Mixed-initiative mission planning considering human operator state estimation based on physiological sensors

Nicolas Drougard[1], Caroline P. Carvalho Chanel[1], Raphaëlle N. Roy[1] and Frédéric Dehais[1]

***Abstract*— Missions involving humans with automated systems become increasingly common and are subject to risk of failing due to human factors. In fact, missions workload may generate stress or mental fatigue increasing the accident risk. The idea of our project is to refine human-robot supervision by using data from physiological sensors (eye tracking and heart rate monitoring devices) giving information about the operator's state. The proof of concept mission consists of a ground robot, autonomous or controlled by a human operator, which has to fight fires that catch randomly. We proposed to use the planning framework called Partially Observable Markov Decision Process (POMDP) along with machine learning techniques to improve human-machine interactions by optimizing the decision of the mode (autonomous or controlled robot) and of the display of alarms in the form of visual stimuli. A dataset of demonstrations produced by remote volunteers through an online video game simulating the mission allows to learn a POMDP that infers human state and to optimize the associated strategy. Cognitive availability, current task, type of behavior, situation awareness or involvement in the mission are examples of studied human operator states. Finally, scores of the missions, consisting in the number of extinguished fires, will quantify the improvement made by using physiological data.**

I. INTRODUCTION

Since more than a century the use of automated systems began to be increasingly common: assembly lines, autopilots in aircrafts, autonomous cars, unmanned vehicles such as drones or ground robots, for military operations or contaminated area and many others. This phenomenon has been amplified by the recent technical advances in artificial intelligence and machine learning, providing even more autonomy to machines. As an example, convolutional networks led to artificial vision [1] and popularized deep learning techniques, which played an important role in the latest successes of decision making algorithms based on reinforcement learning [2] and planning [3].

Despite the impressive progress of autonomy in machines, human operators are still vital in many scenarios. In fact, humans are deemed to be creative, highly flexible (they can handle complex or unknown situations) and able to produce tactical, moral, social and ethical decisions [4], while these qualities are rarely attributed to machines. Humans and machines possess complementary strengths as illustrated by the Fitts list [5], providing a general summary of the best from both humans and machines. Further explanations of the key role of human operators concerns societal reasons. For instance, legal regimes need people for responsibility assessment issues, encouraging human supervision of automated systems. More generally, interactions between humans and machines are inevitably growing due to the increasing use of machines. Despite the drastic change of the human operator role in favor of system autonomy, they are still involved in missions as members of human-machine teams or as supervisors [4].

Considering studies about existing human-machine systems, for example in aviation [6] and [7], it appears that human error represents a major cause of accident. In details human factors are involved in 80% of autonomous aerial vehicles accidents [8]. This fact is due to several constraints experienced by humans during their missions. Stress, high workload, fatigue or boredom, which can be induced respectively by pressure (e.g. cause by a danger), complexity, hardness or duration of their tasks, are some of the main criticalities for humans. These weaknesses may affect human abilities by producing mental confusion, attentional tunneling [9], mind wandering [10], lower vigilance, etc. As a consequence the mission may fail or at least be achieved in a sub-optimal manner.

In mixed-initiative missions, *i.e.* missions benefiting from the skills of both computer and human operator, the last is often thought as omniscient and able to fix any occurring issue during the process. In order to reduce failures of mixed-initiative missions, our project starts by considering human operators as simple members of the human-machine teams, not always reliable because of the constraints indicated above. In other words, this paper proposes to address the problem of improving human-machine missions by computing a supervision strategy taking into account the human operator along with the evolution of the human mental state. To this purpose, the strategy is meant to modify the mental state of the human operator when it is damaging to the mission. It has also to ensure that the current task is appropriate to the human operator's features and mental states. Finally, the machine has to adapt its own behavior in function of the detected human operator's condition and the mission status.

An appropriate supervision strategy has to manage the information given to the human operator, the task allocation between the human and the machine, as well as the machine policy during its own tasks. In other words, our goal is to refine supervision strategy of human-machine teams firstly by providing, or not, appropriate alarms to the human operator, secondly by allocating the tasks that can be carried out by both the human and the machine, and finally by adapting

This work was supported by Dassault Aviation.

[1]Nicolas Drougard, Caroline P. Carvalho Chanel, Raphaëlle N. Roy and Frédéric Dehais are with the Aerial & Spacial vehicle Conception Department, ISAE-Supaero, 31400 Toulouse, France firstname.name@isae-supaero.fr.

actions of the machine according to the human behavior and mental state.

