%$ of the observed duration for the state i lie within two standard deviations.[2]

Therefore, we compute $\mathcal{N}_{d,i}^D$ in (2) as $\mathcal{P}_i(\log(d) \,|\, \boldsymbol{u}_t) \sim \mathcal{N}(\hat{\mu}_{i,t}^D, \hat{\Sigma}_{i,t}^D)$ with

$$
\hat{\mu}_{i,t}^D = \sum_{j=1}^{K^D} \gamma_{i,j}(\boldsymbol{u}_t)\, \tilde{\mu}_{i,j}^D(\boldsymbol{u}_t), \tag{6}
$$

$$
\hat{\Sigma}_{i,t}^D = \sum_{j=1}^{K^D} \gamma_{i,j}(\boldsymbol{u}_t)\Big(\tilde{\Sigma}_{i,j}^D + \tilde{\mu}_{i,j}^D(\boldsymbol{u}_t)\big(\tilde{\mu}_{i,j}^D(\boldsymbol{u}_t)\big)^{\top}\Big) - \hat{\mu}_{i,t}^D(\hat{\mu}_{i,t}^D)^{\top}, \tag{7}
$$

---

[2]Note that the conditional duration probability is characterized by a mean and a variance lying in the log-transformed space of the duration data, therefore an exponential mapping is needed to define the maximum duration as time steps.

where

$$
\tilde{\mu}_{i,j}^D(\boldsymbol{u}_t) = \mu_{i,j}^D + \boldsymbol{\Sigma}_{i,j}^{DU} \boldsymbol{\Sigma}_{i,j}^U{}^{-1} (\boldsymbol{u}_t - \boldsymbol{\mu}_{i,j}^U), \tag{8}
$$

$$
\tilde{\Sigma}_{i,j}^D = \Sigma_{i,j}^D - \boldsymbol{\Sigma}_{i,j}^{DU} \boldsymbol{\Sigma}_{i,j}^U{}^{-1} \boldsymbol{\Sigma}_{i,j}^{UD}, \tag{9}
$$

$$
\gamma_{i,j}(\boldsymbol{u}_t) = \frac{\pi_{i,j}^{\mathcal{D}} \mathcal{N}(\boldsymbol{u}_t \,|\, \boldsymbol{\mu}_{i,j}^U, \boldsymbol{\Sigma}_{i,j}^U)}{\sum_k^{K^D} \pi_{i,k}^{\mathcal{D}} \mathcal{N}(\boldsymbol{u}_t \,|\, \boldsymbol{\mu}_{i,k}^U, \boldsymbol{\Sigma}_{i,k}^U)}. \tag{10}
$$

When it comes to human-robot collaboration, the proposed formulation can be exploited for learning reactive and proactive behaviors. On the one hand, ADHSMM encodes the temporal patterns and sequential information observed during the demonstration phase through its duration probabilities and transition matrix. This feature allows the robot to behave proactively by taking leading actions in case the user does not follow the task plan as experienced in the training phase. On the other hand, ADHSMM also permits the robot to shape the task dynamics by modifying the states duration according to the interaction with the human, and therefore react to the user's actions. These two types of behaviors are driven by the forward variable $\boldsymbol{\alpha}_t$ in (2), which determines the influence of the ADHSMM states at each time step t considering the partial observation $\boldsymbol{\zeta}_{1:t}$, the transition matrix $a_{i,j}$, and the duration model $\mathcal{P}_i(\log(d) \,|\, \boldsymbol{u}_t)$ that takes into account the interaction with the user. The forward variable will next be used to generate trajectory distributions to control the robot during the collaborative task.

