A Probabilistic Model for Cobot Decision Making to Mitigate Human Fatigue in Repetitive Co-manipulation Tasks

Aya Yaacoub, Vincent Thomas, Francis Colas, Pauline Maurice

A Probabilistic Model for Cobot Decision Making to Mitigate Human Fatigue in Repetitive Co-manipulation Tasks

Aya Yaacoub[1], Vincent Thomas[1], Francis Colas[1], Pauline Maurice[1]

***Abstract*—Work related musculoskeletal disorders (WMSDs) are very common. Repetitive motion, which is often present in industrial work, is one of the main physical causes of WMSDs. It uses the same set of human joints repeatedly, which leads to localized joint fatigue. In this work, we present a framework to plan a policy of a collaborative robot that reduces the human fatigue in the long term, in highly repetitive co-manipulation tasks, while taking into account the uncertainty in the human postural reaction to the robot motion and the partial observability of the human fatigue state. We model the problem using continuous-state Partially Observable Markov Decision Process (POMDP), and use a physics-based digital human simulator to predict the fatigue cost of the possible robot actions. We then use an online planning algorithm to compute the optimal robot policy. We demonstrate our approach on a simulated experiment in which a robot repeatedly carries an object for the human to work on, and the object Cartesian pose needs to be optimized. We compare the policy generated with our approach with a random, a cyclic and a greedy (short-term optimization) policy, for different user profiles. We show that our approach outperforms the other policies on all tested scenarios.**

***Index Terms*—Human Factors and Human-in-the-Loop, Human-Robot Collaboration, Planning under Uncertainty**

I. INTRODUCTION

MUSCULOSKELETAL Disorders (MSDs) are injuries affecting the different substructures of the musculoskeletal system. Work-related MSDs (WMSDs) are particularly common. Almost 3 out of 5 workers in the EU-28 region report a WMSD [1]. This urges finding solutions to reduce the risk of developing MSDs at work. Recently, industry has been heading towards an increased integration of collaborative robots (cobots) [2]. When in direct interaction with workers (generally through a co-manipulated object), cobots can influence how workers perform a task, either through force or motion modification. Consequently, multiple studies have investigated exploiting cobots to reduce the risk of developing WMSDs [3]–[5]. They generally focused on alleviating one or more of three common biomechanical risk factors leading to WMSDs: forceful work, non-ergonomic postures, and repetitive motion [6].

Manuscript received: March, 15, 2023; revised: July, 23, 2023; accepted: August, 24, 2023.

This paper was recommended for publication by Editor Aniket Bera upon evaluation of the Associate Editor and Reviewers comments.

This work was partly supported by the French ANR under Grant No. ANR-20-CE33-0004 (project ROOIBOS).

[1] The authors are with Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France. firstname.name@loria.fr

Digital Object Identifier (DOI): see top of this page.

Specifically, several studies focused on guiding the human to operate in an ergonomically-optimal posture via the positionning of the object co-manipulated by the robot [3], [4]. However, numerous industrial tasks are highly repetitive, hence focusing on a single posture, even the most ergonomic one, could induce repeated motions which constantly load the same human joints. Eventually, joints would be subject to localized fatigue, which, in the long term, could lead to WMSD [6]. Actually, some ergonomics studies suggest that motor variability in the task execution –i.e., varying how the motion is performed– might be beneficial to reduce WMSD risk [7]. To limit the effect of repetitive motions, some studies proposed reactively adjusting the robot motion based on the human physical fatigue level to induce changes in the human posture over time [5].

While definitely a step forward, such approaches assume that setting the co-manipulated object pose is sufficient to guide the human to the intended whole-body posture. But due to the kinematic redundancy of the human body, the object pose only partly constrains the human posture. A same pose might trigger different whole-body motions from the human (referred to as *postural reactions* in this work), depending on inter- (e.g. handedness, expertise) or intra-individual (e.g. fatigue) factors. These factors often can only be inferred, which makes the human postural reaction uncertain, i.e. non-deterministic. In addition, the human fatigue level is a physiological state that can hardly be directly measured. Instead, it can be estimated by combining observation of the human motion with a fatigue model, but this process makes fatigue a non-observable variable because of the inevitable uncertainties brought by the modeling. Finally, state-of-the-art approaches on WMSDs risk mitigation in repetitive co-manipulation tasks usually are purely reactive: they adjust the object pose to improve the immediate fatigue, regardless how this affects the long term evolution of fatigue. This does not guarantee that the resulting behavior is optimal in the long term, i.e. after many repetitions of the task.

To address the aforementioned challenges (stochasticity of the human postural reaction, partial observability of the human fatigue, long-term planning), this work proposes an approach based on probabilistic planning. Specifically, we present a framework to plan a policy of a collaborative robot that reduces the human physical fatigue in the long term, in highly repetitive co-manipulation tasks (i.e. tasks comprised of numerous short-duration –from a few seconds to a few minutes– cycles). We model the problem as a long-

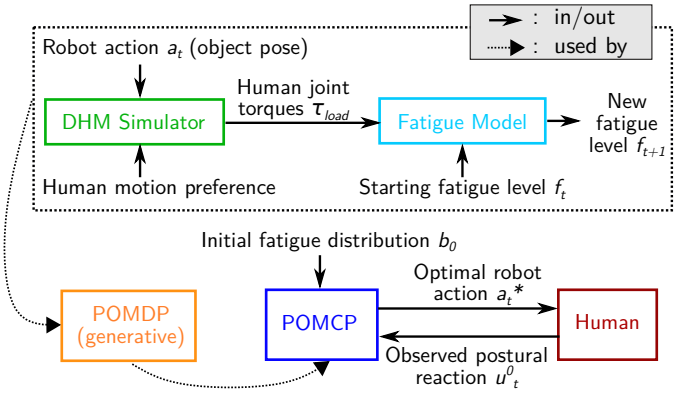


Fig. 1. Description of the framework. The problem is modeled as a POMDP (Section III-A). A DHM simulator (Section III-C) and a fatigue model (Section III-B) form a fatigue estimation block used by the POMDP generative model to evaluate the fatigue cost induced by possible robot actions. Based on this evaluation, the POMCP algorithm (Section III-D) selects the action to execute at each cycle.

