Estimating an Operator’s Cognitive State in Real Time: a User Modeling Approach

Thomas R. Colin, Tina Mioch, Nanja Smets, and Mark A. Neerincx

Abstract—This paper presents a Cognitive Task Load (CTL) model designed to keep track of an operator’s mental workload, both quantitatively (amount of workload) and qualitatively (cognitive state). By integrating this model in a (semi-)autonomous robot, robot’s level of automation and user interface can be attuned to the operator’s state (Every second, the CTL-model updates a diagnosis of the operator’s cognitive state). The CTL-model’s predictions were tested in an Urban Search And Rescue (USAR) setting. Tasks could be well-monitored automatically, but the test showed insufficient workload variations to validate the model. This indicates that participants should be subjected to more "high-pressure" conditions in future trials. These results also constitute evidence that adaptation to an operator’s mental workload is most effective when the operator’s ability to manage his own workload is restricted by the urgency of the tasks.

I. INTRODUCTION

(Semi-)autonomous robots are becoming increasingly capable due to the combined progress of robotics and artificial intelligence. But in most operation fields, the achievement of full autonomy within the near future is not a realistic goal. In the meantime, humans and robots must co-ordinate harmoniously, that is to say, dynamically attune their task performances to each other like team members rather than like a worker and his or her tool. Machines must alert their operators of their results, ask for help in difficult situations or when they are not competent, eg. in [1]. Teams mixing humans and robots should behave increasingly like human teams.

In human teams, teammates adapt their communication patterns to different levels of workload [2]. For example, when a team member is busy, the rest of the team leaves him to his task or helps him, and is less likely to ask for his help. This results in adequate performance under increased workload conditions [3]. In contrast, despite the growing presence of mixed-initiative systems, machines are incapable of understanding and taking into account the mental state of other team members or of their operators. Most robots remain oblivious to the cognitive state of their operator, asking for help when their operator is already struggling to complete his own tasks, or abstaining from doing so when their operator is idle.

One reason for this state of affairs is the difficulty of measuring workload in real time. Dedicated physiological sensors allow for this, but require additional hardware and can be invasive or impractical. In the USAR domain in particular, robustness and simplicity are critical [4]. To comply with these requirements, we rejected physiological measures of workload during real operations, favouring instead an indirect approach (of course physiological measures can be used to validate the models in realistic test settings). We use communication between the user and the robot, along with environmental information already available to the robot, to infer the tasks performed by the operator. This resembles the natural behavior of human team members: despite the lack of EEG or heart rate variability data, team members manage to keep track of their teammates’ workload.

As humans do, the CTL-model aims to detect different types of problems associated with task load. For example, excessive workload can cause cognitive lock-up or overload, and repetitive tasks can cause vigilance issues. The CTL-model should be configured and integrated on a robot or other system; it will then produce, every second, a real-time estimation of the operator’s cognitive state. The robot can make use of this knowledge to take the best course of action or to adopt the best mode of communication.

This paper gives answers to the research questions: is it possible to give a reliable, real time estimation of an operator’s mental workload using task detection (without dedicated physiological sensors)? What limiting factors might occur?

II. BACKGROUND

The cognitive task load (CTL) model developed by Neerincx [5] describes the effects of task allocations on operator performance and mental effort. Cognitive Task Load consists of three metrics:

- Time Occupied (TO) is the amount of time that an operator spends performing tasks, divided by the total time available. It is expressed as a percentage over the time frame that is considered relevant - eg. the previous 120 seconds.
- Task Set Switches (TSS) is the number of times that an operator changed tasks during the relevant time frame.
- Level of Information Processing (LIP) is the type of cognitive processes required by recent tasks. It is expressed as an average of the LIPs of all tasks within the relevant time frame. For each task, LIP is defined in
relation to Rasmussen’s Skills-Rules-Knowledge (SRK) paradigm [6].

When the values for the 3 metrics fall into a certain range (corresponding to a certain region in CTL-space), the operator is diagnosed to be in a certain mental state. For example, \( TO > 80\% \), \( TSS > 5 \) and \( LIP > 2 \) could correspond to a state of cognitive overload. Depending on the region, appropriate measures should be taken: modifying the level of autonomy or adapting the interface.

Figure 1 gives an intuitive representation of the CTL-model as a cube, in 3D space. The dimensions parallel to the edges of the cube correspond to the metrics used. Different regions of the cube correspond to different mental states, including the optimal state (left transparent) and various problem regions (in light or dark gray). The boundaries of these regions are established experimentally and used to configure the model.

Fig. 1. Cognitive load space with four general problem regions [5].

The CTL model was validated for naval ship operators who “have to act in dynamic, critical and high-demand task environments” [7]. Using the model, the performance of an operator could be estimated with 74% accuracy in real conditions.