2.2 Trajectory retrieval using dynamic features

In the field of speech processing, it is common to exploit both static and dynamic features to reproduce smooth trajectories from HMMs [Furui, 1986; Tokuda *et al.*, 1995; Zen *et al.*, 2007a]. This is achieved by encoding the distributions of both static and dynamic features (the dynamic features are often called delta coefficients). In speech processing, these parameters usually correspond to the evolution of mel-frequency cepstral coefficients characterizing the power spectrum of a sound, but the same approach can be used with any form of continuous signals. In robotics, this approach has rarely been exploited, at the exception of the work from [Sugiura *et al.*, 2011] employing it to represent object manipulation movements. We take advantage of this formulation for retrieving a reference trajectory with associated covariance that will govern the robot motions according to the behavior determined by the ADHSMM.

For the encoding of robot movements, velocity and acceleration can alternatively be used as dynamic features. By considering an Euler approximation, they are computed as

$$
\dot{\boldsymbol{x}}_t = \frac{\boldsymbol{x}_{t+1} - \boldsymbol{x}_t}{\Delta t}, \ddot{\boldsymbol{x}}_t = \frac{\dot{\boldsymbol{x}}_{t+1} - \dot{\boldsymbol{x}}_t}{\Delta t} = \frac{\boldsymbol{x}_{t+2} - 2\boldsymbol{x}_{t+1} + \boldsymbol{x}_t}{\Delta t^2}, \tag{11}
$$

where $\boldsymbol{x}_t$ is a multivariate position vector.

By using (11), the observation vector $\boldsymbol{\zeta}_t$ will be used to represent the concatenated position, velocity and acceleration vectors at time step t, namely

$$
\boldsymbol{\zeta}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \dot{\boldsymbol{x}}_t \\ \ddot{\boldsymbol{x}}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ -\frac{1}{\Delta t}\boldsymbol{I} & \frac{1}{\Delta t}\boldsymbol{I} & \boldsymbol{0} \\ \frac{1}{\Delta t^2}\boldsymbol{I} & -\frac{2}{\Delta t^2}\boldsymbol{I} & \frac{1}{\Delta t^2}\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t+1} \\ \boldsymbol{x}_{t+2} \end{bmatrix}. \tag{12}
$$

ζ and x are then defined as large vectors concatenating ζ_t and x_t for all time steps, namely $\zeta = \begin{bmatrix} \zeta_1^\top & \zeta_2^\top & \cdots & \zeta_T^\top \end{bmatrix}^\top$, $x = \begin{bmatrix} x_1^\top & x_2^\top & \cdots & x_T^\top \end{bmatrix}^\top$.

Similarly to the matrix operator (12) defined for a single time step, a large sparse matrix Φ can be defined so that $\zeta = \Phi x$, namely[3]

$$
\overbrace{\begin{bmatrix} \vdots \\ x_t \\ \dot{x}_t \\ \ddot{x}_t \\ x_{t+1} \\ \dot{x}_{t+1} \\ \ddot{x}_{t+1} \\ \vdots \end{bmatrix}}^{\zeta} = \overbrace{\begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \cdot\cdot\cdot & \\ \cdots & I & 0 & 0 & \cdots & \\ \cdots & -\frac{1}{\Delta t}I & \frac{1}{\Delta t}I & 0 & \cdots & \\ \cdots & \frac{1}{\Delta t^2}I & -\frac{2}{\Delta t^2}I & \frac{1}{\Delta t^2}I & \cdots & \\ \cdots & I & 0 & 0 & \cdots & \\ \cdots & -\frac{1}{\Delta t}I & \frac{1}{\Delta t}I & 0 & \cdots & \\ \cdots & \frac{1}{\Delta t^2}I & -\frac{2}{\Delta t^2}I & \frac{1}{\Delta t^2}I & \cdots & \\ \cdot\cdot & \vdots & \vdots & \vdots & \ddots & \end{bmatrix}}^{\Phi} \overbrace{\begin{bmatrix} \vdots \\ x_t \\ x_{t+1} \\ x_{t+2} \\ x_{t+3} \\ \vdots \end{bmatrix}}^{x} .
$$
(13)