term decision making problem under uncertainty and partial observability using the Partially Observable Markov Decision Process (POMDP) framework [8]. We use a physics-based digital human simulator to generate the various human postural reactions (whole-body motion) to all the possible robot actions (object poses). We then use a fatigue model from the literature to evaluate the fatigue of each human joint induced by each simulated motion. The estimated fatigue finally serves to solve the POMDP, i.e. compute the optimal robot policy (which actions the robot should execute), with the Partially Observable Monte Carlo Planning (POMCP) algorithm [9]. Fig. 1 summarizes the different elements of the proposed approach. We test our approach in simulation, and benchmark the policy computed with POMCP against random, alternating and greedy policies.

II. Related Work

Many studies addressed reducing WMSD risk in human-robot collaboration through robot-induced human posture modification [3], [10]. In [3] the co-manipulated object pose corresponding to the most ergonomic human posture is calculated, and the robot presents the object to the human at this optimal pose. Others proposed reactively leading the worker back to the most ergonomic posture when a deviation occurs [11], [12]. In [11], possible deviations from the optimal human posture and their corrective robot actions were identified. Then, in a welding task where the cobot carried the object, the human posture was monitored online, and a robot action executed when a deviation occurred. Other studies planned a start-to-end optimal hand trajectory for each task [13], [14]. In [14], a multi-objective optimization framework was proposed to identify a hand trajectory that simultaneously optimizes a group of ergonomics measures, for a specific user in a given task. The resulting hand trajectory could then be used by a robot in co-manipulation tasks. However, all these studies consider a single time-invariant optimal posture or motion. But focusing on a small set of postures, despite them being ergonomically optimal, risks generating repetitive motion.

Conversely, [5] considered the evolution of the human capacity during a repetitive task. Specifically, they changed the object pose according to the human muscular fatigue level whenever fatigue crossed a threshold, to allow fatigued muscles to recover over the following cycles. Yet, the approach was purely reactive and did not account for the uncertainty of the human reaction to the robot motion.

In a non-WMSDs-related context, human-robot interaction studies proposed to consider the stochastic nature of the human behavior as well as the lack of full observability of the human state using Markov Decision Process [15] or Bayesian inference [16]. In [15], a variant of POMDP (with mixed observability) was used to infer a hidden human adaptability variable that conditioned the human reactions in a task where a human and a robot moved a table together and plan the robot strategy (leading or following the human) accordingly.

Different aspects of the WMSD risk mitigation problem in repetitive co-manipulation have thus been addressed. But to the best of our knowledge, no framework has been proposed to consider all aspects together: long-term planning in repetitive tasks, non-deterministic nature of human movements, and partial observability of the human physical state.

III. Methods

The objective of our framework is to allow planning a policy of a collaborative robot that minimizes the long term human fatigue, taking into account the uncertainty in the human postural reaction, and the partial observability of the human state (fatigue). In the remaining of this work, we illustrate our approach on a task where, at each cycle, the robot brings an object to the human at a certain Cartesian pose. The human then reaches to the object to work on it while the robot continues to hold it at the same pose (e.g. spray painting car pieces, scanning parcels packed by the robot). When done, the robot moves the object away and brings a new one, which corresponds to starting a new cycle [1].

Fig. 2 depicts one cycle of our decision-making problem. Every cycle, the cobot decides the pose where to bring the object, according to its current belief about the human state (fatigue) and the computed policy. Given this pose, the human chooses a motor strategy to reach for the object (referred to as *motion preference* in this work) according to various factors influencing its preference (e.g., fatigue, expertise). The human then executes the corresponding whole-body motion (*postural reaction*) to work on the object, which modifies its fatigue state in a way that is specific to the postural reaction[2]. On the robot side, observing the human postural reaction gives information that is used to update the belief about the human fatigue state.

[1]The proposed approach could readily apply to other kinds of co-manipulation tasks. For instance, for tasks where objects are repetitively co-transported by the human and the robot simultaneously, the robot action used in the POMDP model would be the robot end-effector trajectory during the transportation, instead of the end-effector pose of the fixed object in the current example. Since the DHM simulator allows to simulate whole-body motions, and not only postures, such displacement tasks can also be handled.

[2]Human motion preference is an input of the problem: we cannot choose the motion preference. Rather when selecting the robot action, we need to account for the fact that humans can react in different ways to an robot action, and that these reactions have different consequences on the fatigue.

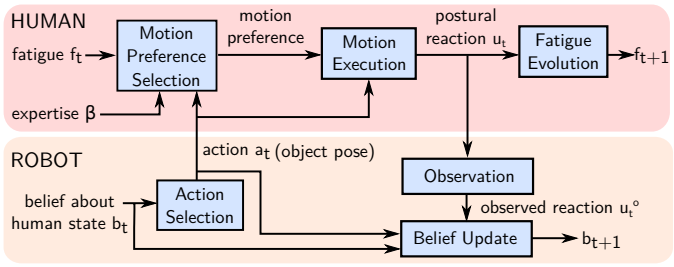


Fig. 2. Workflow of one cycle of the human-robot interaction process. The robot action is selected according to a policy π.

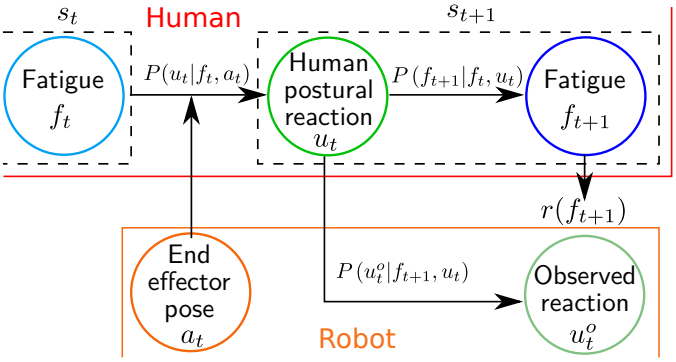


Fig. 3. Definition of a transition from state s_t to state s_{t+1} in the POMDP. For implementation purposes the state was augmented with the human reaction, i.e. $s_t = (f_t, u_{t-1})$.