As the events that occur during a mixed-initiative mission are uncertain due to the human behavior as well as the random dynamics of most environments, the framework of Probabilistic Planning has been selected: this paper considers Markov Decision Processes (MDP) [11] allowing to define the goal of a given mission in terms of rewards valuating states of the system. The optimization of MDPs consists in computing a strategy maximizing the expected sum of rewards over time [12]. More precisely this paper deals with Partially-Observable Markov Decision Processes (POMDP) [13] reflecting possible unobservable states of the process. In fact, human mental states are intrinsically not observable but it appears essential, as explained above, to take them into account to optimally drive mixed-initiative missions. These states can be inferred using physiological data from the human operator. For instance, relevant data can be the gaze position, the heart beats frequency, the electrical activity of the brain or its relative changes in hemoglobin concentration. The processes meant to collect them are respectively called *eye tracking* (ET), *electrocardiography* (ECG), *electroencephalography* (EEG) and *functionals near-infrared* (fNIR). These measures allow for properly estimating the human operator mental state and eventually triggering external action to maximize operator efficiency in the mission.

The POMDP design requires a probabilistic modeling of the human-machine system under study, practically involving the hard task of defining the probability values of some events of interest. The complexity consists in gathering a sufficiently large dataset of demonstrated sequences of human actions on the given interface during the task execution. To obtain precise probability distributions this activity implies time consuming experiment and numerous voluntary operators. In complex human-machine systems, situations with a very low probability of occurrence are not well represented, if not totally missing, in experimental datasets, risking to affect the estimation of the true events probabilities. The frequency of an event appearing in one of these situations is likely to be a poor estimation of the actual probability value because the sample used to compute it is too small (if not empty). To ensure a sufficient amount of demonstrations, we propose here to call on remote volunteers who can participate to our experiments through a website simulating the mission of interest and accessible at the following address: `www.humanrobotinteraction.fr`. Of course, physiological data have to be collected *in situ* with lab's devices, resulting to a smaller dataset. Hence, one of the challenges of the presented work will be to cross-check data (from remote and *in situ* volunteers) in order to define the events probabilities defining the POMDP.

The rest of the paper is organized as follows. Section II is dedicated to the description of the proof of concept mission along with details on its technical implementation, including the website. Section III describes preliminaries and POMDP design methodologies, employing the virtual demonstrations dataset and the physiological collected dataset to enhance next human-machine supervisions. Related conclusions are reported in Section IV.

II. Proof of concept mission

This section is devoted to the description of a mixed-initiative mission which has to be general enough to allow the evaluation of methods meant to improve human-machine supervision. Firstly we propose to consider a mission involving a human-robot team since robots are widely used in practice to manage tasks in dangerous (*e.g.* war), contaminated (*e.g.* nuclear plants) or remote (*e.g.* planet exploration) areas. Secondly it is also desirable that the mission reflects realistic situations in terms of technical equipment and environment in order to prove the usability of the developed methods in practice. As explained at the end of this section the mission will be implemented in our lab as the considered robot is available there and the environment easily reproducible.

Thirdly, as mentioned in introduction, the human mental states seem to be good features that we can build upon to drive human-robot interaction (HRI), hence a good proof of concept mission should produce fluctuations of these states. As a consequence it is important that the considered mission seems dangerous by involving unsafe events and implies pressure and stress due to a quantitative evaluation (score) of the mission achievement. It is also required that the tasks allocated to the human operator are demanding or complex implying cognitive workload. These constraints enable the appearance of undesirable mental states for the mission such as attentional tunneling, mental confusion, mind wandering, low vigilance etc.

Finally let us remind that this study delves into the optimization of mixed-initiative missions. If we want to make the mission achievements "better", the latter should be quantifiable *i.e.* the goal of the mission has to be defined in terms of valuation of system states. This constraint is already raised above with the notion of evaluation, or score, meant to produce pressure to the human operator.

A. A *firefighter robot*

Now that the needed aspects have been explicitly given, let us describe the resulting mission: a firefighter robot is present in a small area with few trees which have a weird tendency to self-ignite for some unknown reason. Through a graphical user interface (GUI) which appears in Figure 1, the human operator gets the position of the robot in a map (bottom center), as well as the video from its camera (top right).