III. IMPLEMENTATION

Our CTL-model, once configured and integrated into a system, produces every second an updated diagnostic of the operator’s cognitive state. In this section, we describe some of the essential features of the implementation: the underlying concepts and formulas, and how the model is configured to domain-specific tasks.

For an intuitive example of how a fully configured implementation works in practice, see Figure 2, in Section IV.

A. Tasks

The CTL-Model bases its estimation of operator workload on the detection of the tasks performed by the operator. A task \( t_i \) has the following properties:

- A mental occupancy value \( o_i \), between 0 and 1. Mental Occupancy (MO) represents the degree of focus required for performing a task and is a refinement of the Time Occupied (TO) metric of Fig. 1. Indeed, some tasks only require intermittent attention (eg. looking in the mirrors while driving a car), and others only require a portion of an operator’s attention (eg. giving a running commentary while performing another task). The more a task requires of the operator’s attention, the higher the mental occupancy.

- A level of information processing value \( l_i \), between 1 and 3. Based on Rasmussen’s SRK paradigm [6], the level of information processing of a task identifies the type of reasoning needed for that task. It distinguishes tasks requiring complex manipulations of mental representations (eg. strategic decisions, knowledge-based) from those relying on reflexes (eg. reacting to a signal, skill-based). Each task is given a value in 1, 2, 3 corresponding respectively to the skill-based, rule-based and knowledge-based levels.

- A set of information domains \( I_i \). An information domain is a category of mental representations involved in performing the task, for example: “using the interface” or “avoiding obstacles”.

- A duration \( d_i \) based on a start time and an end time. If the task is ongoing, the end time is provisionally set to the current time; if the start time is not in the relevant time frame (too old), it is set to the beginning of the time frame.

B. Task modifications

After being detected, tasks can be modified in two ways. First through a task modifier: for example, in an USAR scenario, if a UGV detects a nearby obstacle, the system is considered in a state of “collision risk” and this causes changes to the properties of some relevant tasks (driving is more difficult). Second, in the case of multitasking: if two or more tasks are detected to be simultaneous, their overlapping parts are combined as a single task. For example, an operator can be detected to be talking and to be simultaneously driving the UGV. When this happens, tasks are combined two by two (several iterations may be needed to combine all overlapping tasks), so that the chronologically overlapping parts of \( t_1 \) and \( t_2 \) are deleted, and a task \( t_3 \) is added. The characteristics of the “multitask” \( t_3 \) are calculated as follow:

- Mental occupancy is the sum of its constitutive tasks, with an upper limit of 1: \( o_3 = \max(o_1 + o_2, 1) \).
- Level of information processing is the maximum of its constitutive tasks: \( l_3 = \max(l_1, l_2) \).
- Information domains is the union of the information domains of its constitutive tasks: \( I_3 = I_1 \cup I_2 \).
- The duration corresponds to the duration of the overlap between \( t_1 \) and \( t_2 \).

C. CTL metrics and diagnosis

Once all tasks have been detected, modified if needed, and combined if they overlap, it is possible to evaluate the mental state of the operator. This is achieved by using the three metrics LIP, MO and TSS. We describe how these metrics are calculated in the equations below. In each of these equations,
• \( f \) is the duration in seconds over which cognitive task load is calculated.
• \( T \) is the set of tasks performed by the operator during the last \( f \) seconds, ordered chronologically. \( T = \{ t_1, ..., t_n \} \), with each task \( t_i = \{ t_i, o_i, D_i, d_i \} \).
• \( l_i \) is the level of information processing value of the task \( t_i \).
• \( o_i \) is the mental occupancy of the task \( t_i \).
• \( I_i \) is the set of information domains of the task \( t_i \).
• \( d_i \) is the duration of the task \( t_i \).

1) **Level of Information Processing (LIP):** LIP gives an estimation of the type of mental activity that the operator is performing: is he automatically responding to simple signals, or attempting to figure out the solution to complex problems? LIP is calculated as the average information processing value of tasks, over the past \( f \) seconds. LIP varies between 0 (idle) and 3 (only “knowledge”-tasks).

\[
LIP(T_f) = \frac{\sum_{i=1}^{n} l_i d_i}{f}
\]

2) **Mental Occupancy (MO):** Time Occupied (TO) is the proportion of time during which an operator is performing a task. TO varies between 0 (fully idle) and 1 (fully busy). It is calculated by dividing the sum of the duration of tasks by the total time.

\[
TO(T_f) = \frac{\sum_{i=1}^{n} d_i}{f}
\]

TO can be refined as MO. MO is the proportion of an operator’s attention that was occupied by a task, during a period of time. In particular, in the case of multitasking, MO accounts for the increase in cognitive load caused by the additional mental effort of performing two tasks at once, whereas TO does not.