This section presents the elements used to model and solve the decision-making problem, i.e. generate the action-selection policy that minimizes the accumulated fatigue over repeated cycles (Fig. 1): the POMDP (Section III-A) which models the consequences of a cobot action w.r.t. the objective of minimizing fatigue, the fatigue model (Section III-B) and digital human simulator (Section III-C) which serve to estimate the evolution of fatigue associated with a human postural reaction, and the online algorithm (Section III-D) which generates the action selection policy.

A. Fatigue POMDP

POMDP is a framework used to model the stochastic evolution of a partially observable system state given a system's agent action [8]. The POMDP describing our problem is defined by the tuple $(S, A, \Omega, T, O, R, b_0)$, where:

- S is the state space of the system. At cycle t, the state $s_t = (f_t, u_{t-1})$ contains the current fatigue state f_t and the previous human postural reaction u_{t-1}. In this work, we model fatigue at joint level, and track the fatigue of each human internal degree of freedom (DoF), including a distinction between agonist and antagonist DoF (see Section III-B).
- A is the action space, i.e. the set of all actions $a \in A$ (Cartesian poses) where the robot can choose to bring the object.
- Ω is the observation space, and $o_t \in \Omega$ is an observation of the state s_t. Specifically, $o_t = u^o_{t-1}$ is an observation of the human postural reaction u_{t-1} (whole-body motion) in response to the robot action.
- $T(s, a, s') = P(s'|s, a)$ is the transition probability from state $s \in S$ to $s' \in S$ when action $a \in A$ is selected. The transition function encodes the human fatigue evolution. The transition is done in two steps (Fig. 3). First, given the current system state s_t and the action a_t selected by the robot, the human chooses a motion preference and executes the corresponding postural reaction u_t with probability $P(u_t|s_t, a_t)$ (Section **??**). Second, due to the motion execution, the human fatigue evolves (Section III-B) leading to the new state s_{t+1} with a probability $P(f_{t+1}|s_t, u_t)$. The transition probability is the product of the probabilities of the two steps $T(s_t, a_t, s_{t+1}) = P(s_{t+1}|s_t, a_t) = P(f_{t+1}|s_t, u_t)P(u_t|s_t, a_t)$.
- $O(o, s) = P(o|s)$ is the probability of observing $o \in \Omega$ at $s \in S$, i.e. of identifying, from available motion measurements, a human postural reaction u^o_t at the new state s_{t+1}. Since the postural reaction is part of the system state, it is defined by $P(u^o_t|s_{t+1}) = P(u^o_t|f_{t+1}, u_t) = P(u^o_t|u_t)$. This distribution represents the uncertain identification by the robot of the human postural reaction u within the possible set U (e.g., due to noise in the sensors used to track the human motion).
- R is the reward function such that $r(s, a, s')$ is the reward obtained for the transition from $s \in S$ to $s' \in S$ by $a \in A$. In this work, R is defined using a fatigue cost function which depends on the fatigue value at state s'.
- $b_0 = P(s_0)$ is the initial belief state, i.e. a probability distribution over the set of possible initial fatigue states of the system.

Let $b \in B$ be the belief state, i.e., the probability distribution over the true state of the system, given the past actions $a_{1:t}$ and past observations $o_{1:t}$, such that $b(s) = P(s_t|a_{1:t}, o_{1:t})$. Solving the POMDP consists in computing an optimal policy $\pi : B \to A$ associating to each belief state b the action a to perform (i.e. π is a mapping from action-observation histories to actions) in order to maximize the average discounted accumulated reward $E[\sum \gamma^t r_t] = \sum_{t=0}^{\inf} \gamma^t r_t$ during execution. $\gamma \in [0, 1]$ is the discount factor of the POMDP which weighs the importance of future vs. immediate reward when selecting an action. In this work, the optimal policy π allows to select the robot action (object Cartesian pose) at each cycle, given the current belief about the human fatigue, in order to minimize the predicted accumulated fatigue over the future cycles.

B. Fatigue Model

The transition function of the POMDP encodes the evolution of the human fatigue induced by a whole-body motion executed in response to a robot action. Therefore a fatigue model is needed to evaluate the transition function. Several fatigue models have been proposed in the literature, ranging from biochemical models of the mechanisms underlying muscular activity (e.g., Ca^{2+} cross-bridge mechanism) and representation of the different motor-unit types, to macroscopic models evaluating endurance time at joint level [17]–[19]. Usually, the more detailed and closer to biophysics the models are, the more parameters they involve, which can be hard to

tune. Therefore, in this work we use a macroscopic fatigue model from the literature introduced by Ma et al. [20], which requires only a limited set of parameters. While these authors proposed both a muscular [21] and a joint [20] version of their model, we opt for the joint fatigue model for computational efficiency. Indeed, the joint model requires the values of the joint torques associated to the motion (see Eq. 2), whereas the muscular model requires the values of the muscle forces, which necessitates a more complex and computationally-costly human model.

