

Effectiveness of Physiological and Psychological Features to Estimate Helicopter Pilots' Workload: A Bayesian Network Approach

Patricia Besson, Christophe Bourdin, Lionel Bringoux, Erick Dousset, Christophe Maïano, Tanguy Marqueste, Daniel R. Mestre, Sophie Gaetan, Jean-Pierre Baudry, and Jean-Louis Vercher

Abstract—Despite growing interest over the decades, the question of estimating cognitive workload of operators involved in complex multitask operations, such as helicopter pilots, remains a key issue. One of the main difficulties facing workload inference models is that no single specific indicator of workload exists, so that multiple sources of information have to be inputted to the model. The question then arises as to the nature and the quantity of features to be used for increasing model performance. In this research, done in cooperation with Eurocopter, the effectiveness of physiological, psychological, and cognitive features for estimating helicopter pilots' workload was systematically investigated, using Bayesian networks (BNs). The study took place in two different contexts: a constrained laboratory situation with low ecological validity and a more realistic and challenging situation relying on virtual reality. The constrained conditions of the laboratory study allowed us for testing various combinations of entropy-based physiological, cognitive, and affect features as inputs of BN models. These three different kinds of features are shown to carry complementary information that can be used with advantage by the model. The results also suggest that increasing the number of physiological inputs improves the model performance. The second study aimed at challenging some of these conclusions in a more ecological context, by using the NH90 full-flight simulator of the Helisim company. The results emphasize the problem of accessing the ground truth, as well as the need for an efficient feature selection or extraction step prior to the classification step.

Index Terms—Cognitive science, graph theory, human computer interaction, human factors, intelligent systems.

I. INTRODUCTION

HELICOPTER pilots are involved in complex multitask activities where they constantly have to make quick and appropriate decisions. Advanced systems provide them with some assistance, by automating some processes and by delivering large amounts of information on task context. However,

automated agents impose additional information processing demands that might also increase the level of cognitive workload (denoted as *workload* from now on) and be not relevant neither to the task context nor to the real needs of the helicopter pilots (individualistic approach) [1]. Consequently, it would be more efficient to develop intelligent systems, able to adapt the assistance to the current level of pilots' workload [2]–[5]. Being able to estimate and to characterize operator's workload in real time is, therefore, a prerequisite to adaptive intelligent systems.

However, despite intensive researches in the domain, no reliable estimator of operators' workload has shown up yet. On the one hand, activity analysis and subjective evaluations, conventionally used in human factor studies, are well established approaches, but the methods or information they rely on are not suitable for real-time estimation. These methods may be also considered as too subjective and task dependent, making it difficult to build generic systems. On the other hand, several physiological signals, a continuously available information, have been shown to be sensitive to workload, but no single indicator specific to workload does exist. This makes the problem highly complex: these signals must be exploited jointly, whereas the relationship between these measurements and the “true” subjects' psychophysiological state cannot be directly accessed [6]–[9]. As a result, a wide variety of measurements are proposed as inputs to workload estimation models, raising the questions as to the complexity of these models and the contribution of each measurement to workload estimation from a computational point of view. Furthermore, when dealing with estimation of psychophysiological states, the transfer of the in-lab observations and methodologies to more ecological situations (simulator or real life) is not straightforward (as shown for example in [10] for fatigue evaluation).

The objective of this work was twofold: while using Bayesian networks (BNs) to estimate workload in the context of helicopters' piloting, we aimed at using them as a systematic tool to investigate, from a computational point of view, the aforementioned points. Therefore, the question of the effectiveness, in terms of both quantity and type, of different psychophysiological features on workload estimation is first addressed in a constrained laboratory study with low ecological validity. Second, experiment took place in a full-flight helicopter simulator, and we discuss the difficulties coming out when a more realistic and challenging situation is used. Our choice for BNs was motivated by their flexibility and ability to model complex relationships

Manuscript received October 15, 2012; revised February 20, 2013 and April 26, 2013; accepted June 10, 2013. Date of publication August 6, 2013; date of current version November 26, 2013. This work was supported by Eurocopter. The Associate Editor for this paper was C. Wu.

P. Besson, C. Bourdin, L. Bringoux, E. Dousset, T. Marqueste, D. R. Mestre, S. Gaetan, and J.-L. Vercher are with Aix-Marseille Université, CNRS, ISM UMR 7287, 13288, Marseille cedex 09, France.

