HHAI2022: Augmenting Human Intellect
S. Schlobach et al. (Eds.)

doi:10.3233/FAIA220188

POMDP-Based Adaptive Interaction Through Physiological Computing

Gaganpreet SINGH, [a] Raphaëlle N. ROY [b] and Caroline P. C. CHANEL [b]
[a] *GREYC-CNRS, Caen, France*
[b] *ISAE-SUPAERO, Université de Toulouse, Toulouse, France*

Abstract. In this study, a formal framework aiming to drive the interaction between a human operator and a team of unmanned aerial vehicles (UAVs) is experimentally tested. The goal is to enhance human performance by controlling the interaction between agents based on an online monitoring of the operator's mental workload (MW) and performance. The proposed solution uses MW estimation via a classifier applied to cardiac features. The classifier output is introduced as a human MW state observation in a Partially Observable Markov Decision Process (POMDP) which models the human-system interaction dynamics, and aims to control the interaction to optimize the human agent's performance. Based on the current belief state about the operator's MW and performance, along with the mission phase, the POMDP policy solution controls which task should be suggested -or not- to the operator, assuming the UAVs are capable of supporting the human agent. The framework was evaluated using an experiment in which 13 participants performed 2 search and rescue missions (with/without adaptation) with varying workload levels. In accordance with the literature, when the adaptive approach was used, the participants felt significantly less MW, physical and temporal demands, frustration, and effort, and their flying score was also significantly improved. These findings demonstrate how such a POMDP-based adaptive interaction control can improve performance while reducing operator workload.

Keywords. physiological computing, adaptive interaction, workload and performance estimation, ECG, POMDP.

1. Introduction

Human-Robot Interaction (HRI) is a branch of research devoted to comprehending, designing, and evaluating robotic technologies for use by or with humans [1]. Teleoperation is a kind of HRI [2] that can be described as the operation of a system from a distance without the robot and human being physically or temporally co-located. Usually teleoperated robots perform activities under human supervision to undertake jobs in dangerous or inaccessible settings (*e.g.* space, oceans, or radioactive sites), as well as carrying out operations such as search and rescue missions [3,4,5]. These examples imply robot decision autonomy (*e.g.* navigates, take pictures, identifies targets), which is a general implicit concept of autonomous robots.

According to [1], the simplest method to examine autonomy is to define the level of interaction between humans and robots and the extent to which each one is independent of the other. This idea can be illustrated by a continuous scale of autonomy ranging from direct control, where the system is handled by the human operator, until mixed-initiative

interaction, where it goes towards more autonomous systems collaborating with humans on a peer-to-peer basis. *Adaptive autonomy* refers to autonomous systems that exhibit some degree of flexibility in their behavior [6,7,4,3] assuming that the division of tasks between humans and artificial agents is not fixed, but rather variable [8]. Interestingly, Mixed-Initiative Interaction (MII) systems feature dynamic autonomy and can be seen as a kind of *adaptive autonomy* interaction case.

MII is a flexible interaction framework concept that allows each agent to take charge of the tasks in which they are specialized [9], however such roles are not necessarily immutable. The problem is why, when, and how artificial agents could seize the initiative to perform tasks that, a priori, are expected to be performed by a human agent. Interestingly, [10] suggested to consider the notion of agents' current capabilities in MII frameworks: a control system would delegate a task or allow the *best current agent* to seize control when necessary. It implicitly implies the use of agent monitoring techniques to detect potential drops in agents' performance, cognitive unavailability, or even incapacity.

Human performance and human error are the results of information processing and are subject to variations depending on human affective and cognitive states. Workload, a wide concept that mostly relates to the level of cognitive resource engagement from the operator, directly impacts performance in critical settings as in Manned-UnManned-Teaming (MUM-T) missions for instance [5,11]. This mental state is often analyzed offline (e.g., after mission realization) using questionnaires and/or behavioral measures. However, it can also be estimated in an online fashion using machine learning methods applied to various behavioral and physiological features [12]. This processing pipeline is known as *physiological computing* [13]. Recent literature has proposed online methods [14,15,16] and tools [17] to compute temporal cardiac features on short time windows, such as the Heart Rate (HR) and the Heart Rate Variability (HRV; e.g. rMSSD). These metrics are known to be impacted by workload: an increase in mental workload has been linked to an increase in HR and a decrease in HRV [18]. The use of such tools to estimate the human operator's mental workload during interaction seems to be a promising venue.

In this context, the present work paves the way towards such a Mixed-Initiative Human-Robot Interaction (MI-HRI) control system by proposing a formal framework to efficiently drive the interaction between a human operator (e.g., pilot) and a team of Unmanned Aerial Vehicles (UAVs). The objective of the interaction control policy is to improve human performance in a MUM-T context. The policy chooses to present (or not) tasks to the operator based on an online estimation of her mental workload and current performance. Due to the non-deterministic nature of this dynamic system, the evident partial observability of the human mental workload, and the interest in enhancing the human operator's performance, this work is built upon the Partially Observable Markov Decision Process (POMDP), an elegant framework to model and optimize sequential decision-making under uncertainty and partial observability.