During the demonstration phase of a collaborative task, the collected dataset $\{\zeta_t\}_{t=1}^N$ with $N = \sum_m^M T_m$ is composed of M trajectory samples, where the m-th trajectory sample has T_m datapoints. This dataset is encoded by an ADHSMM, which can also provide a given sequence of states $s = \{s_1, s_2, \ldots, s_T\}$ of T time steps, with discrete states $s_t \in \{1, \ldots, K\}$. So, the likelihood of a movement ζ is

$$
\mathcal{P}(\zeta|s) = \prod_{t=1}^T \mathcal{N}(\zeta_t|\mu_{s_t}, \Sigma_{s_t}),
$$
(14)

where μ_{s_t} and Σ_{s_t} are the center and covariance of state s_t at time step t. This product can be rewritten as

$$
\mathcal{P}(\zeta|s) = \mathcal{N}(\zeta|\mu_s, \Sigma_s),
$$
(15)

with $\mu_s = \begin{bmatrix} \mu_{s_1} \\ \mu_{s_2} \\ \vdots \\ \mu_{s_T} \end{bmatrix}$ and $\Sigma_s = \begin{bmatrix} \Sigma_{s_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{s_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{s_T} \end{bmatrix}$.

By using the relation $\zeta = \Phi x$, we then seek during reproduction for a trajectory x maximizing (15), namely

$$
\hat{x} = \arg\max_x \; \log \mathcal{P}(\Phi x \mid s).
$$
(16)

The part of $\log \mathcal{P}(\Phi x \mid s)$ dependent on x takes the form $c = (\mu_s - \zeta)^\top \Sigma_s^{-1} (\mu_s - \zeta) = (\mu_s - \Phi x)^\top \Sigma_s^{-1} (\mu_s - \Phi x)$. A solution can be found by differentiating the above objective function with respect to x and equating to 0, providing the trajectory (in vector form)

$$
\hat{x} = \left(\Phi^\top \Sigma_s^{-1} \Phi \right)^{-1} \Phi^\top \Sigma_s^{-1} \mu_s,
$$
(17)

with the covariance error of the weighted least squares estimate given by

$$
\hat{\Sigma}^x = \sigma \left(\Phi^\top \Sigma_s^{-1} \Phi \right)^{-1},
$$
(18)

where σ is a scale factor.[4]

[3]Note that a similar operator is defined to handle border condi-

Figure 2: Experimental setting of the handover and transportation task. The teacher on the left demonstrates the skill to the robot, while the person on the right is the collaborator. After learning, the robot reproduces the collaborative behavior as demonstrated by the teacher. **Top:** First phase of the demonstration in which the robot reaches for an object (a screwdriver in this setup) that is delivered by the user. The hand position is tracked using optical markers. **Bottom:** Second phase of the demonstration showing the transportation of the object towards its final location (black area).

2.3 Optimal Controller for HRC

The resulting Gaussian $\mathcal{N}(\hat{x}, \hat{\Sigma}^x)$ forms a trajectory distribution that will be used to control the robot motion during the collaborative task. Specifically, once a reference trajectory $\hat{x}$ has been obtained, the optimal controller for human-robot collaborative tasks proposed in [Rozo *et al.*, 2015] is used to track this reference. Such an optimal feedback controller allows the robot to plan a feedback control law tracking the desired state within a minimal intervention control principle. Formally, the problem is stated as finding the optimal input ν that minimizes the cost function

$$
J_t = \sum_{n=t}^\infty (x_n - \hat{x}_t)^\top Q_t (x_n - \hat{x}_t) + \nu_n^\top R_t \nu_n,
$$
(19)

where the matrices Q_t and R_t are weighting matrices that determine the proportion in which the tracking errors and control inputs affect the minimization problem. Here, we take advantage of the variability observed during the demonstrations to adapt the error costs in (19) in an online manner by defining $Q_t = \left(\hat{\Sigma}_t^x \right)^{-1}$, and by setting R_t in accordance to the application and motors used in the experiment (set as constant diagonal matrix in the experiments reported in this paper).

3 Handover and Transportation Task

In order to show how the proposed approach can be exploited in HRC scenarios, we consider a collaborative task in which

tions, and that Φ can automatically be constructed through the use of Kronecker products.