\[
MO(T_f) = \frac{\sum_{i=1}^{n} d_i o_i}{f}
\]

3) **Task Set Switching (TSS):** TSS accounts for the workload associated with switching attention from one task to another. It is measured relative to the information domains of each task. Switching between very similar tasks is considered less demanding than switching between very different tasks: switching between two tasks with identical information domains has no impact on TSS, whereas switching between two tasks that have no common information domains adds 1 to TSS.

\[
TSS(T_f) = \sum_{i=1}^{n-1} \frac{|(I_i \cap I_{i+1})^c|}{|I_i \cup I_{i+1}|}
\]

For example, considering a case in which the operator has only switched tasks once (there are only two tasks detected within \( f \)):

\[ T_f = \{ t_1 \text{ such that } I_1 = \{ “looking around” \}, \]
\[ t_2 \text{ such that } I_2 = \{ “driving”, “looking around” \} \]

Then:

\[
TSS(T_f) = \frac{|(\{ “driving” \} \cap \{ “looking around” \})^c|}{|\{ “driving” \} \cup \{ “looking around” \}|}
\]

\[
TSS(T_f) = 0.5
\]

4) **Diagnosis:** The operator is considered to be in a problem state if the values for the above metrics fall within certain ranges, corresponding to certain regions in the multidimensional space formed by the metrics. For example, the model can be configured so that when \( LIP(T_f) > 2, MO(T_f) > 0.85, \) and \( TSS(T_f) > 5 \), the operator’s state is set to “overload”.

**D. Task analysis and calibration**

The model must be configured and calibrated for the specific tasks for which it is used. In order to do this, relevant information regarding the tasks must be gathered:

1) A task analysis is performed to identify the tasks performed by the operator and the mental occupancy, level of information processing and information domains for each task.

2) **Experimentation is used to determine the optimal time frame \( f \) and the problem regions. It also allows for more precise estimations of the tasks’ properties.**

The implementation of the CTL-model implies making a compromise between the constraints of task analysis, task detection and the calculation of the CTL dimensions. The task analysis should produce the most meaningful possible description of an operator’s tasks. However the real-time detection of these tasks is dependent on the available sensors, on processor power and AI. When a meaningful task cannot be detected, or when two distinct tasks cannot be distinguished, the closest detectable approximation is used. For example, while we would like the model to distinguish task-relevant talking from irrelevant discussions, we settle for the detection of talking of any kind.

Section IV provides an example of a configuration of the system.

**IV. Testing**

Fig. 3. The NIFTi robot, the tunnel accident, and a firefighter operating the robot.

To test the model, data from an December 2011 experiment was used. Firefighters operated the NIFTi robot [8] to gain an overview of a car accident in a tunnel. The experiment was set in a high-fidelity environment at the training center of the National Fire Department of Italy (SFO).
A. Procedure

Ten firefighters took part in the experiment, of which six produced data we could use. Participants received 30 minutes of training with the robot, after which they performed a battery of tests. Finally, they had to explore a tunnel in which a car accident was simulated. They had no direct vision of the tunnel; instead, a robot (see Figure 3), controlled by the participants, was deployed to gather information. The participants were asked to answer the following questions:

- Are there cars in the tunnel? If so, where are they?
- How is the layout of the situation?
- Are there victims? And if there are, how many were there, and where?
- Have you seen fire and/or dangerous substances, depicted by pictures of warning signs? If so, where?

During the mission, the following data were recorded:

- The operators were prompted every 2 minutes to estimate their current workload on a scale from 1 (none at all) to 5 (far too much). In addition, the reaction time to answer this question was logged.
- The operator’s heart variability was measured with a belt around the chest.
- The beginning and end times of user tasks was marked by using the Observer software [9].

For more information on the experiment, please see [10].

B. Model configuration

A task analysis was performed to determine the main detectable tasks performed by the operator of the UGV and their characteristics. A summary of this task analysis can be found in Table I.

Figure 2, above, gives an example execution of the model using this configuration, based on a task pattern typical of the participants of the end-user experiment.

C. Processing the data

The model normally makes a qualitative diagnostic about the state of the operator (eg. “optimal workload” or “vigilance risk”, as shown in Figure 2). But for the purpose of this test, the model was configured to provide an estimation of the operator’s mental workload (MWL), from the 3 dimensions of the CTL-model (LIP, MO, TSS). To achieve this, all metrics were projected on the $[0, 1]$ interval, so that all points would fit within a cube of dimensions 1, 1, 1. We conjectured that mental workload is highest when all metrics are high, and that each metric, independently of the others, can only have a limited impact on mental workload. To achieve this, we used the following equation (Figure 4 provides a visualization):

$$\text{MWL} = \frac{d_{\text{origin}}}{d_{\text{diagonal}}}$$

As the experiment took place before the CTL-model was implemented, the test was performed off-line. The output of the CTL-model regarding workload (MWL) was then compared to the results of physiological and subjective indicators of mental workload. It was judged pertinent to use multiple...
metrics, as the current literature is inconclusive regarding which heart rate metrics are most efficient (compare [11] and [12]). Furthermore, physiological metrics and subjective metrics may provide complementary data [13]. In this study, the following indicators of workload were used:

- Average heart rate over the previous 120 seconds (HR).
- Ratio of low frequency (0.07Hz-0.15Hz) to high frequency (0.15Hz-0.5Hz) heart rate variability (L/H).
- Subjective mental workload as evaluated by the participants themselves (SWL).