Related work. A similar POMDP-based interaction control approach was proposed for a robotic teleoperation scenario [19], in which POMDP model parameters were approximated based on experimental data. However, their model was only evaluated in simulation. In [20] an elegant POMDP-based technique was suggested for estimating different types of human operators (safe or efficient) based on the sequence of actions experimentally observed. The obtained POMDP policy adapted the robot behavior (e.g. speed) in function of the operator type estimation. In [21] a POMDP-model that anticipates a human's state of mind during a human-robot collaboration task was proposed. The human-

related estimation is based on human activity recognition, without physiological inputs, which gives very relevant features in some settings, but less so in remote operation scenarios where the operator is not co-located with the robot. [4] proposed an integrated POMDP-based system applied in a multi-UAV supervision scenario, where the operator was confronted with a dual-task paradigm. The system considered a sort of human state monitoring that accounted for the non-deterministic human behavior: the response time to UAVs' requests and the *availability* of the human operator (measured with an eye-tracker). The human operator's performance on a secondary task was improved while preserving the overall system performance in the primary task. Note in almost all these previous works, the POMDP interaction model was hand-tuned (except in [19]-[20]).

Hypothesis. Based on these previous studies, we hypothesize that adapting the interaction using an approach that exploits both the human agent's mental workload estimation and current performance will increase the long-term performance of the human operator while lowering their mental workload as measured through subjective, behavioral, and physiological metrics.

Contributions. The contribution of the present work is three-fold: (i) the human state estimation is not only based on behavioral features but also on physiological features. In detail, this estimation is introduced as a human (mental) state observation into the POMDP model, and its confidence is modelled as the observation function, enabling to tackle the uncertainty inherent to such a mental state monitoring; (ii) interaction data acquired in a previous study was exploited to learn the MUM-T interaction dynamics defining the POMDP state transition function. In this way, probabilities concerning mission events and human performance are not hand-tuned; (iii) the proposed system was evaluated following an experimental protocol that included 13 participants. In accordance with the literature, the achieved results show the proposed approach to be a promising venue for MUM-T applications. As far as the authors know, no previous work (even without mental workload estimation) has ever proposed a similar long-term reasoning adaptive strategy in the same human-UAVs interaction settings.

2. MUM-T scenario and previous work achievements

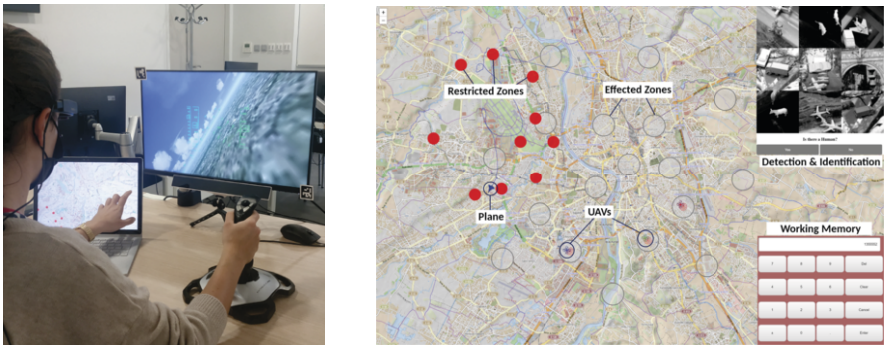
A MUM-T scenario was implemented, and a specific study [22] was run to characterize and estimate mental workload in this scenario based on physiological signals. We describe this scenario and the former results that constitute the basis of the present work.

2.1. MUM-T search and rescue mission

The MUM-T scenario considered is a search and rescue mission where pilots have to interact with the UAVs through the U-track application while performing the flight using the Aerofly FS2 flight simulator (see Figures 1a and 1b) providing a virtual environment with wreckage areas with destroyed 3D buildings to illustrate a disaster area.

The main search and rescue mission was composed by 3 sub-tasks:

- Detection and identification sub-task: It simulates the process of UAV requests in the form of pop-ups in the U-track application (Fig. 1b). A beep sound alerted the request arrival, and the pilot had to search in a group of 9 gray scale images if there was any human present, and had to answer with Yes or No.



(a) A participant. (b) The U-track application.

Figure 1. (a) A participant performing the flying task while interacting with the UAVs via the tactile tablet. (b) The U-track application (custom interface) launched in a tactile tablet. Top right corner: detection and identification sub-task using images from the Norb database [23]. Bottom right corner: working memory sub-task. Red zones: restricted zones to avoid. The pilot can visualize their position and UAVs' in real-time.

- Pilot flying sub-task: Pilots had to remember the ATC heading instructions and maintain that heading in the Aerofly simulator using the joystick. Restricted zones had to be avoided (see Fig. 1b).

- Working memory sub-task: It mimics Air Traffic Control (ATC) instructions played in the form of audio messages of headings to be followed. The pilot was also required to memorize UAV communication channels and recall them in the U-track application using a numpad (Fig. 1b).

Participants were encouraged to put the highest priority on the detection and identification sub-task, as well as to avoid flying over restricted areas. Keeping the right heading announced in ATC commands is followed by decreasing priority. The last priority was given to the working memory sub-task (i.e. recall of UAV communication channels).

2.2. Human agent workload estimation

In a former study [22], 14 participants played the MUM-T scenario described above to determine relevant physiological features for mental workload assessment. For that, workload level (low: L, and high: H) could be varied between experimental blocks in a pseudo-random manner through tasks' difficulty modulation: (i) the number of communication channel inputs (i.e., 1 UAV in L *vs* 2 UAVs in H); and, (ii) the restricted zones placement (i.e., far from the plane path in L *vs* close to the path in H). Each participant performed 4 missions: 2 missions with a L workload level and 2 with a H level. These levels were evaluated using a self-reported questionnaire, behavioral and physiological measures: cerebral activity measurements with an electroencephalogram (EEG), cardiac activity with a electrocardiogram (ECG), and oculomotor activity with an eye-tracker (ET). Statistical analysis was conducted to determine the impact of workload levels on features computed using a short time window (i.e. 6-second). To achieve human agent mental workload estimation, several feature combinations and classifiers were evaluated.