C. Maïano was with the Aix-Marseille Université, CNRS, ISM UMR 7287, 13288, Marseille cedex 09, France. He is now with the Cyberpsychology Laboratory, Department of Psychoeducation and Psychology, Université du Québec en Outaouais, Gatineau, QC J98X 3X7, Canada.

J.-P. Baudry is with the Human Factors and Cockpit Design Department, Eurocopter, 13725 Maignane, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2013.2269679

among random variables (RVs), capturing both qualitative and quantitative knowledge [11], [12].

As we just mentioned, many psychophysiological features and machine-learning techniques have been proposed to infer cognitive states, such as workload, distraction, or stress [13]. Task performance analyses, sensorimotor features (gaze, head movements, etc.), or physiological measurements have been shown to be useful [9], [14]–[16]. Thus, electrocardiogram (ECG), electromyogram (EMG), skin conductance (SC), and respiration were used in [17] to infer the level of stress experienced by drivers using linear discriminant analysis. In [18], the authors developed an artificial neural network (ANN) taking electroencephalogram (EEG), electrooculogram (EOG), and respiration as inputs to assess workload levels. ECG, EEG, and EOG were also used in [19] to derive an information-theoretic indicator of cognitive state. A support vector machine and an ANN were applied to workload estimation in [20], using EEG, SC, respiration, and heart rate (HR) data. These works might also rely on reaction times (RT) to a secondary task, shown to be a valuable cognitive indicator of resource allocation and, indirectly, of workload [3], [6].

Additionally, several works have addressed the problem of inferring users' affects from physiological signals [21], [22]. These works take place in the field of affective computing, which aims at endowing machines with emotional skills, in particular, the ability of perceiving and adapting to the user's current affective state to improve the efficiency of human-machine interfaces. At a first glance, affect and workload estimations may seem unrelated problems, but we can see at least two reasons why they should be considered jointly. First, affective computing is based on findings suggesting that humans' emotional intelligence provides them with *the capacity to reason about emotions* and *of emotions to enhance thinking* [23]. As a result, the subjects' affective states are also subjective and mostly related to their own perception or analysis of the context at hand, as well as to their ability to face the situation. Thus, Isen [24] showed that positive affects enhance problem solving and decision making. Moreover, according to cognitive appraisal theory, psychological stress results from assessment of a discrepancy between situational demand and subject's resource [25], [26]. In that sense, affects and workload are obviously linked. Second, physiological signals are exploited in models inferring either workload or affects, which means that these signals carry information related to both elements. Being able to distinguish between workload-related and affect-related information would allow for noise filtering and lead to improved model performance.

Based on the literature, we used physiological, cognitive, and affective features in our study, the ground truth being provided by subjective measurements. We let aside some features commonly used to infer operators' cognitive states, such as behavioral (e.g., action commands), performance (e.g., flight data), or sensorimotor (e.g., eye movements) features, because they require to be interpreted in connection with some knowledge about the current surrounding conditions to make sense. In the specific context of helicopters' piloting intended here, the relationships between pilot's actions and helicopter's motions show substantially larger degrees of freedom than those attached

to car driving, or even, airplane flying contexts, for example, making the interpretation of these behavioral or sensorimotor features very complex. Therefore, we decided not to use them in this work. We also decided not to measure EEG (incompatible to date with helmets wore by helicopter pilots) since we wanted to use noninvasive and minimally intrusive sensors.

Thus, in the laboratory study, we used EMG, HR, SC, and respiration. The RT to a secondary task (dual-task paradigm) was chosen as a cognitive feature. As previously stated, RT is expected to vary when more mental (cognitive) resources are demanded by the task. Finally, the positive and negative affect state scale (PANAS) [27], [28] evaluated subjectively the participants' affective state during the experiment.

Different BN structures, built from expert knowledge, were tested, using in turn several combinations of these features. Their performance was evaluated in terms of two criteria to be jointly optimized: the accuracy (i.e., how close from the workload level the model prediction is) and the diversity (i.e., the ability of the model to be functional for different subjects). Indeed, the literature points out the problem of large differences among individuals, when trying to estimate psychophysiological states (see [29]). In addition, notice that part of the results related to this study has been presented in [30] and [31].