The battery charge level of the robot decreases with time and with respect to its actions: when the robot is in the charging station, represented by a red square on the ground, the battery recharges. If the battery is empty and the robot is not on the red square, the mission fails and is finished. All the information related to the robot is summarized on the bottom left of the GUI. The volume of water contained by the robot is not unlimited: to recharge in water, the robot has to be in the water station represented by a blue square on the ground and the associated tank has to contain enough water. For that,

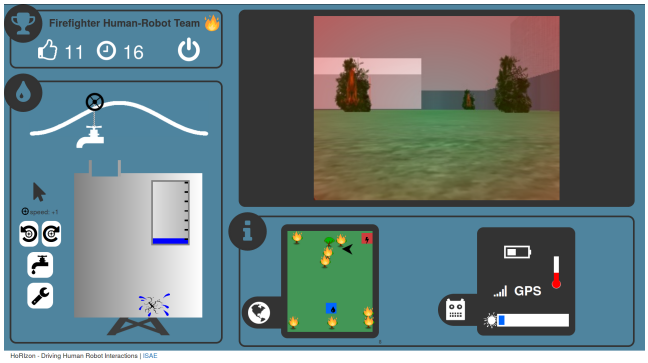

Fig. 1. Graphical user interface of the Firefighter Robot mission.

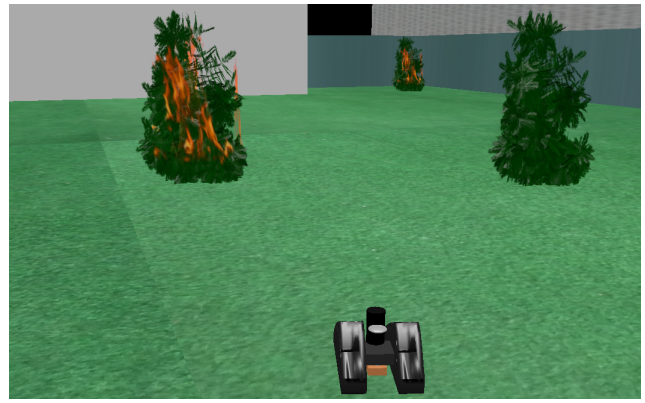

Fig. 2. Simulation of the robot in its environment with MORSE

the human operator has to fill this tank using the buttons on the left-side of the interface: a tap, which can move horizontally by actioning a wheel (top buttons), fills the tank when it is in the middle (which is an unstable equilibrium). To actually fill the tank, the button bellow (black tap) turns on the tap for few seconds. Some leaks may appear on the tank during the mission emptying it: the button below (black wrench) can be used to fix them.

With the help of this robot, the goal of the mission is to fight as many fires as possible in a limited amount of time (top left information on the GUI). The robot is controlled by the arrows (navigation) and the space bar (shoot water) of the keyboard but it can become autonomous at any time. Finally the temperature of the robot increases when it is too close to flames and the mission terminates when it is too hot. The presence of fires is supposed to be felt as a danger by the operator. Temperature and battery managements as well as the score (number of extinguished fires) and the remaining time should imply stress and pressure. Finally, as observed during pre-testing, both tasks (robot control and water management) are complex enough to generate cognitive workload.

B. Implementation details

A simulation of the robot and its environment has been realized with MORSE[1] [14] as shown in Figure 2. It allows remote volunteers to participate in the experiment by performing the mission on the website `www.humanrobotinteraction.fr`. Thanks to this website the size of the resulting dataset of realized missions is higher than with experiments in the lab with *in situ* volunteers.

The control of the real robot as well as its own simulation use the Orocos[2] library [15]. The GUI (and the full website) was implemented using angular-fullstack[3] framework so that the real robot can be linked either to the real robot or to the simulation. It makes the missions based on the simulation as similar as possible to the missions with the real robot.

[1] `https://www.openrobots.org/morse/doc/stable/morse.html`
[2] `http://www.orocos.org/`
[3] `https://github.com/angular-fullstack/generator-angular-fullstack`

Now that a proof of concept mission and the way to record more data on it have been described, let us have a look on the proposed method to improve the achievement of a human-robot mission.