The main results of this former study [22] were: i) a significant impact of mental workload on all measures; and, ii) a higher intra-subject classification accuracy (75%) reached using ECG features alone, or in combination with EEG and ET ones when a classical training-test pipeline was used (i.e., training and testing sets created by randomly picking data samples among the 4 missions). However, when a more realistic procedure was employed (i.e. training set at the beginning of the session, testing set at the end,

mimicking a calibration for a new user), the classification accuracy generally dropped to chance level. Indeed, when this realistic procedure was considered, inter-subject classification results based only on ECG temporal features gave the best average accuracy results (59.8%).

These previous achievements are the basis of the present work. More specifically, the confusion matrix of the inter-subject classifier using ECG features is used as part of the observation function of the POMDP model that aims to control the MI-HRI system. Moreover, pilot behavioral data and expert knowledge (i.e., pilot performance scores, mission-related events, mental workload levels) and mission phases were also exploited to approximate the POMDP transition function for this MI-HRI framework.

3. Adaptive Interaction Control System

3.1. Reminder on the POMDP formalism

A POMDP is a tuple $\langle S, A, \Omega, T, \mathcal{O}, R, \gamma, b_0 \rangle$, where, S is a set of states; A is a set of actions; Ω is a set of observations; $T : S \times A \times S \to [0;1]$ is a transition function denoting the probability $T(s',a,s) = p(s'|a,s)$ of reaching state $s' \in S$ given the action $a \in A$ is performed in state $s \in S$. $\mathcal{O} : \Omega \times S \to [0;1]$ is an observation function such that $\mathcal{O}(o,s) = p(o|s)$, denoting the probability of observing $o \in \Omega$ given state $s \in S$; $R : S \times A \to \mathbb{R}$ is a reward function associated with a state-action pair; γ is the discount factor; and, b_0 is the initial probability distribution over states, we have $b_0(s) = p(s_0 = s)$. At each decision time step t, the agent takes an action $a \in A$, receives an observation $o \in \Omega$ and updates its *belief state* $b \in \Delta$ (Δ denotes the *belief state space*) using the Bayes' rule:

$$b_a^o(s') = \varphi^{-1} p(o|a,b) = \varphi^{-1} p(o|s') \sum_{s \in S} p(s'|s,a) b(s) \tag{1}$$

with, $\varphi = \sum_{s'} p(o|s') \sum_s p(s'|s,a) b(s)$. The belief state can be see as a complete information history, as it concatenates all the action-observation sequences such as $b_t(s) = p(s_t = s | o_t, a_{t-1}, o_{t-1}, \ldots, a_0)$.

In order to act efficiently, the POMDP solving step consists in finding a policy $\pi^* : \Delta \to A$, which maximizes a performance criterion, the Value Function, generally defined as the expected discounted sum of rewards:

$$V^{\pi^*}(b) = \max_{\pi} \mathbb{E}^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r(b_t, \pi(b_t)) \middle| b_0 = b \right] \tag{2}$$

Interestingly, V^{π^*} satisfies the Bellman equation ($V^{\pi^*} = V^*$): $V^*(b) = \max_{a \in A} [r(b,a) + \gamma \sum_{o \in \Omega} p(o|a,b) V^*(b_a^o)]$, where $r(b,a) = \sum_{s \in S} R(s,a) b(s)$. The optimal policy π^* can be extracted from the Bellman equation when it converges for all belief states. In practice, the exact optimal solution of an infinite horizon POMDP is hard to be achieved [24]. However, recent POMDPs solving algorithms, as HSVI2 [25] or SARSOP [26] are able to compute policies by approximating a ε-optimal value function in reasonable time.

3.2. POMDP model for adaptive interaction control and parameters learning

3.2.1. States, observations and actions.

The POMDP state set is composed by the Cartesian product between a set of state variables ($S : S_{M_s} \times S_{H_p} \times S_{H_{wL}}$): the mission-related state (S_{M_s}), the performance of the hu-

man pilot (S_{H_p}), and the workload level of the human pilot ($S_{H_{wL}}$). In detail, we have S_{M_s}: {*f, fpr, frar, fpr_rar*}, symbolizing an only piloting situation, piloting and pop-up request arrival, piloting and UAV radio channel setting request, and piloting along with pop-up and UAV radio channel setting requests, respectively; S_{H_p}: {*low_perf, high_perf*}, expressing a low (or high) human pilot performance at the current mission phase; and, $S_{H_{wL}}$: {*low_workload, high_workload*}, denoting a low (or high) mental workload level.

The POMDP observation set is composed by the Cartesian product between a set of observation variables ($\Omega : O_{M_s} \times O_{H_p} \times O_{H_{wL}}$): the mission-related observation (O_{M_s}), the observation of the performance of the human pilot (O_{H_p}), and the observation of the workload experienced by the human pilot ($O_{H_{wL}}$). Specifically, we have O_{M_s}: {*of, ofpr, ofrar, ofpr_rar*} symbolizing the observation of same mission-related state values as for the S_{M_s} state variable; O_{H_p}: {*olow_perf, ohigh_perf*}, expressing the observation of low (or high) human performance; and, $O_{H_{wL}}$: {*obs_low_workload, obs_high_workload*}, denoting the observation of low (or high) mental workload.

The POMDP action set A is defined as $A = \{ns\_pr, ns\_rar, ns\_pr\_rar, d_n\}$, the which relate to not show a pop-up request to the pilot, to not send a radio channel setting request, to not show pop-up and not send radio channel setting request, and a do nothing action being equivalent to a *keep monitoring* action, respectively.