In the virtual reality study, the experimental protocol took place in the NH90 full-flight simulator of the Helisim company. Physiological measurements (EMG, HR, SC, and respiration) were collected during two offshore mission scenarios, and BNs were used to infer the pilot's workload. In this second study, the level of workload was directly estimated and compared with the self-evaluation made online by the subjects.

Sections II Sections III Sections IV describe the laboratory study. Section II describes the experimental protocol used to collect representative data. Section III presents the BN models tested on the data set, i.e., the performance being assessed in Section IV, in terms of two criteria to be jointly optimized: diversity and accuracy. The virtual reality study performed on the helicopter simulator is presented and discussed in Section V.

II. LABORATORY EXPERIMENT

A. Subjects and Material

Ten subjects (nine males and one female, aged 30 ± 10.7 years) with normal or corrected to normal sight and hearing, participated in the first experiment, which aimed at investigating the problem of workload estimation in a controlled environment.

The subjects sat in darkness, facing a standard 24-in monitor, where graphical dynamic flying scenes generated by the home-grown ICE software were displayed. An experimenter's computer was used to acquire all the data synchronously, using the Captiv Software (TEA, France). These data were made of the simulation data (e.g., aircraft position) sampled at 100 Hz and of the physiological data acquired at a sampling rate of 2048 Hz using the FlexComp Infinity sensors and encoder (Thought Technology Ltd., Canada). The subjects wore stereo headphones, so that they could hear the prerecorded instructions (the instructions' tone and content were then strictly identical



Fig. 1. Screenshot of a typical flying scene created by our home-grown software ICE. The ratio of hit rings over the total number of rings in the trajectory appeared on the cockpit dashboard (e.g., 1/60), and a green or red light indicated whether the last ring had been hit or missed.

for each subject) and the task-related noises such as the engine noise (leading to greater immersion) or the possible alarms.

B. Procedure

Using a standard joystick, the subjects were asked to pilot an aircraft and to do their best to follow a trajectory defined by 60 rings, alternatively red and yellow (see Fig. 1). They were instructed to fly through the center of each ring, without missing any if possible. The trajectories varied only along the vertical dimension. The aircraft's speed was maintained constant at the same predefined value for all the trajectories. The score, which is defined as the ratio of hit rings over the total number of rings in the trajectory, appeared on the cockpit dashboard. A green or red light indicated whether the last ring was hit or missed.

The experiment was organized in five sessions of six trials. Each trial lasted approximately¹ 90 s. In the first three sessions (labeled S1D1, S2D2, and S3D3), the subjects were presented with three different trajectories of increasing difficulty (D1, D2, and D3). The trajectory difficulty was an independent variable meant to manipulate the task workload requirement. It was varied by changing the vertical distance between two successive rings, while keeping their depth distance constant. In the last two sessions (labeled S4D1 and S5D3), the subjects were asked to fly again on the simplest and hardest D1 and D3 trajectories and to try to beat their own mean scores over these trajectories. Moreover, a strident alarm sound was emitted in case of a missed ring. This challenge and the alarm were introduced in order to maintain the subjects' motivation and involvement in the task.

For each of the five sessions, a dual-task paradigm was introduced. Two geometrical shapes (a square or a triangle) appeared on the screen during 1 s, at pseudo-random positions (the ring apparition zone was avoided and the same number of targets appeared in each of the four screen quarters, with eccentricity values ranging from 10° to 25°) and at pseudo-random times (no apparition while the ring was crossed and a minimum time interval of 1.5 s between two successive targets). The subjects had to press a button on the joystick with the forefinger, as quickly as possible, in response to the square

target apparition. They should not react to a triangle target. Fig. 1 shows a typical screen shot of the simulated scene.

C. Dependent Variables

Performance on the primary (percentage of hit rings) and secondary tasks (false and good detection rates; RT) were recorded. The physiological variables comprised the following measurements:

- HR, estimated from the ECG by the Captiv software, using R-R intervals;
- Root mean squares (RMS) of the *flexor digitorum* EMG (RMS1) of the dominant arm and of the right *trapezius descendens* EMG (RMS2);
- Tidal volume (Vt): respiration measured through chest expansion;
- SC, measured using electrodes placed on the first and little fingers of the nondominant hand (temperature in the room equal to 19.33 ± 0.98 °C).