III. POMDP for Driving Human-Machine Interaction

This study proposes to use the framework of POMDP [13] to compute a supervision strategy for HMI mission improvement. This section starts with a short description of this framework and the definition of a POMDP in the context of the presented mission. Follows a general method for learning the parameters of such a POMDP from available datasets. Finally other avenues based on expert assumptions and related works are described.

A. POMDP framework in HMI context

A Partially Observable Markov Decision Process (POMDP) is used to model an agent in a probabilistic system whose states generate observations but are not directly observable. It allows to compute optimal successive actions in terms of reward expectation. More formally, it can be defined by a tuple $\langle \mathcal{S}, \mathcal{A}, \Omega, \mathcal{T}, \mathcal{O}, r, b_0 \rangle$ where:

- $\mathcal{S}$ is the finite set of unobservable states,
- $\mathcal{A}$ is the finite set of possible actions,
- Ω is the finite set of observations,
- $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0,1]$ is the transition function: $\forall (s,s') \in \mathcal{S}^2$, $\forall a \in \mathcal{A}$, $\mathcal{T}(s,a,s') = \mathbf{p}\left(s' \mid s,a\right)$ *i.e.* the probability that the next state is equal to s' when the current state is s and the chosen action is a,
- $\mathcal{O} : \mathcal{A} \times \mathcal{S} \times \Omega \rightarrow [0,1]$ is the observation function: $\forall a \in \mathcal{A}$, $\forall s' \in \mathcal{S}$, $\forall o' \in \Omega$, $\mathcal{O}(a,s',o') = \mathbf{p}\left(o' \mid s',a\right)$ *i.e.* the probability of observing o' if action a is chosen leading to state s',
- $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function defining the goal of the mission by valuating state-action pairs, $r : \mathcal{S} \rightarrow \mathbb{R}$ the terminal reward function,
- $b_0 : \mathcal{S} \rightarrow [0,1]$ is the initial belief state: $\forall s \in \mathcal{S}$, $b_0(s)$ is the probability that the initial state is s.

The *belief* state at a time step $t \in \mathbb{N}$ is the name given to the best estimation of the hidden state: $b_t(s) =$

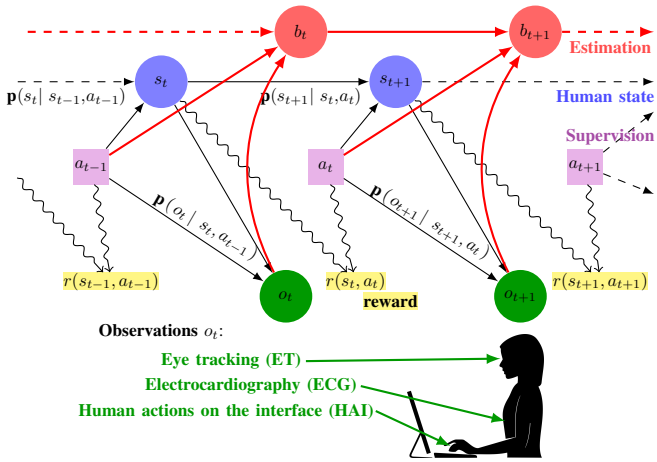

Fig. 3. Graphical representation of a POMDP for driving HMI

$\mathbb{P}(s_t = s \mid a_0, o_1, \ldots, a_{t-1}, o_t)$. It can be computed recursively after executing $a_t \in \mathcal{A}$ and the reception of a new observation $o_{t+1} \in \Omega$:

$$b_{t+1}(s') = \frac{\mathcal{O}(a_t, s', o_{t+1}) \sum_{s \in \mathcal{S}} \mathcal{T}(s, a_t, s') b_t(s)}{\sum_{(s,s') \in \mathcal{S}^2} \mathcal{O}(a_t, s', o_{t+1}) \mathcal{T}(s, a_t, s') b_t(s)}.$$

Optimal strategies are usually based on the belief state *i.e.* given a horizon $T \in \mathbb{N}$, an optimal strategy can be defined as a sequence of functions $(d_t)_{t=0}^{T-1}$ such that $d_t : b_t \mapsto a_t \in \mathcal{A}$ and maximizing

$$\mathbb{E}_{s_0 \sim b_0} \left[\sum_{t=0}^{T-1} r\Big(s_t, d_t(b_t)\Big) + r(s_T) \right].$$

A graphical representation of a POMDP is depicted in Figure 3.