3.2.2. State transition function.

The transition function is based on some conditional independence assumptions and on expert knowledge. The transition function is defined as:

$$\begin{aligned} T(s', a, s) &= p(S'_{M_s}, S'_{H_p}, S'_{H_{wL}} | a, S_{M_s}, S_{H_p}, S_{H_{wL}}) \\ &= p(S'_{H_p} | S_{H_p}, S_{H_{wL}}, S_{M_s}) p(S'_{H_{wL}} | S_{H_{wL}}) p(S'_{M_s} | a, S_{M_s}) \end{aligned}$$

To learn the state transition function, all the data collected in the previous study (see Sec. 2) were used. However, before detailing the learning method, the pre-processing step operated on the available dataset is hereafter presented.

Dataset preprocessing. First, data were separated following workload conditions of the previous experiment. And, for each mission and workload condition, data were epoched into 6-second windows, regrouping all available measurements. Then, a performance score was attributed to each 6-second window (denoted w_s^i, where i represents the i-th window). This score was defined in function of each sub-task priority and in function of the score achieved by the participant in each sub-task given the workload condition: how well the ATC heading command is followed while avoiding restricted zones, or if the pilot had identified a person was present in images, or how well radio channel was recalled. The performance score is formalized as a weighted score as:

$$w_s^i = \omega_1 \mathbf{f}_{score} + \omega_2 \mathbf{f}_{penalty} + \omega_3 \mathbf{pop}_{score} + \omega_4 \mathbf{ra}_{score} + \omega_5 \mathbf{c}_{penalty} \tag{3}$$

where, $\sum_{j=1}^{5} \omega_j = 1$. The $\mathbf{f}_{score}$ is the score obtained in the flying sub-task as: $\mathbf{f}_{score} = \sum_{n=0}^{k} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x_n-\mu)^2}{2\sigma^2}}$, with x_n one sample of the aircraft heading observed every 500ms in this i-th 6-second time window (it means we dispose $k = 12$ samples per time window), μ the actual heading command for this i-th window, and σ the standard deviation set to 20 degrees[1]; $\mathbf{f}_{penalty}$ is a penalty defined as the total number of times the pilot enters in a

[1] when x_n was out of the range of $\mu \pm \sigma$, the heading sample value x_n was set to zero.

restricted zone during this i-th time window, measured every 50ms, and multiplied by a coefficient, such as: $\mathbf{f}_{penalty} = \beta \sum_{n=0}^{m} \mathbb{I}_{restricted}$, with $m = 120$, $\beta = 0.5$, and $\mathbb{I}_{restricted}$ being the indicative function; $\mathbf{pop}_{score}$ is the score obtained for a right (or wrong) identification; $\mathbf{ra}_{score}$ is the score for a right (or wrong) UAV radio channel recall; and finally $\mathbf{c}_{penalty}$ is the penalty associated to the workload condition (high workload level being bigger). Each sub-task score was normalized between $[-1, 1]$.

Once these scores were computed for each time window, a median score was calculated for each mission-related state and workload condition combination. These median scores were then used to separate the data between *low_perf* and *high_perf* balanced sets. We assumed then, the 6-second window as the POMDP decision time step.

Transition function learning. The $p(S'_{H_p}|S_{H_p}, S_{H_{wL}}, S_{M_s})$ probability function was approximated based on those preprocessed data partitioned in function of the workload level, mission-related state and human pilot performance. Note, S_{H_p} is assumed to be independent from the action, because during the previous study all requests were triggered to the human pilot as soon as they appeared. Specifically, the frequency of change of the S_{H_p} state variable was calculated in function of the previous state value ($S_{H_p}, S_{H_{wL}}, S_{M_s}$).

For the human workload level state variable $S_{H_{wL}}$, the function $p(S'_{H_{wL}}|S_{H_{wL}})$ was defined based on a different a priori knowledge. It is assumed that a low (resp. high) workload level remains at least for 5 minutes following the experimental protocol defined for this work (see Sec. 4.2). So that, as a POMDP decision (i.e. action) is triggered each 6s, there is a chance of $1/50$ that the workload level will change, and $49/50$ that it remains the same. These probabilities are independent of the action.

The mission-related state variable S_{M_s} is assumed to be independent of the others state variables (i.e. mission events do not depend on the workload level or the performance). Thus, the $p(S'_{M_s}|a, S_{M_s})$ probability function was approximated by computing the frequency of appearance of each event using all available data (i.e. all 6-second time windows). For example, the probability that a pop-up request appears with respect to the mission-related state was approximated by the frequency that this event appeared given the previous mission state.

We highlight that this data-driven process indicate the transition function for the d_n action (e.g. *do nothing*), such as $p(S'_{M_s}|a = d_n, S_{M_s})$, because in the previous study all events where triggered to participants as soon as their appeared. To design the transition function adapted to each action that *does not show a particular request*, a specific normalization process was introduced. This process builds upon the fact that is less likely to reach a mission-related state where the pop-up event is present if the action *do not show pop-up request* is played. More specifically, to define the transition function for the *do not show pop-up request* action (*ns_pr*), such as $p(S'_{M_s}|a = ns\_pr, S_{M_s})$, we used the $p(S'_{M_s}|a = d_n, S_{M_s})$ probability function as initial support, then, the probability of the mission-related states where a pop-up event is present (*fpr* and *fpr_ra*) was set to δ, as:

$$\hat{p}(S'_{M_s}|a = ns\_pr, S_{M_s}) = \begin{cases} \delta, \text{ if } S'_{M_s} = fpr \text{ or } S'_{M_s} = fpr\_ra \\ p(S'_{M_s}|a = d_n, S_{M_s}) - \delta N_s, \text{ otherwise} \end{cases} \quad (4)$$

where, $\delta \in [0, 0.01]$ and N_s is the number of mission-related states where the considered event is presented (pop-up in this example). Then, the probability function was normalized to sum-up to one: $p(S'_{M_s}|a = ns\_pr, S_{M_s}) = \Phi^{-1} \hat{p}(S'_{M_s}|a = ns\_pr, S_{M_s})$ with $\Phi = \sum_{S'_{M_s}} \hat{p}(S'_{M_s}|a = ns\_pr, S_{M_s})$. An equivalent process was applied for the other actions considering the others missions-related events.

Table 1. Table reporting $p(O_{H_{wL}}|S_{H_{wL}})$. The confusion matrix of the classifier is used as is.

		$O_{H_{wL}}$	
		obs_low_workload	*obs_high_workload*
$S_{H_{wL}}$	*low_workload*	0.636	0.364
	high_workload	0.342	0.658

Observation function. The observation function is defined as: $\mathcal{O}(o,s) = p(O_{H_{wL}}|S_{H_{wL}})$ $p(O_{M_s}|S_{M_s})p(O_{H_p}|S_{H_p})$. For the functions $p(O_{M_s}|S_{M_s})$ and $p(O_{H_p}|S_{H_p})$ we assume that S_{M_s} and S_{H_p} are fully observable state variables, because the system knows which events are presented in a given decision time step, and, the human performance score given the mission phase can be computed online without any ambiguity ($O_{M_s} = S_{M_s}$, $O_{H_p} = S_{H_p}$).

The function $p(O_{H_{wL}}|S_{H_{wL}})$ was defined thanks to the previous study [22]. To be compliant with the fact that the participants of the present study may differ from the participants of the previous study, the inter-subject pipeline (see Sec. 2.2) was chosen. The classifier that reached the best inter-subject accuracy based on ECG features - Heart-Rate (HR) and temporal Heart-Rate Variability (HRV) - was used, and its confusion matrix defined the $p(O_{H_{wL}}|S_{H_{wL}})$ function, as shown in Table 1.

Reward function. The reward function $R(s,a)$ is defined as the average among all 6s windows performance scores (Eq. 3) belonging to a given state ($S_{M_s}, S_{H_p}, S_{H_{wL}}$). Additionally, based on expert knowledge we have decided to slightly modify this reward with an additional penalty in the specific case were the human-pilot faces more than one subtask, presenting a low performance and high workload level. The main idea here was to incite the POMDP policy to keep participants away from high workload and low performing states. We can note that this performance is independent of the action. Yet it does not prevent the system to find a relevant policy - the POMDP policy based on the current belief state will induce the MI-HRI system to future states that favor gains.

Offline policy generation. The POMDP model was solved using the SARSOP algorithm [26] ($\gamma = 0.95$ and $\varepsilon = 0.01$). The policy was executed during the experiment.

4. Materials and Methods

4.1. Participants

Thirteen healthy volunteers (11 males; 25.8 y.o. $\pm$ 3.5) gave their written consent for anonymous data collection storage and processing. No previous piloting experience from participants was required. The experiment was authorized by the local ethical committee (CER Univ. Toulouse ID 2019-137).

4.2. Experimental Protocol

The experimental protocol was designed to evaluate the POMDP-based interaction control policy with regards to human performance and mental workload management when compared with a non adaptive interaction system. To that end, all developments were integrated and played online: extraction and processing of physiological features, monitoring of human mental state, and adaptation of the interaction. Two mission runs were counter-balanced between participants, one where the POMDP-based adaptive policy was played, and one where no adaptive action was taken. In the non adaptive condition, all UAV and system requests were triggered to the human pilot as soon as they appeared.

Regarding the experimental protocol (Fig. 2), first the participants were invited to give their informed consent. Then, they performed a training session to familiarize them-

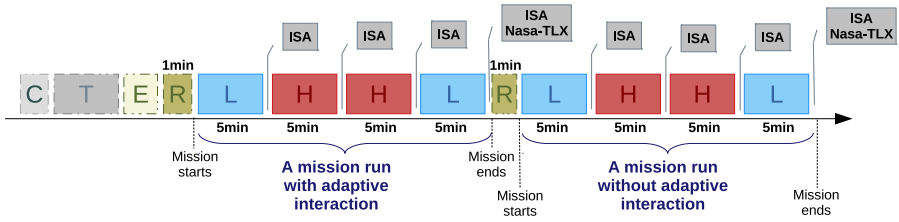


Figure 2. Experimental protocol timeline. C: consent form signature, T: task training, E: ECG set up, R: resting state (ECG baseline), L: low workload block, H: high workload block. Mission runs, with or without adaptive interaction policy were counter-balanced between participants; workload blocks were counter-balanced within mission runs with either the L-H-H-L or the H-L-L-H pattern. Note an experiment took around 45 minutes.

selves with the task - incl. the management of ATC commands - and were equipped with the ECG device. A mission block started after a resting block (1 min), which allowed us to record the ECG baseline to compute normalized HR and HRV features. During a mission run, workload was manipulated per blocks of 5 minutes: the variation followed L-H-H-L or H-L-L-H patterns, and was pseudo-randomized between participants.