Psychological data were also collected at the end of each session. The subjects self-assessed their own workload during the performed task, using the NASA Task Load Index (TLX) scale [32]. The NASA TLX asks the subjects to rate their perceived workload on six different subscales. At the end of the experiment, these six components were matched two by two, and the subjects had to choose for each couple which component best described the workload in the performed task. Each component score can thus be weighted accordingly to the number of times it has been chosen in the matching phase. In the present experiment, the NASA TLX rates on the six subscales were weighted and summed for each session to result in a single *Workload Index* (WI) (normalized on [0, 1]) per session. The subjects also completed the PANAS questionnaire [27], [28] after each session. This instrument was used to provide information on participants' affective reactions to the experiment. It comprised the two affects' scale: positive (i.e., interested, excited, etc.) and negative (i.e., distressed, scared, etc.). The 20 items of this instrument were rated on a five-point Likert scale from: (1) = "very slightly or not at all" to (5) = "Extremely". The ratings for each participant were normalized between 0 and 1, and the positive affect scores (PA) reversed ($1 - PA$) for all the scores to be interpretable in a consistent way. These normalized negative and reversed positive scores were then summed and normalized on [0,1], resulting in the *Affect Index* (AI) used in the following of the paper.

III. COMPUTATIONAL ANALYSIS

A. Selection of Output and Input Features

Analysis of variance (ANOVA) statistical tests show that the workload level has been effectively manipulated using our experimental paradigm: the WI scores increased with difficulty, with significant differences between sessions ($F(4, 26) = 12.284$, $p = 0.000$). Hence, variations observed in the physiological signals can be expected to effectively correspond to variations of workload. Notice that the interested readers are referred to [30] and [31], for more detailed statistical analyses of the experimental data.

¹Although the speed was maintained constant, the duration of each trial was not necessary the same, since the aircraft's trajectory could be more or less sinusoidal.

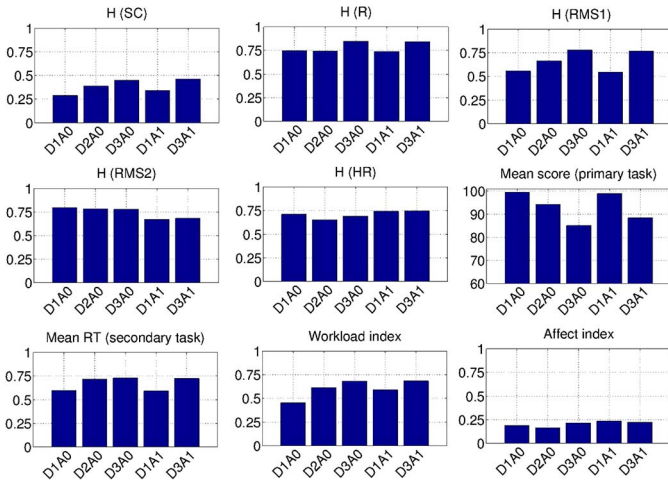


Fig. 2. Mean physiological feature values (entropy H of the physiological signals, in bits), performance on the primary task (in %), RT (in seconds), workload and affect indexes, for each session performed by subject 4.

We expected (and we observed) variations in the physiological signals as the subjects' cognitive state varied. The physiological signals we recorded are mainly under the control of the autonomous nervous system, which regulates the body's homeostasis through successive activations of the sympathetic and parasympathetic systems (resulting in mobilization or slowing down of the organism) [33]. These variations can be captured by Shannon's entropy of the physiological data, which is a measure of the average uncertainty in RV X [34]. The input for our model is then the entropy of the physiological data.

Before estimating the entropy values, the noise in the raw signals was first smoothed using a low-pass median filter. The first and last seconds of each trial signal were also removed to avoid possible starting and ending effects. Then, the data were normalized between 0 and 1, taking the minimal and maximal values observed on the first three sessions (used as training sessions). Entropy values were estimated on 15-s-long windows slid by 5 s along the signals, using a histogram of 41 bins that ranged on $[0, 1]$. Therefore, there were about 90 values per session and per subject. Entropy values were also normalized between 0 and 1, by taking the maximal and minimal values over the first three sessions for each subject. For simplification purposes, the RV denoting the entropy features were named as the acronyms of the corresponding physiological data (SC, Vt, HR, RMS1, and RMS2).