Thus, we represent fatigue at joint level, and define it as a temporary loss in torque production capacity of a joint [22]:

$$f^j(t) = 1 - \frac{\tau^j_{cem}(t)}{\tau^j_{max}} \tag{1}$$

where f^j is the instantaneous fatigue of joint j, τ^j_{cem} is the current maximum exertable torque of joint j, and τ^j_{max} is the nominal maximum exertable torque (in the absence of fatigue). The evolution of τ^j_{cem} associated with a human movement is given by the model [20], which distinguishes two modes, fatigue (τ^j_{cem} decreases) and recovery (τ^j_{cem} increases):

$$\frac{\mathrm{d}\tau^j_{cem}}{\mathrm{d}t} = \begin{cases} -k\frac{\tau^j_{cem}}{\tau^j_{max}}\tau^j_{load} & \text{if } \tau^j_{load} > \alpha^j\tau^j_{max} \\ & \text{(fatigue)} \\ R(\tau^j_{max} - \tau^j_{cem}) & \text{if } \tau^j_{load} < \alpha^j\tau^j_{max} \\ & \text{(recovery)} \end{cases} \tag{2}$$

where k and R are respectively the fatigue and recovery coefficients of the joint, τ^j_{load} is the exerted human joint torque (i.e. torque generated by the human posture or motion) at joint j, and α^j determines the threshold to switch between fatigue and recovery modes.

Despite not using a muscle model, we still account for the fact that joints are actuated by both agonist and antagonist muscles which act in opposite directions. We therefore consider the positive and negative joint torques of each human internal DoF separately, and assign to each joint an agonist and an antagonist fatigue value, such that the fatigue state $f_t \in \mathbb{R}^{2N}$, where N is the number of human internal DoFs.

Given the fatigue model in Eq. 2, the fatigue at the end of a cycle f_{t+1} only depends on the previous fatigue f_t and on the human postural reaction u_t via the time-series of τ_{load} throughout the motion. Thus, in the POMDP transition function, $P(f_{t+1}|s_t, u_t) = P(f_{t+1}|f_t, u_t)$ is deterministic: $P(f_{t+1}|f_t, u_t) = 0$ unless f_{t+1} matches the value computed by the fatigue model.

C. Digital Human Simulator

In order to compute the fatigue evolution with the model presented in the Section III-B, the time-series of human joint torques $\tau_{load}(t)$ throughout the motion are needed. We use a Digital Human Model (DHM) simulator [23] to compute the joint torques associated with each possible human postural reaction u_t, i.e. each pair of possible robot action and human motion preference, and thereby predict the effect on fatigue of the different robot actions. The physics-based simulator allows to generate dynamically-consistent whole-body motions and associated joint torques from high-level descriptions of the task to perform and the motion preference.

The DHM used in this work is a tree-like chain of 17 rigid bodies linked together by 16 compound joints, for a total of $N = 39$ internal DoFs (7 per arm, 7 per leg, 11 for the back and head), plus 6 non-actuated DoFs for the free-floating base. Each DoF is a revolute joint controlled by a single actuator. The model is scalable in height and mass to adapt to any specific user's morphology.

The DHM motion is computed by solving a Linear Quadratic Programming (LQP) optimization problem to find the actuation variables (joint torques) which enable to follow some objectives at best (e.g., hand trajectory), while respecting dynamic and biomechanical constraints [24]. The LQP problem is formulated as follows [23]:

$$\begin{aligned} &\arg\min_{\boldsymbol{\chi}\in\mathbb{X}} \sum_i \omega_i \Delta_i(\boldsymbol{\chi}) \\ &s.t. \begin{cases} M(\mathbf{q})\dot{\boldsymbol{\nu}} + \mathbf{C}(\mathbf{q},\boldsymbol{\nu}) + \mathbf{g}(\mathbf{q}) = S_a\boldsymbol{\tau} - \sum_k J^T_{c_k}(\mathbf{q})\mathbf{w}_{\mathbf{c_k}} \\ G\boldsymbol{\chi} \leq \mathbf{h} \end{cases} \end{aligned} \tag{3}$$

where $\boldsymbol{\chi} = (\boldsymbol{\tau}^T, \dot{\boldsymbol{\nu}}^T, {\mathbf{w}_{\mathbf{c_k}}}^T)^T$ with $\boldsymbol{\tau}$ the joints torques, $\mathbf{q}$ the generalized coordinates of the system (joint angles), $\boldsymbol{\nu}$ the generalized velocity combining the free-floating base twist and the joint velocities $\dot{\mathbf{q}}$, and $\mathbf{w}_{\mathbf{c_k}}$ the contact wrench of the *k-th* contact point. The equality constraint is the equation of motion, with M the inertia matrix of the system, $\mathbf{C}$ the vector of Coriolis and centrifugal forces, $\mathbf{g}$ the gravity forces, S_a the actuation selection matrix due to the free-floating base, and J_{c_k} the Jacobian of contact point c_k. G and $\mathbf{h}$ are used to define the biomechanical limits (joint positions, velocities and torques), and the contact existence condition for each contact point according to the Coulomb friction model (feet/ground contact). The objective function is a weighted sum of LQP tasks $\Delta_i(\boldsymbol{\chi})$ formulated as the squared error between a desired (Cartesian or joint) acceleration and the system current acceleration[3]. For each LQP task, the desired body/joint acceleration is computed with a proportional-derivative formulation from the goal position, velocity and acceleration.

In this work, we use the following LQP tasks:

- A center of mass (CoM) Cartesian position task to control the horizontal position of the CoM and thereby maintain balance.
- A hand Cartesian pose task to simulate the human reaching to the object brought by the robot and working on it. For the reaching phase, a minimum-jerk trajectory between the hand initial pose and the object pose is used as reference trajectory.
- A low-level whole-body posture (i.e. joint position) task with a natural reference posture (standing, arms along the

[3]LQP acceleration tasks are used for position or trajectory tracking of body segments or joints. In this work, we only use LQP acceleration tasks, but since the external contact wrenches are part of the optimization variables, the presented LQP formulation allows to handle force tasks as well. Thus, while in the present example there is no force exchange between the human and the robot, the DHM simulator and the proposed approach can readily apply to co-manipulation activities involving significant human-robot interaction force.

body) to solve the kinematic redundancy in a way that generates reasonably human-like motions.

Additional LQP tasks are defined to simulate different human motion preferences, i.e. different motor strategies (by activating/deactivating these tasks, or by changing their weights w_i). The LQP formulation allows to solve the human kinematic redundancy using high-level description of the human motion preference (e.g., which joints to preferably move w.r.t. others, or which body segment orientation to preferably maintain) without having to specify the motion of each joint individually.