In the context of the Firefighter Robot mission presented in the previous section it is possible to define the set of actions as the Cartesian product of $\mathcal{A}_D$ the set possible visual alarms (display of a window on the user interface), $\mathcal{A}_A = \{autonomous, manual\}$ defining if the robot is autonomous or controlled manually (task allocation), and $\mathcal{A}_R = \{forward, backward, left, right, shoot\}$ the set of possible robot actions when autonomous: $\mathcal{A} = \mathcal{A}_D \times \mathcal{A}_A \times \mathcal{A}_R$. Some possible visual alarms present in A_D are "low battery", "too-high temperature", "60 seconds before the end of the mission", "robot's tank will soon be empty (2 shoots left)", "robot is in autonomous mode", "robot is in manual mode" and "ground tank's water level is low". Note that simulated missions with remote volunteers are used for learning purposes. Thus during these missions, if a situation described by an alarm in $\mathcal{A}_D$ is experienced, this alarm is displayed with probability 0.5. Decision of the autonomy ($a \in \mathcal{A}_A$) is also randomized. However an expert strategy is used for robot actions: they can be optimized using RL once the human behavior can be simulated.

As explained in [16], identifying visible and hidden variables reduces computational time of POMDP solving. It is also convenient here to make explicit which data is observable and may be used as input of the computed strategy:

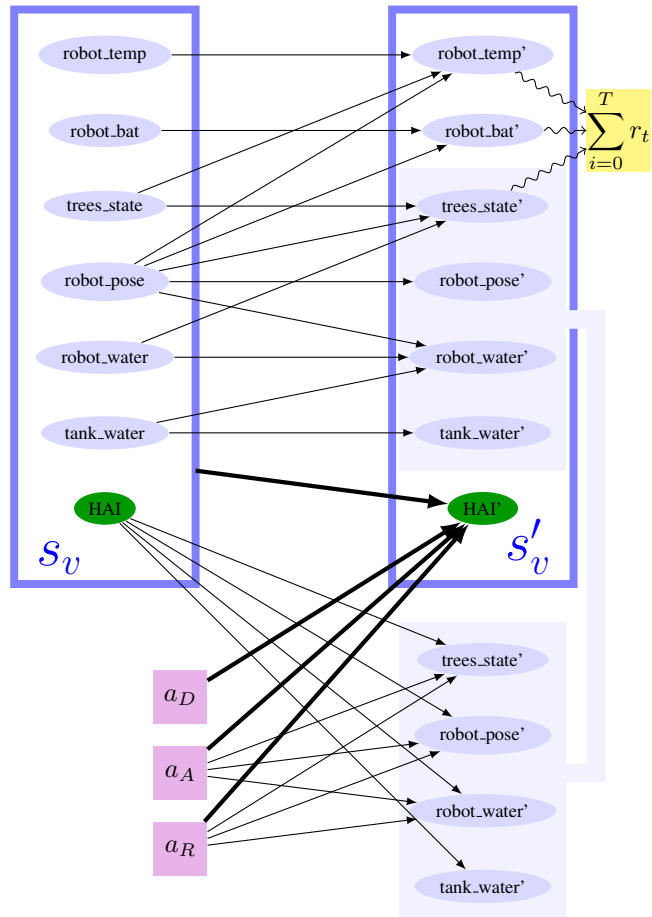

Fig. 4. Graphical representation of the variables describing the system states. If the name of a given current state variable is s, the corresponding next state variable is denoted by s'. For clarity four next state variables, depending on both system variables and human-machine actions, are duplicated. Missing arrows represent independencies. For instance, robot temperature only depends on the states of the trees, its pose (proximity to trees on fire) and its previous temperature.

the state space can be written as the Cartesian product of a visible state space $\mathcal{S}_v$ and a hidden state space $\mathcal{S}_h$: $\mathcal{S} = \mathcal{S}_v \times \mathcal{S}_h$. The visible state space $\mathcal{S}_v$ contains the directly visible states of the system: robot, trees and container states, made explicit in Figure 4, along with the human actions on the user interface (HAI) *i.e.* the clicks on buttons and the use of keyboard keys. More precisely and as shown in this figure a state $s_v \in \mathcal{S}_v$ is a tuple of variable assignments $s_v = \left(s_v^i\right)_{i=1}^{7} =$ (robot_temp, robot_bat, trees_state, robot_pose, robot_water, tank_water, HAI). The reward is a function of trees_state since the goal of the mission is to extinguish fires: $r(s_v) = r(\text{trees\_state}) = \#\{\text{safe trees}\}$. While the reward at each time step only depends on variable trees_state, the final sum of rewards is also affected by the variables robot_temp and robot_bat. Indeed when robot_bat is zero or robot_temp is greater than a given threshold, the robot becomes inoperable insuring the most possible expected number of trees on fire during all the next mission steps.