It could be observed that, generally, variations of mean entropy features were consistent with variations of performance on the primary task, RT, and WI (see Fig. 2 showing subject 4's features as an illustrative example). However, we observed the variations of the physiological data to be idiosyncratic, e.g., for some subjects (as for the subject 4), the mean entropy values of SC increases with the difficulty level, whereas it decreases for some others. As a result, the models were individual (trained and tested on each subject separately).

B. Model Definition

BN models were defined to infer the subject's WI on each session, from different kinds of features.

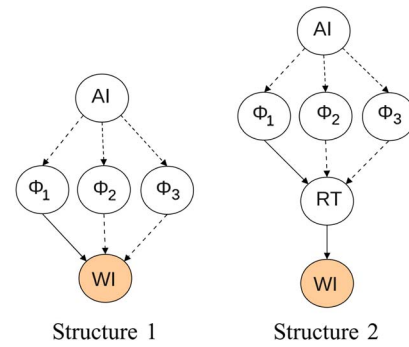


Fig. 3. BN models inferring the WI value from physiological features Φ_1 , Φ_2 , and Φ_3 either directly (Structure 1) or via RT (Structure 2). The models were tested with or without an AI node as parent of the Φ nodes (thus, this edge is dashed on the graph). There could also be either one, two, or three physiological nodes in the model.

To examine the impact of the number and type of physiological inputs on the models' performance, different classifiers were tested, each taking one, two, or three² of the possible physiological features SC, Vt, HR, RMS1, and RMS2 as inputs. The generic denomination Φ_i , with $i \in \{1, 2, 3\}$, is used to refer to these physiological nodes. The structure of these models where WI, the estimate of the subject's workload, was a direct child of the physiological nodes is referred to as *Structure 1*. Then, the impact of adding to the model a cognitive feature, namely, RT, was evaluated. An RT node could be added to the model at different positions. However, in our opinion, RT should be a child of physiological nodes because we consider that probability of changes in RT is conditioned on changes in body arousal. Thus, models with *Structure 2* were also tested, where WI was a child of RT, which was itself a child of the physiological nodes. Finally, we assessed how the affective state of the subject could impact the physiological data, and hence, estimation of the WI score. As stated in Section I, changes in the affects are known to induce physiological changes [35]. Therefore, an AI node was also introduced as a possible parent of the physiological nodes.

The different models are presented in Fig. 3. Notice that, if some more complex structures were tested in [30], only naive BNs were tested here, for the primary question we addressed was the effectiveness of psychophysiological inputs to estimate workload, not the study of the "true" relationships between these features.

The joint probability density functions (pdfs) described by the BNs were estimated on the training set using histograms with the following parameters (RV take on values in $[0, 1]$, but RT, taking on values in $[0, +\infty[$): five bins of width 0.2 for the physiological RV, 20 bins of width 0.05 for WI and AI, and 16 bins of width $\exp^{(0.2)}$ for RT, with the first bin being centered on $\exp^{(-3.7)}$ and the last bin taking all the values greater than $\exp^{(-0.9)}$. For each subject, the training set was made of the data collected on the three first sessions S1D1, S2D2, and S3D3. The testing set was made of the two last sessions S4D1 and S5D3. Both the learning and inference stages were implemented using the Bayes Net Toolbox for MATLAB [36]. Because there were some missing data (HR in

²Our sample sizes were not large enough to let us test models with more physiological nodes.

particular could not be reliably recorded sometimes, and there was not necessary one RT value per measurement window), the expectation maximization algorithm was used (with a stopping criterion of ten iterations).

C. Assessing the Models' Performance

The performance of the models was assessed by looking at the differences in the estimated WI between S4D1 and S5D3 sessions. Rather than assessing the ability of the models to infer *raw workload level* itself, we focused on how good they were in inferring the *workload change* between successive tasks: in the context of defining adaptive intelligent systems, estimating a workload change opposite to reality may drive the system to undertake actions opposite to those required by the operator's state, with dramatic consequences, and should be absolutely avoided. The model output was deemed as correct if the observed and inferred WI were evolving the same way, that is, if the performance index ρ , defined as follow, was positive:

$$\rho = \text{sign}(\Delta) \cdot \text{sign}(\Delta^*) \cdot \left| \frac{\Delta^*}{\Delta} \right|, \quad \text{if } \Delta^* < \Delta \quad (1)$$

$$= \text{sign}(\Delta) \cdot \text{sign}(\Delta^*) \cdot \left| \frac{\Delta}{\Delta^*} \right|, \quad \text{else} \quad (2)$$

where Δ was the difference between the subjects' WI on sessions S4D1 and S5D3, and Δ^* was the difference between the predicted WI on these two sessions. The quality of the model performance was given by the distance to 1 (the closer, the better). A fine analysis of the model's false detections was useless, since we wanted this false detection rate to be null.