4.3. Data Acquisition, Analysis and Policy Execution

Subjective feedback. At the end of each workload block, participants were asked to answer the Instantaneous Self Assessment (ISA) of workload [27], and at the end of each mission the NASA-TLX [28] was given to evaluate the quality of interaction condition. Participants were also asked how engaged they had felt during the previous mission using a home-made scale and the question: *"How engaged and focused were you?"*

Behavioral features. The scoring for pilot performance used here was computed following the weighted score defined in Eq. 3 as explained in Sec. 3.2. This score was then translated as a POMDP observation variable O_{H_p} (*low_perf* or *high_perf*) following the same median scores values obtained for database preprocessing when the state transition function was approximated.

Physiological features. Physiological data were acquired and recorded thanks to Lab Streaming Layer (LSL) streams[2]. Streams were created and published in a local network, to perform an online processing of the features, and also to record them with LabRecorder[3]. The ECG acquisition device (Faros eMotion 360°) can be configured to provide online ECG raw data and RR intervals (based on R-peak detection), which were broadcasted using a Bluetooth protocol. The Faros Streamer application[4] allows streaming these data through LSL. Additionally, a script based on a library proposed by Aura Healthcare[5] was used to compute online ECG temporal features, as Heart-rate (HR) and Heart-rate Variability (HRV), from RR intervals. At each 6-second non-overlapping window, HR and HRV features were computed and normalized with respect to the 1-minute resting block. They were then provided as feature vector to the classifier for mental workload estimation. The chosen classifier following the former work was a Linear Discriminant Analysis (LDA) with shrinkage of the covariance matrix. The output of this classifier is the value of the observation variable $O_{H_{wL}}$ used in the model.

[2]https://labstreaminglayer.readthedocs.io/index.html
[3]https://github.com/labstreaminglayer/App-LabRecorder
[4]https://github.com/bwrc/faros-streamer-2
[5]https://github.com/Aura-healthcare/hrv-analysis

POMDP policy execution. The POMDP policy solution was executed in parallel with other applications using a dedicated script. Starting from the initial belief state, the action to be performed is computed following the policy and launched. Then, 6 seconds after the action execution, the observation $o = (O_{M_s}, O_{H_p}, O_{H_{wL}})$ is received which is used to update the belief state with the help of Bayes' rule (Eq. 1). It works as an online monitoring of the human (mental) state. Note that during the experiment the POMDP policy does not observe the workload being manipulated, it receives the label $O_{H_{wL}}$ predicted as output of the classifier, and knows the frequency of appearance (each 5min) defined through the transition function. Then, based on the current belief state, the POMDP action is chosen following the policy again, and after 6 seconds a new observation is received, and so on.

4.4. Statistical Analysis

In order to assess the impact of the experimental conditions, statistical analyses were performed on all data using the Statistica software. Subjective measures acquired through the NASA-TLX questionnaire were analyzed for each scale using paired t-tests when the data were normally distributed, and using a Wilcoxon test when they were not. All other measures as ISA questionnaire, mean performance and mean cardiac features HR and HRV were analyzed using two-way repeated measure ANOVAs (i.e 2 load conditions $\times$ 2 types of interaction: adaptive *vs* non-adaptive) after homoscedasticity and normality assumptions verification. Tukey *post-hoc* tests were performed for each statistically significant main effect and interaction effect.

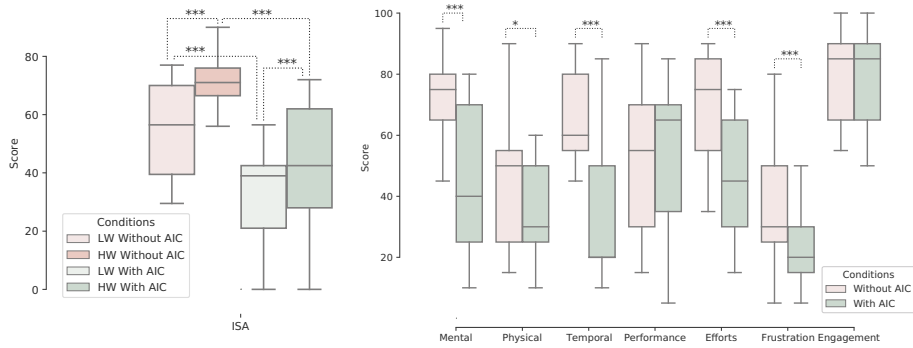
5. Experiment Results

5.1. Adaptive interaction impact on subjective feedback.

Participants' workload as reported through the ISA questionnaire was significantly lower in the low workload condition than in the high one, and lower in the adaptive condition compared to the non-adaptive condition (respectively $F(1,12) = 30.96$, p-value < 0.01, $\eta^2 = 0.72$, and $F(1,12) = 79.14$, p-value < 0.01, $\eta^2 = 0.87$). Moreover, the interaction between workload and adaptive condition was significant, with a stronger increase in reported workload with workload in the non-adaptive condition than in the adaptive one ($F(1,12) = 8.21$, p-value < 0.05, $\eta^2 = 0.41$). These results are illustrated in Fig. 3a. Regarding participants' workload scales from the NASA-TLX questionnaire, their reported mental demand, physical demand, temporal demand, effort and frustration significantly decreased with the adaptive interaction policy (respectively $t = 5.17$ with p-value < 0.001, $t = 2.24$ with p-value < 0.05, $W = 2.00$ with p-value < 0.001, $t = 4.56$ with p-value < 0.001, $t = 2.99$ with p-value < 0.05). No significant effect was found for reported performance and engagement (Fig. 3b).