B. Using a simple MDP

A first idea could be to learn the parameters of the resulting MDP (a POMDP with visible states $\mathcal{S}_v$ and no observations) whose graphical model is depicted in Figure 4. Since lack of arrows represents independencies in a Bayesian Network [17], the transition function of the process can be computed from marginal ones $\mathbf{p}\left(s_v' \mid s_v, a\right) = \prod_{i=1}^{7} \mathbf{p}\left((s_v^i)' \mid s_v^i, a\right)$. The dynamics of HAI variable, represented with thick arrows in Figure 4, is unknown and should be defined using the dataset of demonstrations resulting from the online mission (available at `www.humanrobotinteraction.fr`). Denoting the first six variables by "non_human", probabilities $\mathbf{p}\left((s_v^7)' \mid a, s_v\right) = \mathbf{p}\left(HAI' \mid a, HAI, \text{non\_human}\right)$ are simply computed as the ratio between the number of transitions from $(HAI, \text{non\_human})$ to HAI' selecting action $a \in \mathcal{A}$, and the number of events $(a, HAI, \text{non\_human})$, both numbers being counted over every transitions of each demonstration. Solving the computed MDP leads to an optimal strategy δ^* defining at each time $t < T$ an action according to the current state: $\delta_t^* : S_v \to \mathcal{A}$.

Obviously the spaces of the state variables $\left(s_v^i\right)_{i=1}^{7}$ have to be discretized according to the size of the dataset in order to make the ratio defining the estimated probabilities statistically significant. While non_human variables have a straightforward discretization, more work is necessary for variable HAI: even if the MDP step $t \in \mathbb{N}$ is often incremented, say at each second, the human operator may interact continuously with the system and thus provide a very specific stream. The finite space of HAI values can be built using clustering techniques [18] or by considering sound features of the stream (e.g. number/length of each interaction with the interface). In the same way, the ET and ECG streams from human operators are not used as such: common metrics of these streams, according to the literature, define the set of observations of the POMDPs described in the next section: for instance, current area of interest (AOI), frequency of saccades and durations of fixations are usual metrics for ET streams. Inter-beat intervals and standard deviation of them should be sufficient information from ECG streams to infer mental states [19].

C. Learning an accurate POMDP

As discussed in introduction, human states may influence human actions which affect system states and thus rewards. Taking them into account should thus improve the strategy for driving human-robot interactions. These human states can be represented by hidden states $s_h \in \mathcal{S}_h$ of the POMDP and estimated using streams of ET and ECG from human operator. Note that variable HAI can be also considered as an observation of the human state [20]. As POMDPs are nothing more than Hidden Markov Models (HMMs) enhanced with actions and rewards, the Baum-Welch algorithm [21] can be almost directly applied to learn unknown transition and observation functions as well as hidden states [22]. These unknown functions are represented by thick arrows in Figure 5: $\mathbf{p}\left(s_h' \mid \text{non\_human}, s_h, a\right)$,

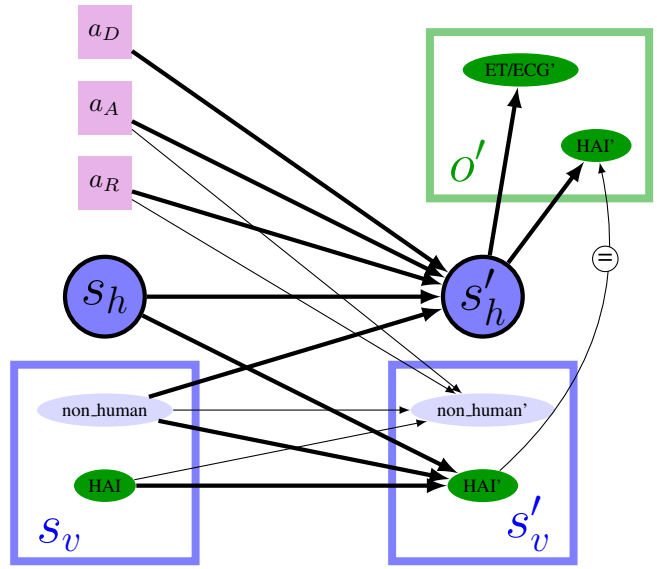

Fig. 5. POMDP taking into account hidden states $s_h \in \mathcal{S}_h$ associated to the human operator: variable HAI is dependent on the hidden states whose dynamics change according to the supervision actions $a \in \mathcal{A}$. These states can be inferred using ET, ECG and HAI streams.