For each model, we looked at the performance over the subject set. Thus, we wanted the maximum number of subjects to be correctly detected, with an accuracy as close as possible to 1. This is a two-variable optimization problem: the percentage of subjects correctly detected assessed the *diversity* S of the model, whereas the normalized area under the ρ curve, plotted as a decreasing function of S , defined the *accuracy* θ of the model

$$\theta = 10 \cdot \frac{\sum \rho}{S}, \quad \theta \in [0, 1]. \quad (3)$$

In the next section, we will use both S and θ indicators for determining which inputs can be of some interest, not only for accurately estimating workload but also for this accurate estimation to be maintained across individuals (a very important point, yet often overlooked).

IV. MODELS' PERFORMANCE ON LABORATORY EXPERIMENTAL DATA

A. Impact of Adding Physiological, Cognitive, and Affective Inputs on Models' Performance

Most of the models perform well in either the *diversity* (S) or *accuracy* (θ) criteria, as shown in Fig. 4. However, we were interested in classifiers efficient in both dimensions simultaneously. These are referred to as the *good models*, i.e., the

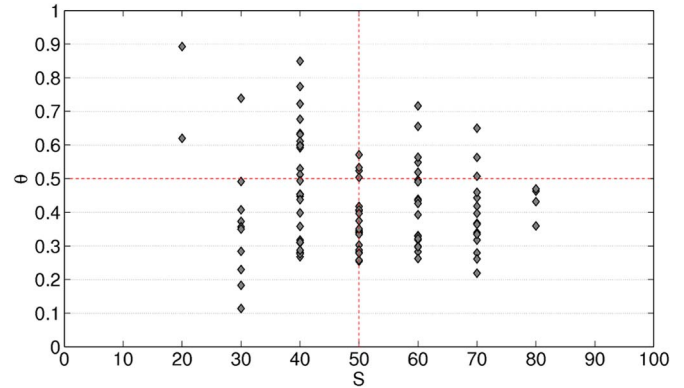


Fig. 4. Performance of the models in terms of diversity (S score) and accuracy (θ score). The *good models* (efficient in both dimensions simultaneously) lie in the upper right-hand side quarter of the graph.

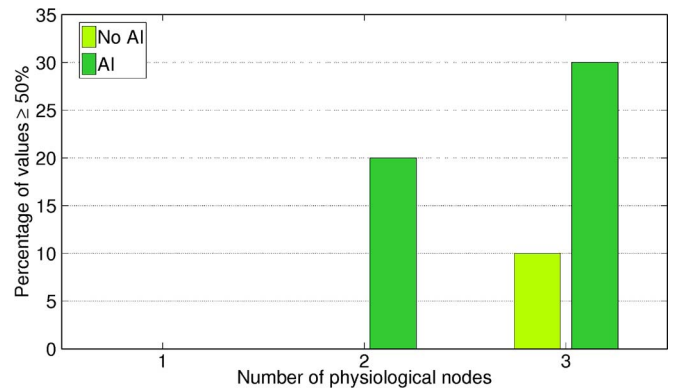


Fig. 5. Percentage of good models (with performance greater than 50% for both accuracy and diversity criteria) for the different physiological node numbers (no distinction is made between the two possible BN structures). Models with or without the AI node are compared.