[4]Equations (17) and (18) describe a trajectory distribution, and can be computed efficiently with Cholesky and/or QR decompositions by exploiting the positive definite symmetric band structure of the matrices.
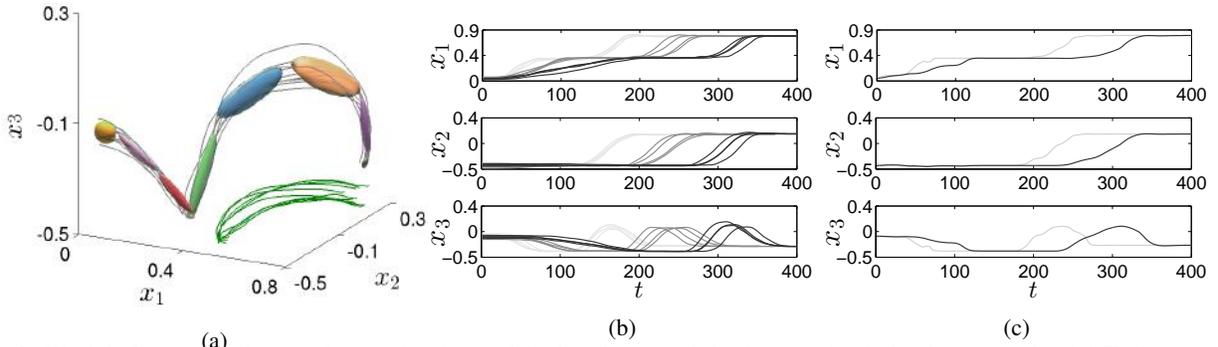
Figure 3: Model, demonstrations and reproductions of the handover task in the reactive behavior scenario. **(a)** Task-space view of the model, robot end-effector and human hand trajectories depicted in gray and green solid lines, respectively. **(b)** Evolution in time of the demonstrations. **(c)** Reproductions of the skill. Darker lines display slower human hand motion.
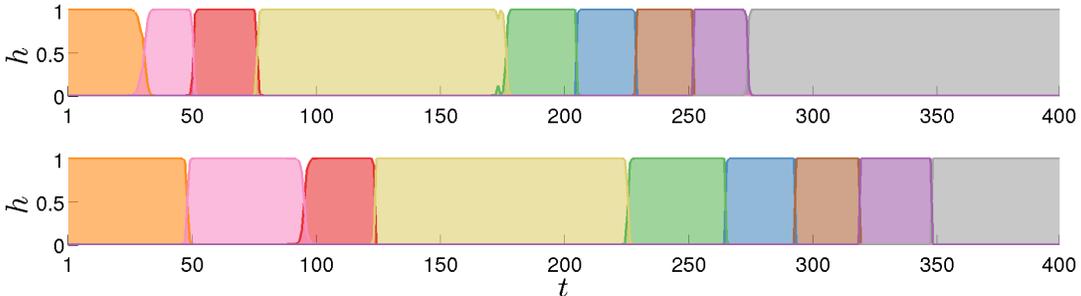
Figure 4: Evolution in time of the rescaled forward variable $h_{t,i}$ during the handover task. The temporal dynamics of the task is autonomously adapted according to the interaction with the user.

the robot role is to first reach for an object that is delivered by the user, and then transport it along a given path to attain a final location. The first part of the task should thus be conditioned by the human motion, namely, the state durations for this phase of the task should vary according to the user hand position measured when he/she is bringing the object to the location where the robot will grab it. The second part of the task occurs when the robot takes the object and transports it towards the final location. Here, the robot motion is expected to be independent from the human motion.

A Barrett WAM robot is used in this experiment. In the demonstration phase, the gravity-compensated robot is kinesthetically guided by the teacher while cooperatively achieving the task with a person, as shown in Figure 2. A human teacher first shows the robot how to approach the object location based on the user motion, and how to transport the object to the final location. The collaborator's hand position is tracked with a marker-based NaturalPoint OptiTrack motion capture system, composed of 12 cameras working at a rate of 30 fps. The position of the robot is defined by Cartesian position x, while the external input u, conditioning state duration, corresponds to the human hand position x^H.