$\mathbf{p}\left(\text{HAI'} \mid \text{non\_human}, \text{HAI}, s_h, a\right)$ and $\mathbf{p}\left(o' \mid s_h'\right)$ where $o' = (ET', ECG', HAI')$. The other transition functions are known and are not updated when using Baum-Welch algorithm. Let us recall that the size of the dataset produced by remote volunteers is far higher than the dataset of demonstrations with available ETG/ET streams (less volunteers *in situ*). Baum-Welch algorithm can be used on both type of demonstrations: when ET/ECG streams are available, all the parameters are updated, otherwise parameters associated to ET/ECG streams do not change. Note that spectral methods can also be used [23] to learn a HMM.

The POMDP can be also enhanced using existing classifiers or expert knowledge on the process: for instance [9] proposes a classifier using ET/ECG streams to detect attentional tunneling of the human operator. It can be used to directly define a human variable (tunneling) and set (or at least initialize) the observation function with the error probabilities as done in [24]. Next three examples explicitly define hidden state variables using observable data, making the learning method as simple as learning an MDP (see Section III-B). The first example is used in [25]: types of human operators are defined by clustering the set of HAI sequences. The transition function of this human state variable is the identity function since the human operator is supposed to be the same during all the mission. The second example is used in [26] defining the current human task (or unavailability) using the current AOI provided by ET stream. The last idea is to define a human state variable at step t as the sum of rewards received from t to $t + d$ with $d \in \mathbb{N}$. Indeed we are particularly interested in human states influencing the success of the mission.

Solving the resulting POMDP leads to an optimal strategy which can be compared to other strategies (e.g. from the

MDP or an expert) by proceeding in validation tests with *in situ* volunteers and evaluating the averaged sums of rewards of each strategies.

D. Reinforcement Learning

Each reward is observable as a function of a visible state variable (namely trees_state). Hence it seems interesting to use Reinforcement Learning (RL) to optimize the choice of actions $a \in \mathcal{A}$ over time. The dataset of demonstrations with available ECG/ET streams is unfortunately too small for the RL framework. However RL algorithms for POMDPs such as [27], [28], [29] can be used with the large dataset from the website or directly on it (only with HAI observations). Moreover when an appropriate model has been built, RL can be used to optimize robot actions $a_R \in \mathcal{A}_R$ by simulating human operator interaction with the system.

IV. Conclusion

This paper describes a mission used to prove that POMDP and machine learning frameworks can improve human-robot interactions and mixed-initiative missions by using physiological sensors. The mission is proposed in a website to produce a sizable dataset for learning. A general way to optimize the supervision actions has been proposed, as well as specific enhancements. We hope that this project will highlight the need for taking into account particular mental states or situations, through physiological data from human operator. It should also demonstrate the need for training the human-machine supervising system (through a remote simulation as a website). Considering that each robot action can be independently autonomous, *i.e.* that $\mathcal{A} = \mathcal{A}_A^{\mathcal{A}_R} \times \mathcal{A}_D$ is set as a perspective.

References

[1] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[3] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," in *ECML*, vol. 6. Springer, 2006, pp. 282–293.

[4] W. D. Nothwang, M. J. McCourt, R. M. Robinson, S. A. Burden, and J. W. Curtis, "The human should be part of the control loop?" in *Resilience Week (RWS), 2016*. IEEE, 2016, pp. 214–220.

[5] J. C. F. de Winter and D. Dodou, "Why the fitts list has persisted throughout the history of function allocation," *Cognition, Technology & Work*, vol. 16, no. 1, pp. 1–11, Feb 2014. [Online]. Available: http://dx.doi.org/10.1007/s10111-011-0188-1

[6] D. A. Wiegmann and S. A. Shappell, "A human error approach to aviation accident analysis: The human factors analysis and classification system," 2003.