D. Solving the POMDP

Given the continuous nature of the fatigue variable and the dimension of the fatigue state vector $f_t \in \mathbb{R}^{2N}$ with $N = 39$, the state space S is continuous and high dimensional. We therefore use the Partially Observable Monte-Carlo Planning (POMCP) algorithm which allows solving a POMDP with a continuous and high dimensional state space [9]. With such a state space, it is complex to solve the problem at hand by using an explicit POMDP model. To overcome this limitation, POMCP relies on a generative model of a POMDP. The generative representation is described in Fig. 4: when an action a is performed from state s, the generative model samples a future state, a received observation, and a reward according to the probability distributions defining the transition and observation functions, and the reward function.

Using the generative model, POMCP follows a Monte-Carlo Tree Search approach (sampling-based tree search) to evaluate the actions of the POMDP at some belief state b. Starting from b, POMCP grows the tree by performing multiple descents down the tree. Each descent branches the tree following a sequence of sampled actions (using Upper Confidence Bound UCB1) and observations (using the POMDP generative model) and stops when a leaf is reached. Along each descent, every call to the generative model returns a reward that is stored for later action evaluation. The optimal action to execute at belief state b is finally selected based on the discounted accumulated rewards obtained across all the descents branching from this action.

The computation of an optimal policy with POMCP relies on two parameters: the number of descents per belief tree, and the exploration/exploitation parameter c of UCB1. A larger number of descents improves the evaluation of the reward associated with each action. c balances the width/depth ratio of the tree: for a given number of iterations, the tree is explored deeper when c decreases. For a properly chosen c, the POMCP policy converges to the optimal POMDP policy as the number of descents from b tends to infinity (thus a large number of descents is required for convergence). But given the limited time budget when performing online planning, the number of descents is constrained. We therefore use a variant (here referred to as POMCP-Max) introduced initially as MaxUCT for a fully-observable problem in [25], which aims at easing the time-consuming high number-of-descents convergence condition.

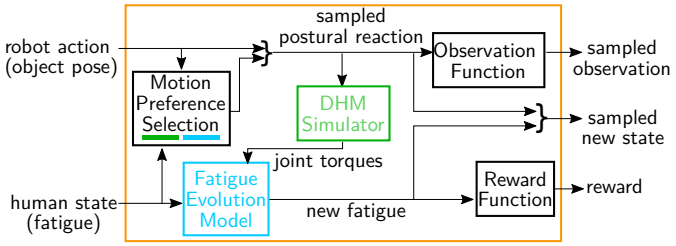


Fig. 4. Generative function of the POMDP. Given a state and a robot action, the Motion Preference Selection function estimates the probability of the human selecting each motion preference $P(u_t|f_t, a_t)$. A preference is then sampled according to the resulting distribution. The fatigue state is then evolved according to the sampled postural reaction, using the DHM simulator and the fatigue model. Finally, a new state, an observed reaction and a reward are sampled according to the transition, observation and reward functions respectively.

IV. Experiments

We validate our approach with a proof-of-concept experiment in simulation. This section describes the scenario and the specific model parameters.

A. Scenario

We consider a repetitive task comprised of $16\,s$ long cycles. At each cycle, the robot brings an object and the human reaches to it, then the human works on the object (supported by the robot) with a $5\,kg$ manual tool in its right hand for $6\,s$, and finally the robot takes the object away while the human goes back to the initial neutral posture (standing idle, see attached video). The actual working phase is representative of fine manipulation on the object, which, from a whole-body perspective consists in maintaining the same working posture (fingers motion not simulated). In this scenario, there is no significant human-robot interaction force (e.g., scanning the object), though the proposed approach allows to handle tasks with interaction forces.

B. Fatigue Model Parameters

We use values from the literature for the fatigue and recovery rates (Eq. 2), suggested and validated in [20], [21]: $k = 0.01667s^{-1}$ and $R = 0.04s^{-1}$ for all joints. The threshold α^j for switching between the fatigue/recovery modes is set to 10% of the nominal maximum exertable torque of each joint (such that it is below the permanent endurance limit recommended in [26]). The nominal maximum exertable torque values τ^j_{max} in both agonist and antagonist directions are the ones mentioned in [27]. Note that while we use average values from the literature for all the fatigue parameters here, user-specific values could be used instead. However the user-specific measurement process is out of scope of this work.

C. DHM Simulator Parameters

In this work, we manually define two human motion preferences, referred to as *arm reaction* and *back reaction*. The *arm reaction* corresponds to keeping the back as vertical as possible (avoid bending) in order to reduce the low-back torque (the reaching motion is done mainly with the arm).

The *back reaction* corresponds to keeping the upper-arm as vertical as possible in order to reduce the shoulder torque (reaching then involves significant back motion). These two motion preferences are implemented in the DHM simulator as LQP Cartesian orientation tasks, respectively on the back segment (*arm reaction*) and on the right upper-arm segment (*back reaction*). The weights w_i for the LQP tasks are: 1000 for the CoM task, 10 for the right hand task, 1 for the whole-body posture task, 1 for the back orientation task (active only in the *arm reaction*), and 10 for the upper-arm orientation task (active only in the *back reaction*).