During the demonstration phase, the first part of the task was demonstrated by showing three different human motion velocities labeled as low, medium and fast. We collected four demonstrations for each velocity level, totaling twelve demonstrations, and afterwards trained a model of nine components ($K = 9$, selected empirically), under the assumption of a left-right topology. Each datapoint consists

of the robot position x_t and velocity $\dot{x}_t$ at each time step t of the demonstration, therefore the observation vector is defined as $\zeta_t = [x_t^\top, \dot{x}_t^\top]^\top$ in this experiment. We model the state duration using a GMM with $K^D = 2$ (selected empirically), trained by using the dataset $\{\xi_{i,m}\}_{m=1}^{M_i}$ with $\xi_{i,m} = [u_m^\top, \log(d_{i,m})]^\top$, where u_m corresponds to the hand position x^H recorded while the system is in state i, $\log(d_{i,m})$ is the log-transformed duration given by the number of consecutive time steps that the system stays in state i, and M_i is the number of datapoints in the demonstration sequences in which state i was visited.

3.1 Results

Reactive behaviors

Figure 3 shows the model, demonstrations and reproductions of the collaborative task when the robot is acting in a reactive manner to the human input. In Figure 3a, we show a 3D view of the model in the workspace of the robot, as well as the human input during the demonstrations. The model successfully encodes the local correlations between the task space variables. Figure 3b depicts the demonstrations provided to the robot through kinesthetic teaching, with lighter lines corresponding to a faster approach towards the human hand before the handover occurs. In Figure 3c we show the skill reproductions for two different hand velocities. We can see that the movement is correctly regenerated in both situations. Finally, Figure 4 shows the forward variable in the two cases that we considered during the reproductions. Firstly, we see that the sequence of states is correctly generated in both
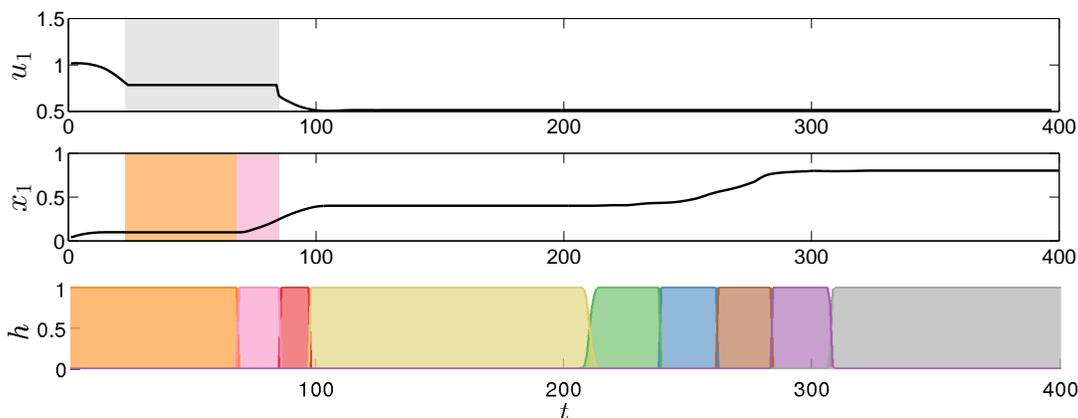
Figure 5: Evolution in time of the proactive behavior variables. **Top:** First dimension of the user hand position. The hand does not move during the period marked by the shaded area. **Middle:** Dimension x_1 of the robot end-effector position in task space. The state transition (shaded areas) marks the start of the proactive behavior. **Bottom:** Rescaled forward variable $h_{t,i}$.

scenarios, matching what one would expect to be an accurate task space trajectory of the end-effector for the considered skill. Note that this was achieved by taking advantage of the probabilistic modeling of temporal variability employed by ADHSMM, through state transition and state duration probabilities. Secondly, we observe that the duration of the first three states is strongly correlated to the human hand motion since the duration shortens when the hand moves faster (first row), resulting in a faster approach of the end-effector to the human hand. The influence of the hand movement in the remaining states is negligible, as expected.