[7] G. Li, S. P. Baker, J. G. Grabowski, and G. W. Rebok, "Factors associated with pilot error in aviation crashes," *Aviation, space, and environmental medicine*, vol. 72, no. 1, pp. 52–58, 2001.

[8] K. W. Williams, "A summary of unmanned aircraft accident/incident data: Human factors implications," *U.S. Department of Transportation, Federal Aviation Administration, Civil Aerospace Medical Institute*.

[9] N. Regis, F. Dehais, E. Rachelson, C. Thooris, S. Pizziol, and C. T. Mickaël Causse, "Formal detection of attentional tunneling in human operator-automation interactions," in *IEEE Trans. Human-Machine Systems*, ser. THMS '14, vol. 44, 2014, pp. 326–336.

[10] G. Durantin, F. Dehais, and A. Delorme, "Characterization of mind wandering using fnirs," *Frontiers in systems neuroscience*, vol. 9, 2015.

[11] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[12] R. Bellman, "The theory of dynamic programming," *Bull. Amer. Math. Soc.*, vol. 60, no. 6, pp. 503–515, 11 1954. [Online]. Available: http://projecteuclid.org/euclid.bams/1183519147

[13] R. D. Smallwood and E. J. Sondik, *The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon*. INFORMS, 1973, vol. 21.

[14] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan, "Modular open robots simulation engine: Morse," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 46–51.

[15] H. Bruyninckx, "Open robot control software: the orocos project," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 3. IEEE, 2001, pp. 2523–2528.

[16] S. C. W. Ong, S. W. Png, D. Hsu, and W. S. Lee, "Planning under uncertainty for robotic tasks with mixed observability," *Int. J. Rob. Res.*, vol. 29, no. 8, pp. 1053–1068, July 2010.

[17] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," in *Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine*, Aug. 1985, pp. 329–334. [Online]. Available: http://www.amazon.com/Bayesian-networks-self-activated-evidential-University/dp/B00071DFAE

[18] V. Jääskinen, V. Parkkinen, L. Cheng, and J. Corander, "Bayesian clustering of dna sequences using markov chains and a stochastic partition model," *Statistical applications in genetics and molecular biology*, vol. 13, no. 1, pp. 105–121, 2014.

[19] R. N. Roy, S. Charbonnier, and A. Campagne, "Probing ecg-based mental state monitoring on short time segments," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 6611–6614.

[20] N. Drougard, "Exploiting Imprecise Information Sources in Sequential Decision Making Problems under Uncertainty," Theses, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Dec. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/tel-01296749

[21] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[22] S. M. Siddiqi, G. J. Gordon, and A. W. Moore, "Fast state discovery for hmm model selection and learning," in *Artificial Intelligence and Statistics*, 2007, pp. 492–499.

[23] D. Hsu, S. M. Kakade, and T. Zhang, "A spectral algorithm for learning hidden markov models," *Journal of Computer and System Sciences*, vol. 78, no. 5, pp. 1460–1480, 2012.

[24] P. E. U. d. Souza, C. P. C. Chanel, and F. Dehais, "MOMDP-based target search mission taking into account the human operator's cognitive state," in *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, ser. ICTAI '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 729–736. [Online]. Available: http://dx.doi.org/10.1109/ICTAI.2015.109

[25] S. Nikolaidis, R. Ramakrishnan, K. Gu, and J. Shah, "Efficient model learning from joint-action demonstrations for human-robot collaborative tasks," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 189–196. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696455

[26] T. Gateau, C. P. C. Chanel, M.-H. Le, and F. Dehais, "Considering human's non-deterministic behavior and his availability state when designing a collaborative human-robots system," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 4391–4397.

[27] T. Jaakkola, S. P. Singh, and M. I. Jordan, "Reinforcement learning algorithm for partially observable markov decision problems," in *Advances in neural information processing systems*, 1995, pp. 345–352.

[28] S. Ross, J. Pineau, B. Chaib-draa, and P. Kreitmann, "A bayesian approach for learning and planning in partially observable markov decision processes," *Journal of Machine Learning Research*, vol. 12, no. May, pp. 1729–1770, 2011.

[29] K. Azizzadenesheli, A. Lazaric, and A. Anandkumar, "Reinforcement learning of POMDPs using spectral methods," in *Conference on Learning Theory*, 2016, pp. 193–256.