D. POMDP Parameters

a) Action space: We define an action set A composed of two robot actions (i.e. two object poses): *action high* where the robot presents the object to the human at chest level (position w.r.t. center of the feet: $0.5\,m$ forward, $0.25\,m$ rightward, $1.2\,m$ high), and *action low* where the object is presented at knee level (position w.r.t. center of the feet: $0.4\,m$ forward, $0.15\,m$ rightward, $0.5\,m$ high). The orientation of the object is kept similar in both actions.

b) Transition function: $T(s, a, s')$ depends on the human postural reaction probability $P(u_t|s_t, a_t)$. In this proof-of-concept experiment, we define two motion preferences (Section IV-C), which, together with the two robot actions, generate four different postural reactions u_t. We consider that a human selects a motion preference based on the fatigue induced by the postural reaction to be executed, and on what we call the human expertise β (however, it does not depend on the previously selected postural reaction u_{t-1}). An expert ($\beta = \infty$) is able to predict the fatigue that a postural reaction will induce, therefore at each cycle s/he always chooses the least fatiguing motion preference w.r.t. his/her current fatigue state and the given object pose. Conversely, a novice ($\beta = 0$) is unaware of the consequences of his/her choice and selects any motion preference with the same probability. We formulate this using a Boltzmann distribution:

$$P(u_t|s_t, a_t) = P(u_t|f_t, a_t) = \frac{e^{-\beta\, c(u_t, f_t)}}{\sum_{u \in U} e^{-\beta\, c(u, f_t)}} \tag{4}$$

where $c(u_t, f_t) = C(f_{t+1})$ is the fatigue cost induced by reaction u_t starting from fatigue state f_t, computed based on the fatigue state at the end of the motion f_{t+1} (see Eq. 5).

c) Observation Function: The present experiment is conducted in simulation, therefore there is no sensor noise and we use a deterministic observation function $P(u_t^o|u_t) = 1$ if $u_t^o = u_t$, and 0 otherwise.

d) Reward Function: R is defined by the fatigue cost at the end of a cycle, so $r(s, a, s') = r(f_{t+1}) = -C(f_{t+1})$ with ($2N$ being the total number of agonist/antagonist DoFs):

$$C(f_t) = \sum_{j \in joints} \frac{(f_t^j)^2}{2N} \tag{5}$$

e) Initial belief: We test our framework with two different initial believes b_0: a single state belief (i.e. no uncertainty on the initial fatigue state) where the human is fully relaxed (zero fatigue on all joints), and a multiple state belief where the human could be fully relaxed, half-fatigued on the back flexion joint, or half-fatigued on the right hip joint with probabilities 0.2, 0.4 and 0.4 respectively. We test the single state b_0 for both novice and expert users, and the multiple state b_0 for a novice user only.

With a single state b_0 and an expert user, the problem is fully deterministic. In all other conditions (multiple state b_0 or novice user), the problem is non-deterministic because of the uncertainty on the initial fatigue state, or on the motion preference that the human will select at each cycle.

f) Discount factor: We use $\gamma = 0.95$ in this work.

E. Benchmark

We compare the performance obtained when executing the optimal policy computed with our approach (both with regular POMCP and POMCP-Max) with three other policies:

- *Random:* randomly returns an action $a \in A$ at every cycle (baseline policy). Note that with only two robot actions, random could perform well especially with an expert who always chooses the least fatiguing reaction.
- *Alternating:* alternates between the two robot actions regardless of the human fatigue state.
- *Greedy:* returns, at each cycle, the action with the lowest expected fatigue cost after one single cycle.

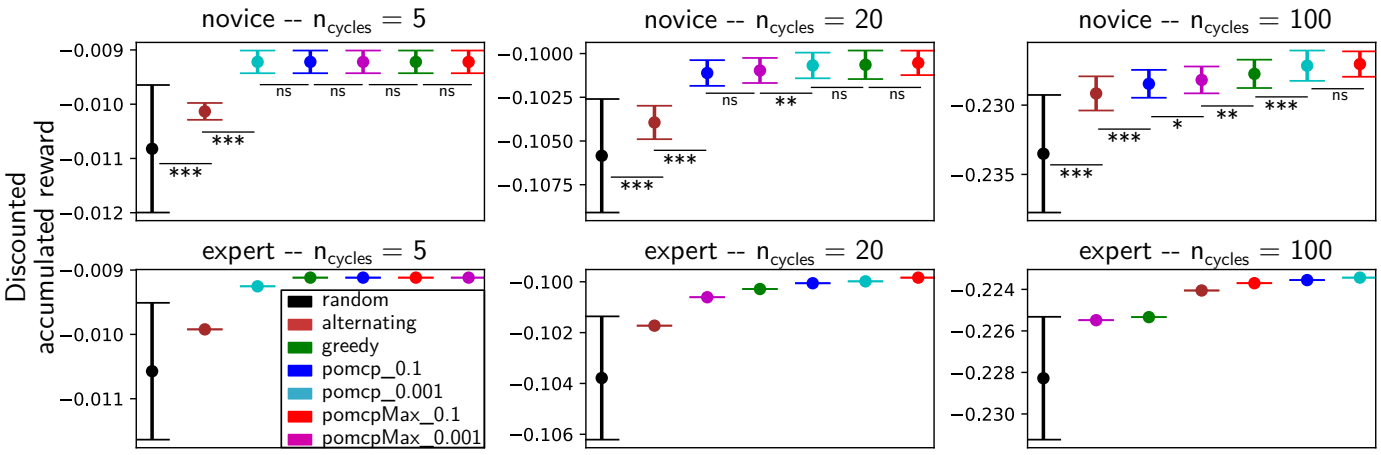
We compute the optimal policy with POMCP and POMCP-Max using 5000 descents in the search tree, and test 2 values (0.1 and 0.001) for the exploration/exploitation factor.

For each test condition (initial belief and human expertise), we test three different numbers of cycles for the full task, $n_{cycles} \in \{5, 20, 100\}$. For each number of cycles, we execute each policy 100 times using the POMDP generative model to account for the stochasticity of the problem. We use the discounted accumulated reward $\sum_{t=0}^{n_{cycles}} \gamma^t r_t$ over each execution as performance metrics. We perform t-test comparisons with a 5% significance level on the reward values to compare the different policies[4].