Proactive behaviors

In addition to human-adaptive reactive behaviors, the proposed approach can also be used to generate proactive behaviors that remain consistent with the expected temporal evolution of the task. To showcase this property, we portray a scenario where the human stops moving while reaching the object (Figure 5, top). This illustrates a situation in which a new person would be interacting with the robot and would not know enough about the task to lead the cooperation. Consequently, the behavior that one would like to observe in the robot would be that it provides clues about how previous users proceeded in similar situations. In the proposed approach, the robot will take the initiative to proceed with the movement after some time (in case this duration lasts unexpectedly longer than in past experiences), and will guide the user toward the next step of the task (Figure 5, middle). This occurs at most when the duration of state 1 exceeds its maximum value $d_1^{\max}$ and the model transits to state 2 (Figure 5, bottom). This mechanism can be exploited to let the robot help new users proceed with their roles in the collaboration by showing its intent in the cooperation, i.e., showing the way in which the task is believed to be continued (see Figure 5, before $t = 100$).[5] A video showing the results of this experiment is available at http://programming-by-demonstration.org/IJCAI-IML2016/

This mechanism currently has some limitations. In the example, the robot had to stop only once before the user could understand what to do. If the user had not understood that

his/her role was to hand the object to the robot, the cooperation would have failed because the robot would have finished the task without having the object in its hand. A possible way to increase the number of clues that the robot provides to the human before continuing the task on its own could be to add more states to the model to let the user better understand the intent of the movement. This would make the robot reach the handover pose in approximately the same time, but with a greater number of state transitions, i.e., discrete movements towards the target, providing the user with more information about the movement. Similarly, additional sensory information could be used to verify that a valid situation occurs before moving on to the next part of the task.

4 Conclusions and Future Work

This paper introduced an approach allowing collaborative robots to learn reactive and proactive behaviors from human demonstrations of a collaborative task. We showed that reactive behaviors could include the modulation of the temporal evolution of the task according to the interaction with the user, while proactive behaviors could be achieved by exploiting the temporal patterns observed during the learning phase. These collaborative behaviors can be exploited to extend the robot capability to assisting tasks in which both interaction and temporal aspects are relevant. Indeed, the probabilistic nature of the proposed ADHSMM allows the robot to react to different human dynamics, which is beneficial for collaborating with distinct partners. The proposed proactive behavior allows the robot to take the lead of a task when it is appropriate (namely, according to the task dynamics previously experienced in the demonstrations), which can be exploited to communicate its intention to the user. The approach is also compatible with task-parameterized movement encoding, in which states are modulated by external parameters, reducing the need for learning new trajectories [Silvério *et al.*, 2015]. We plan to extend the proposed learning model to situations in which the transitions between the model states also depend on the interaction with the partner, which will allow the robot to learn more complex collaborative tasks.

---

[5] In this experiment, a time step approximately lasts 0.04 seconds.

References

[Amor *et al.*, 2013] H. B. Amor, D. Vogt, M. Ewerton, E. Berger, B. Jung, and J. Peters. Learning responsive robot behavior by imitation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 3257–3264, 2013.

[Billard *et al.*, 2008] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Robot programming by demonstration. In B. Siciliano and O. Khatib, editors, *Springer Handbook of Robotics*, pages 1371–1394. Springer, 2008.

[Evrard *et al.*, 2009] 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 (Humanoids)*, pages 399–404, 2009.

[Furui, 1986] S. Furui. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(1):52–59, 1986.

[Kulvicius *et al.*, 2013] T. Kulvicius, M. Biehlc, M. Javad Aein, M. Tamosiunaite, and F. Wörgötter. Interaction learning for dynamic movement primitives used in cooperative robotic tasks. *Robotics and Autonomous Systems*, 61(12):1450–1459, 2013.