5.2. Adaptive interaction impact on performance.

Participants' mean performance only revealed a significant decrease with workload ($F(1,12) = 56.84$, p-value < 0.01, $\eta^2 = 0.52$): participants had a better score in the low workload condition than in the high one (Fig. 4a). Regarding participant's mean flying score only, it was significantly lower in the high workload condition than in the low workload condition ($F(1,12) = 56.84$, p-value < 0.05, $\eta^2 = 0.32$), as well as significantly higher in the adaptive than in the non-adaptive condition ($F(1,12) = 136.55$, p-value < 0.01, $\eta^2 = 0.53$).



(a) Subjective workload via ISA. (b) Subjective workload via NASA-TLX.

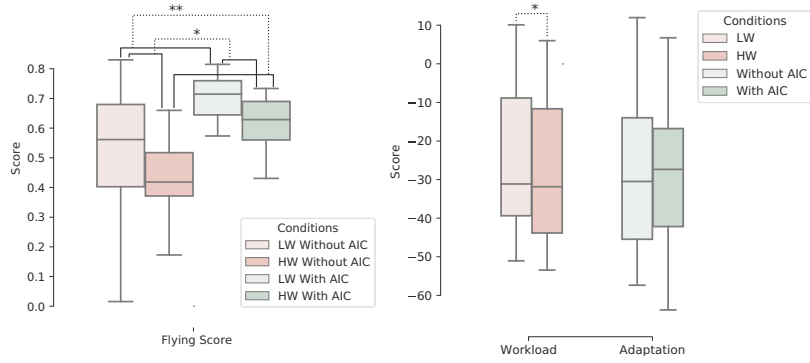
Figure 3. Subjective results (***: p-value < 0.001, **: p-value < 0.01, and *: p-value < 0.05). LW/HW: low (resp. high) workload. AIC: adaptive interaction control.

5.3. Adaptive interaction impact on physiological features.

Participants' normalized HRV significantly decreased with an increase in workload ($F(1,12) = 904.92$, p-value < 0.05, $\eta^2 = 0.43$) - see Fig. 4b. No significant effect of adaptive condition or interaction with workload was found. Moreover, even though participants' normalized HR was higher in the non-adaptive than adaptive conditions, and in the high workload than in the low workload conditions, these effects were not significant.

6. Discussion

This study strongly differs form earlier 2000's studies [8,29] that proposed reactive and threshold-based adaptations that did not consider a long-term performance reasoning strategy, as our POMDP-based approach does. Based on the belief state, the POMDP policy decides which task should be triggered to the human operator in order to maximize her long-term performance. In this sense, the belief state can be see as an online monitoring of the human (mental) state: it estimates the couple workload-performance. This study also improves similar previous approaches [21,4,20,19,8] that have proposed adaptive interaction frameworks based POMDPs and human (mental) state estimation. Interaction data from a former experiment [22] were used to learn the POMDP model dynamics, rather than manually tuning it such as previously done by [4], or approached



(a) Flying Score. (b) Normalized Heart Rate Variability.

Figure 4. Behavioral and physiological results (**: p-value < 0.01, and *: p-value < 0.05). LW/HW: low (resp. high) workload. AIC: adaptive interaction control.

by simulation as in [21]. Also, the experimental campaign presented in this study allowed to obtain participant's feedback to evaluate the approach, unlike [19].

The subjective and behavioral results have confirmed the ability of the implemented adaptive interaction behavior to mitigate mental demand, physical demand, temporal demand, frustration and effort. Indeed, in accordance with the literature [8,30], the results indicate that participants have experienced significantly less workload in adaptive interaction condition then in non-adaptive condition, as evaluated using the ISA and NASA-TLX questionnaires. Moreover, the flying score revealed that participants performed significantly better in conditions involving adaptive interaction control.

The expected general effect of workload on cardiac features was present, but only significant for the normalized HRV measure that decreased with an increase in workload in accordance with the literature [16,18]. Contrary to our expectations, no significant effect of the adaptive condition was found on cardiac features. This could reflect a high engagement level from the participants in the adaptive condition which helped them achieve a high performance. Following [31], the fact that cardiac features were not modified by the adaptive condition while subjective (decreased reported workload) and behavioural metrics (increased flying score) indicated a lowering of mental workload may clearly reflect that participants maintained their engagement during the task. This is a very positive point for adaptive interaction since no disengagement from participants was observed.

Limitations. Possible limitations of this work are: the low number of participants, their lack on piloting experience, the use of a wearable, although not invasive, ECG device and the few variables considered in the POMDP model. Note, the binary variables used may prevent the model to capture more fine-grained differences in performance and in workload.

7. Conclusion and Future Work

This work proposes a framework to drive the interaction between a human operator and a team UAVs in a MUM-T scenario. The contribution is the design and the evaluation of a long-term interaction strategy which favors human performance. The design was possible thanks to model parameters learned from available interaction data and to an online behavioral and physiological-based human (mental) state monitoring. The results reveal that: (i) the participants felt significantly less load, and an improved quality of interaction when the proposed adaptive approach was used compared with a non-adaptive system; (ii) participant's flying score was also improved. However, in MUM-T scenarios a focus on team performance may be preferred. Hence, further work is needed to include UAVs' states and scores into the POMDP reward function to improve the expected team performance. The classification pipeline should also be improved, for instance thanks through the use of additional sensors, or thanks to transfer learning and an online POMDP definition and solving. Yet this study demonstrates how adaptive interaction can improve performance while reducing operator workload, hence paving the way for better MUM-T.

Funding

This work has been financially supported by the CASAC Chair[6] (2016-2021).