It was expected that the metrics of the CTL-model would correlate positively with all indicators. In order to account for the effects of learning and fatigue, time elapsed was controlled for.

**D. Results**

The indicators of workload used to control the value provided by the model, SWL, HR and L/H, have almost no significant correlations ($p > 0.1$) between one another. The only exception is participant 5, for whom two indicators (HR and L/H) consistently indicate opposite amounts of workload. Overall, SWL, HR and L/H turned out to be inconsistent indicators of mental workload. This can be caused in part by the difference between physiological indicators (HR and L/H) and self-assessment (SWL), in part by the high level of noise inherent to a complex experiment, and in part due to the insufficient quantity of data (only between 13 to 17 data points could be used per participant).

For some participants, MWL correlated with HR and L/F. Participants 3 and 6 in particular were correctly modelled, according to physiological data (for participant 6, more precise data was available and confirmed the correlation between MWL and HR). But this did not succeed for others, eg. participant 4. A closer look showed that participant 4 was very focused when he considered a task important, causing TSS to be low when workload was high and *vice versa*: it is this personal, high-level strategy that accounts for the results of the model for this participant.

MLW did not correlate significantly with SWL for any of the participants. This could be due to the lack of effort that some participants put in answering the SWL prompt: prioritizing their main task (exploration), they would answer the prompt dismissively. For example participant 5 gave a constant answer to the prompt despite having among the highest standard deviations for both HR and L/F, suggesting he underwent high variations in workload.

**V. CONCLUSION AND DISCUSSION**

**A. Conclusion**

The three indicators of mental workload, SWL, HR and L/H, did not correlate sufficiently with each other to provide a solid estimate of mental workload, and to act as a baseline value against which MWL could be compared. A longer and higher-pressure experiment will provide more data points and a higher workload variation for such a validation.

Some conclusions can still be drawn from the experiment. The CTL-model was previously successfully tested in a dynamic environment [7]: operators had to solve urgent tasks as soon as they appeared, leaving little time for establishing any strategy. In contrast, the environment in our experiment was mostly static, allowing operators to create high-level strategies. The firefighters could decide to focus on one task at a time, to stop all (observable) activity for a while to think of a different approach, etc., which probably resulted in a stable workload. For example, if an external cause of mental workload (a distraction not measured by the CTL-model) occurred, firefighters could reduce their activity to cope with it. The model would then detect no change in workload (or maybe even a decrease where there is none).
In short, in this experiment the participants had a lot of freedom to adapt their behavior based on workload, which was not the case in a previous experiment. The behavior of some participants (eg. participant 4) suggests that the validity of the model is limited to cases in which workload is imposed on the operator: this is the case in highly dynamic environments in which reaction time matters. In a static environment, it may be better to trust operators to manage their own mental workload optimally.

B. Discussion

We presented a model of an operator’s cognitive state. Using task data only, the model classifies the operator’s cognitive state among several possible cognitive states, based on which a robot can adapt its behavior to improve team performance. The experiment however didn’t allow for a good evaluation of the model, as the control indicators (heart rate, ratio of low frequency to high frequency heart rate variations, and subjective workload) did not correlate with one another. The behavior of some participants indicated that they were engaging in high-level strategies, some of which affected our metrics in undesired ways. This lead us to conclude that the model performs best on task patterns that are imposed by the environment, rather than decided by the operator.

C. Future work

The results call for a new experiment to validate the model, taking advantage of the lessons learned: in a more demanding setting, modeling in priority activity and inactivity that is imposed on the operator (rather than chosen by him). An even more realistic environment, with a more advanced robot, may solve some of the issues brought by real-life mental workload estimation: if the operator is receiving orders, and if the robots asks more questions or asks for help more often, workload may be less dependent on the operator’s personal strategy. Furthermore, we will adapt the model to the behavior of individual operators through machine learning. This would allow for more precise estimations, and possibly for taking into account behavior associated with individual strategies. Given better results in estimating the operator’s cognitive states, the next step is to adapt the interface and autonomy of a robot to the mental workload of its operator, in real time, and measure performance gains of the team within which the robot is operating.

REFERENCES