[Leggetter and Woodland, 1995] C.J. Leggetter and P.C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*, 9(2):171–185, 1995.

[Li *et al.*, 2015] Y. Li, K. P. Tee, W. L. Chan, R. Yan, Y. Chua, and D. K. Limbu. Continuous role adaptation for human-robot shared control. *IEEE Transactions on Robotics*, 31(3):672–681, 2015.

[Maeda *et al.*, 2015] G. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, and J. Peters. A probabilistic framework for semi-autonomous robots based on interaction primitives with phase estimation. In *Intl. Symp. on Robotics Research*, pages 1–16, 2015.

[Medina *et al.*, 2011] J. Medina, M. Lawitzky, A. Mortl, D. Lee, and S. Hirche. An experience-driven robotic assistant acquiring human knowledge to improve haptic cooperation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 2416–2422, 2011.

[Mitchell *et al.*, 1995] C. Mitchell, M. Harper, and L. Jamieson. On the complexity of explicit duration HMM's. *IEEE Trans. on Speech and Audio Processing*, 3(3):213–217, 1995.

[Nose *et al.*, 2007] T. Nose, Y. Kato, and T. Kobayashi. Style estimation of speech based on multiple regression hidden semi-markov model. In *INTERSPEECH*, pages 2285–2288, 2007.

[Paraschos *et al.*, 2013] A. Paraschos, C. Daniel, J. Peters, and G. Neumann. Probabilistic movement primitives. In *Neural Information Processing Systems (NIPS)*, pages 2616–2624, 2013.

[Rabiner, 1989] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.

[Rozo *et al.*, 2015] 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 *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 1024–1030, 2015.

[Rozo *et al.*, 2016] L. Rozo, S. Calinon, D. G. Caldwell, P. Jiménez, and C. Torras. Learning physical collaborative robot behaviors from human demonstrations. *IEEE Transactions on Robotics*, pages 1–15, 2016.

[Silvério *et al.*, 2015] J. Silvério, L. Rozo, S. Calinon, and D. G. Caldwell. Learning bimanual end-effector poses from demonstrations using task-parameterized dynamical systems. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 464–470, 2015.

[Sugiura *et al.*, 2011] K. Sugiura, N. Iwahashi, H. Kashioka, and S. Nakamura. Learning, generation, and recognition of motions by reference-point-dependent probabilistic models. *Advanced Robotics*, 25(6–7):825–848, 2011.

[Tokuda *et al.*, 1995] K. Tokuda, T. Masuko, T. Yamada, T. Kobayashi, and S. Imai. An algorithm for speech parameter generation from continuous mixture HMMs with dynamic features. In *Proc. European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 757–760, 1995.

[Wilcox *et al.*, 2012] R. Wilcox, S. Nikolaidis, and J. Shah. Optimization of temporal dynamics for adaptive human–robot interaction in assembly manufacturing. In *Robotics: Science and Systems (R:SS)*, pages 1–8, 2012.

[Yamagishi and Kobayashi, 2005] J. Yamagishi and T. Kobayashi. Adaptive training for hidden semi-Markov model. In *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 365–368, 2005.

[Yamazaki *et al.*, 2005] T. Yamazaki, Naotake, J. Yamagishi, and T. Kobayashi. Human walking motion synthesis based on multiple regression hidden semi-Markov model. In *IEEE Intl. Conf. on Cyberworlds (CW)*, pages 452–459, 2005.

[Yu and Kobayashi, 2006] S. Yu and T. Kobayashi. Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model. *IEEE Trans. on Signal Processing*, 54(5):1947–1951, 2006.

[Yu, 2010] S. Yu. Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215–243, 2010.

[Zen *et al.*, 2007a] H. Zen, K. Tokuda., and T. Kitamura. Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences. *Computer Speech and Language*, 21(1):153–173, 2007.

[Zen *et al.*, 2007b] H. Zen, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura. A hidden semi-Markov model-based speech synthesis system. *IEICE Trans. on Information and Systems*, E90-D(5):825–834, 2007.