[6] https://www.isae-supaero.fr/fr/isae-supaero/mecenat-relations-avec-la-fondation-isae-supaero/chaire-dassault-aviation-casac/

References

[1] Goodrich MA, Schultz AC. Human-robot interaction: a survey. Now Publishers Inc; 2008.

[2] Sheridan TB. Human–robot interaction: status and challenges. Human factors. 2016;58(4):525-32.

[3] Reich F, Heilemann F, Mund D, Schulte A. Self-scaling human-agent cooperation concept for joint fighter-UCAV operations. In: Advances in Human Factors in Robots and Unmanned Systems. Springer; 2017. p. 225-37.

[4] Gateau T, Chanel CPC, Le MH, Dehais F. Considering human's non-deterministic behavior and his availability state when designing a collaborative human-robots system. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2016. p. 4391-7.

[5] Strenzke R, Uhrmann J, Benzler A, Maiwald F, Rauschert A, Schulte A. Managing cockpit crew excess task load in military manned-unmanned teaming missions by dual-mode cognitive automation approaches. In: AIAA guidance, navigation, and control conference; 2011. p. 6237.

[6] Durand B, Godary-Dejean K, Lapierre L, Crestani D. Inconsistencies Evaluation Mechanisms for an Hybrid Control Architecture with Adaptive Autonomy. In: Control Architectures of Robots; 2009. .

[7] Luck M, d'Inverno M, Munroe S. Autonomy: Variable and generative. In: Agent Autonomy. Springer; 2003. p. 11-28.

[8] Parasuraman R, Cosenzo KA, De Visser E. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. Military Psychology. 2009;21(2):270-97.

[9] Allen JE, Guinn CI, Horvtz E. Mixed-initiative interaction. IEEE Intelligent Systems and their Applications. 1999;14(5):14-23.

[10] Chanel CPC, Roy RN, Drougard N, Dehais F. Mixed-Initiative Human-Automated Agents Teaming: Towards a Flexible Cooperation Framework. In: International Conference on Human-Computer Interaction. Springer; 2020. p. 117-33.

[11] Schulte A, Donath D, Honecker F. Human-system interaction analysis for military pilot activity and mental workload determination. In: IEEE International Conference on Systems, Man, and Cybernetics; 2015. p. 1375-80.

[12] Roy RN, Drougard N, Gateau T, Dehais F, Chanel CP. How Can Physiological Computing Benefit Human-Robot Interaction? Robotics. 2020;9(4):100.

[13] Fairclough SH. Fundamentals of physiological computing. Interacting with computers. 2008;21(1-2):133-45.

[14] Rajavenkatanarayanan A, Nambiappan HR, Kyrarini M, Makedon F. Towards a Real-Time Cognitive Load Assessment System for Industrial Human-Robot Cooperation. In: IEEE International Conference on Robot and Human Interactive Communication (RO-MAN); 2020. p. 698-705.

[15] Chanel CPC, Wilson MD, Scannella S. Online ECG-based features for cognitive load assessment. In: IEEE International Conference on Systems, Man and Cybernetics (SMC); 2019. p. 3710-7.

[16] Roy RN, Charbonnier S, Campagne A. Probing ECG-based mental state monitoring on short time segments. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013. p. 6611-4.

[17] Champseix R. Aura Healthcare project; 2018. [Online; accessed September-2021]. https://aura-healthcare.github.io/hrv-analysis/.

[18] Heard J, Harriott CE, Adams JA. A survey of workload assessment algorithms. IEEE Transactions on Human-Machine Systems. 2018;48(5):434-51.

[19] de Souza PEU, Chanel CPC, Dehais F. MOMDP-Based Target Search Mission Taking into Account the Human Operator's Cognitive State. In: 27th IEEE International Conference on Tools with Artificial Intelligence, (ICTAI); 2015. p. 729-36.

[20] Nikolaidis S, Ramakrishnan R, Gu K, Shah J. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In: 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE; 2015. p. 189-96.

[21] Görür OC, Rosman B, Sivrikaya F, Albayrak S. Social cobots: Anticipatory decision-making for collaborative robots incorporating unexpected human behaviors. In: ACM/IEEE International Conference on Human-Robot Interaction; 2018. p. 398-406.

[22] Singh G, Chanel CPC, Roy RN. Mental Workload Estimation Based on Physiological Features for Pilot-UAV Teaming Applications. Frontiers in Human Neuroscience. 2021 August;15(692878):1-20.

[23] LeCun Y, Huang FJ, Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

vol. 2; 2004. p. II-104.
[24] Papadimitriou CH, Tsitsiklis JN. The complexity of Markov decision processes. Mathematics of operations research. 1987;12(3):441-50.
[25] Smith T, Simmons RG. Point-Based POMDP Algorithms: Improved Analysis and Implementation. In: Proc. UAI; 2005. .
[26] Kurniawati H, Hsu D, Lee WS. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In: Proc. RSS; 2008. .
[27] Tattersall AJ, Foord PS. An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics. 1996;39(5):740-8.
[28] Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Advances in psychology. vol. 52. Elsevier; 1988. p. 139-83.
[29] Parasuraman R. Designing automation for human use: empirical studies and quantitative models. Ergonomics. 2000;43(7):931-51.
[30] Parasuraman R, Mouloua M. Automation and human performance: Theory and applications. Routledge; 2018.
[31] Dehais F, Lafont A, Roy R, Fairclough S. A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. Frontiers in Neuroscience. 2020;14:268. Available from: https://www.frontiersin.org/article/10.3389/fnins.2020.00268.