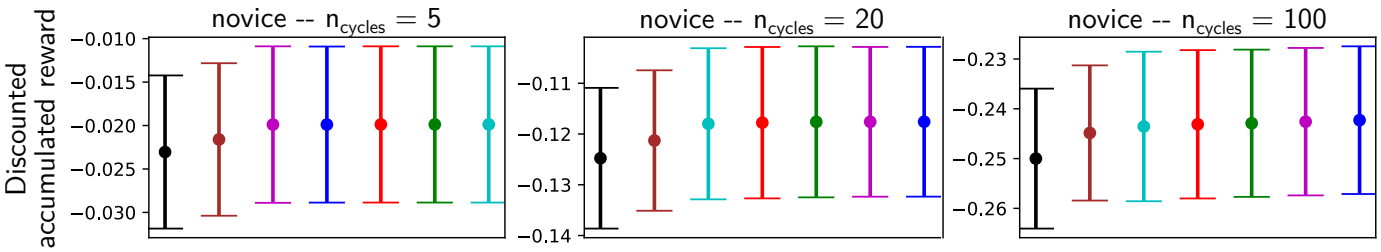
V. Results

a) Single state initial belief: Fig. 5a displays the distribution of the discounted accumulated reward for each policy and each condition (novice/expert and 3 values of n_{cycles}). For a novice user, with $n_{cycles} = 5$, the greedy policy and all POMCP versions are not significantly different from one another, and all significantly outperform the alternating and random policies. With $n_{cycles} = 20$, greedy, POMCP ($c = 0.001$) and POMCP-Max ($c = 0.1$) have equivalent performance and significantly outperform the other policies. With $n_{cycles} = 100$, POMCP ($c = 0.001$) and POMCP-Max ($c = 0.1$) have equivalent performance and significantly outperform all other policies.

[4]For the single state b_0 with an expert user, we ran 100 trials only for the random policy, since with all other policies the problem is fully deterministic and a single trial is sufficient to evaluate the reward. For this condition, we therefore compare the policies using the reward value of the single trial, and mean value over the 100 trials for the random policy.



(a) Single state initial belief, novice and expert users.



(b) Multiple state initial belief, novice user.

Fig. 5. Discounted accumulated reward for all tested policies, for the different test conditions. The reward is computed using the opposite of the fatigue cost (Eq. 5), hence the negative values, and better performing policies have reward values closer to zero. For a novice, the plots show the average and standard deviation over the 100 trials. Stars represent the statistically significant differences (*ns*: not significant). For the sake of clarity, policies are ordered according to performance, and we only display significance between adjacent policies. For an expert, the problem is deterministic except for the random policy, so the plots show the one single reward value per policy, and the mean and standard deviation over 100 trials for the random policy. *pomcp* stands for the regular POMCP, and *pomcpMax* for the POMCP-Max variant. Numbers next to POMCP and POMCP-Max are the values of the exploration/exploitation factor c.

For an expert user, with $n_{cycles} = 5$, the greedy policy, POMCP ($c = 0.1$) and POMCP-Max (both c values) have the same and best performance, shortly followed by POMCP ($c = 0.001$). With $n_{cycles} = 20$, POMCP-Max ($c = 0.1$) performs better than all other policies, followed by POMCP (both c values) and then greedy. With $n_{cycles} = 100$, POMCP (both c values) and POMCP-Max ($c = 0.1$) have close performance, and outperform all other policies. Notably, POMCP-Max ($c = 0.001$) shows low performance.

b) Multiple state initial belief: Fig. 5b displays the discounted accumulated reward for each policy, for a novice user. The differences between the policies are not statistically significant (except for the comparison with the random policy), because of the large standard deviation. Yet, it terms of means, POMCP ($c = 0.001$), POMCP-Max ($c = 0.1$) and greedy have the same and best mean when $n_{cycles} = 5$, shortly followed by POMCP ($c = 0.1$) and POMCP-Max ($c = 0.001$). With $n_{cycles} = 20$, POMCP ($c = 0.1$), POMCP-Max ($c = 0.001$) and greedy have the same and best mean. With $n_{cycles} = 100$, POMCP ($c = 0.1$) has the best mean, followed by POMCP-Max ($c = 0.001$) and then greedy.

VI. Discussion

In our proof-of-concept experiment, the policies computed with the POMCP algorithms allowed to reduce long-term human fatigue compared to the other tested policies. Our results do not allow to conclude on one best-performing POMCP version (though the regular POMCP algorithm seems slighlty less sensitive to the tuning of the exploration/exploitation factor c compared to POMCP-Max), but this was not the main focus of our work. Overall, POMCP policies showed better performance than policies which did not consider the human fatigue to select the robot action (random and alternating) in all tested conditions. For tasks with a small number of cycles, the greedy policy performed as well as our approach, which was expected since greedy considers the short term consequences of an action. Conversely, in highly-repetitive tasks (large number of cycles), our long-term planning approach outperformed greedy. Importantly, our approach showed good performance on a variety of scenarios: fully-deterministic problem (expert user with single-state initial belief b_0) which shows the benefit of long-term planning even without uncertainty, stochastic problem with full-observability (novice user with single-state b_0), and stochastic problem with partial observability (multiple-state b_0). This demonstrates the genericity of our framework.

Yet, the proof-of-concept example used in this work has some limitations. The main one is that we manually defined human motion preferences and associated probability distribution $P(u_t|s_t, a_t)$. While we chose them to be sensible w.r.t. how a human could react, we cannot guarantee that they

match real human behaviors. In a future work, we will address this limitation by identifying human motion preferences from human data. This is indeed a required step to validate our approach on a real human subject experiment. The set of possible robot actions was also manually defined among the robot workspace and might not be optimal.

However, the goal of this paper was to present a proof-of-concept of our framework. The different elements can readily be adapted to handle more complex scenarios. For instance, a stochastic observation function can be used to account for uncertainty in the identification of the human postural reaction in a real-world experiment. Additional user specificities or preferences can also be introduced as hidden variables to be inferred during the interaction process. For instance, uncertainty could be introduced on the parameters of the fatigue model (thereby making the fatigue evolution model non-deterministic) to further personalize the approach.

VII. Conclusion

We proposed a framework to mitigate long-term human physical fatigue in highly repetitive human-robot collaborative tasks, which takes into account the uncertainty in the human postural reaction to the robot motion, and the partial observability of the human fatigue. Our POMDP-based approach showed promising results on a proof-of-concept example, in a variety of user-related scenarios. In future work, we intend to validate our framework on a real human subject experiment, and test it on different co-manipulation tasks involving significant interaction forces and more complex human and robot motions, such as co-displacing heavy objects with a robot.

References

[1] de Kok, Jan and Vroonhof, Paul and Snijders, Jacqueline and Roullis, Georgios and Clarke, Martin and Peereboom, Kees and van Dorst, Pim and Isusi, Iigo, *Work-related Musculoskeletal Disorders: prevalence, costs and demographics in the EU*. Publications Office, 2019.

[2] R. Davies, "Industry 4.0: Digitalisation for productivity and growth | Think Tank | European Parliament," 2015.

[3] B. Busch, G. Maeda, Y. Mollard, M. Demangeat, and M. Lopes, "Postural optimization for an ergonomic human-robot interaction," in *Int. Conf. on Intelligent Robots and Systems*, 2017, pp. 2778–2785.

[4] W. Kim, J. Lee, L. Peternel, N. Tsagarakis, and A. Ajoudani, "Anticipatory Robot Assistance for the Prevention of Human Static Joint Overloading in Human-Robot Collaboration," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 2017.

[5] L. Peternel, C. Fang, N. Tsagarakis, and A. Ajoudani, "A selective muscle fatigue management approach to ergonomic human-robot co-manipulation," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 69–79, 2019.

[6] European Agency for Safety and Health at Work., "E-fact 9 - Work-related musculoskeletal disorders (MSDs): an introduction | Safety and health at work EU-OSHA," 2007.

[7] D. Srinivasan and S. E. Mathiassen, "Motor variability in occupational health and performance," *Clinical Biomechanics*, vol. 27, no. 10, pp. 979–993, 2012.

[8] A. W. Drake, "Observation of a Markov process through a noisy channel," Thesis, Massachusetts Institute of Technology, 1962.

[9] D. Silver and J. Veness, "Monte-Carlo Planning in Large POMDPs," in *Advances in Neural Information Processing Systems*, vol. 23, 2010.

[10] A. Merikh-Nejadasl, I. El Makrini, G. Van De Perre, T. Verstraten, and B. Vanderborght, "A generic algorithm for computing optimal ergonomic postures during working in an industrial environment," *Int. J. of Industrial Ergonomics*, vol. 84, p. 103145, 2021.

[11] A. Shafti, A. Ataka, B. U. Lazpita, A. Shiva, H. Wurdemann, and K. Althoefer, "Real-time Robot-assisted Ergonomics," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 1975–1981, iSSN: 2577-087X.

[12] A. M. Zanchettin, E. Lotano, and P. Rocco, "Collaborative Robot Assistant for the Ergonomic Manipulation of Cumbersome Objects," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6729–6734, iSSN: 2153-0866.

[13] L. v. der Spaa, M. Gienger, T. Bates, and J. Kober, "Predicting and Optimizing Ergonomics in Physical Human-Robot Cooperation Tasks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE, 2020, pp. 1799–1805.

[14] W. Gomes, P. Maurice, E. Dalin, J.-B. Mouret, and S. Ivaldi, "Multi-Objective Trajectory Optimization to Improve Ergonomics in Human Motion," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, 2022.

[15] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 618–634, 2017.

[16] A. Bestick, R. Pandya, R. Bajcsy, and A. D. Dragan, "Learning human ergonomic preferences for handovers," in *International Conference on Robotics and automation (ICRA)*. IEEE, 2018, pp. 3257–3264.

[17] E. Rashedi and M. A. Nussbaum, "A review of occupationally–relevant models of localised muscle fatigue," *International journal of human factors modelling and simulation*, vol. 5, no. 1, pp. 61–80, 2015.

[18] T. Xia and L. A. F. Law, "A theoretical approach for modeling peripheral muscle fatigue and recovery," *Journal of biomechanics*, vol. 41, no. 14, pp. 3046–3052, 2008.

[19] J. Ding, A. S. Wexler, and S. A. Binder-Macleod, "A predictive model of fatigue in human skeletal muscles," *Journal of applied physiology*, vol. 89, no. 4, pp. 1322–1332, 2000.

[20] L. Ma, W. Zhang, D. Chablat, F. Bennis, and F. Guillaume, "Multi-objective optimisation method for posture prediction and analysis with consideration of fatigue effect and its application case," *Computers & Industrial Engineering*, vol. 57, no. 4, pp. 1235–1246, 2009.

[21] L. Ma, D. Chablat, F. Bennis, and W. Zhang, "A new simple dynamic muscle fatigue model and its validation," *International Journal of Industrial Ergonomics*, vol. 39, pp. 211–220, 2009.

[22] N. R. Council and I. o. Medicine, *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*. Washington, DC: The National Academies Press, 2001.

[23] P. Maurice, V. Padois, Y. Measson, and P. Bidaud, "Digital human modeling for collaborative robotics," in *DHM and Posturography*. Elsevier, 2019, pp. 771–779.

[24] J. Salini, V. Padois, and P. Bidaud, "Synthesis of complex humanoid whole-body behavior: A focus on sequencing and tasks transitions," in *Int. Conf. on Robotics and Automation*. IEEE, 2011, pp. 1283–1290.

[25] T. Keller and M. Helmert, "Trial-Based Heuristic Tree Search for Finite Horizon MDPs," *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 23, pp. 135–143, 2013.

[26] W. Rohmert, "Ergonomics: concept of work, stress and strain," *Applied Psychology*, vol. 35, no. 2, pp. 159–180, 1986, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1464-0597.1986.tb00911.x.

[27] P. Maurice, "Virtual ergonomics for the design of collaborative robots," Ph.D. dissertation, Mechanical, Accostic, Electronic and Robotics Doctoral school of Paris, University of Pierre and Marie Curie, 2015.