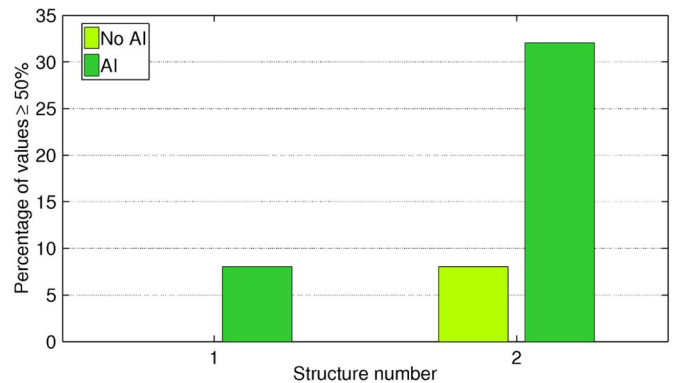


Fig. 6. Percentage of good models (with performance greater than 50% for both accuracy and diversity criteria) for the different structures (whatever the number of physiological nodes). Models with or without the AI node are compared.

models with a performance greater than 50% over the sets for both diversity and accuracy and lying in the upper right-hand corner of the graph in Fig. 4. The percentage of *good models* for different structures and input variables are presented in Figs. 5 and 6. It can be observed that increasing the number of physiological nodes increases the model performance (see Fig. 5), as well as adding RT to the model (i.e., using structure 2: see Fig. 6).

The highest improvement was obtained by adding an AI node (i.e., information about the subjects' affective states). Thus, with an AI node, 20% of the models fit the *good models* criterion with two physiological nodes only, and this number reached 30% if three physiological nodes were used. Without the AI node, three physiological nodes were required to get some good models (10% only), and Fig. 6 shows that these models also contained the RT node (structure 2). Finally, when structure 2 was used (whatever the number of physiological nodes) and an AI node was included, the percentage of good models reached 32%.

B. Discussion of the Laboratory Study Results

Different numbers and types of inputs have been tested and compared in terms of two criteria to be jointly optimized: the accuracy and the diversity of the classifier.

The most striking result is that taking into account subjects' affective state increases performance on workload estimation. Indeed, increasing the number of physiological inputs and adding cognitive information (RT) in the model increases the number of *good models*, as could be expected from the literature (see, e.g., [6] and [8]). However, coupling these conditions with some knowledge about the subjects' affects (through the use of features derived from the PANAS questionnaire outputs) outstandingly improves the models' performance.

Since some studies (see, e.g., [35]) have shown changes in physiological signals to be related with changes in affective states, we suggest that providing the model with information about affective states helps it getting rid of the physiological variations unrelated to subjects' workload changes. Of course, in this exploratory work, the information about the subjects' affect states are collected at the end of the task. Therefore, as it is, the model cannot be used as a workload predictor. Nevertheless, it is worth noticing that, in all the measurements we tested, only the physiological measurements were continuously available. From this perspective, they are essential to the model since they may lead to the nearest real-time estimation. Hence, the main lesson learned from this study is to point out that the modeling process, in conjunction with the experimental paradigm, should be designed such that the physiological variations related to workload can be dissociated from the variations related to other psychological states (the relationships between psychological and physiological data being not one to one [9]). Advanced feature extraction processes could be of some help on that matter.

At that stage, the physiological features we used, i.e., the entropy of the signals, appear to correctly catch the information related to workload variations. Each of the five proposed physiological features carries information related to the workload, since each of them appeared in one of the *good models*. Therefore, a model including all these data would probably outperform the proposed classifier. However, since we were training subject-dependent models, our sample sizes were too small to deal with a classifier with five physiological nodes; hence, we were not able to check this hypothesis.

This study also stresses the importance of adding cognitive features to the input set. Including the RT in the model yields better workload prediction, at the expense of a slightly more

task-dependent method, since it requires a secondary task to be performed. However, there are a lot of situations where routine tasks can be used to infer RT values.

These are only preliminary results, and refinements should be brought to the models, as well as tests on larger sample sizes (which should result in improved models' performance). In addition, a deeper (subject-by-subject) analysis of the results should be carried out, in order to check whether some combinations of specific physiological features are better for some categories of subjects (labile versus stabile for example; see [37]). Since affects varied between subjects, this subject-by-subject analysis could also bring to light some relationships between physiological and psychological data, beyond workload.

Nevertheless, as expected, this first laboratory study gave us some clues about the added value of different kinds of features to the models' performance. It also validated the potential of an approach relying on BNs to infer the operators' workload, while methodically examining the process.

V. VIRTUAL REALITY STUDY

A. Experiment

A second experiment has taken place in the NH90 full-flight simulator of the Helisim society. This simulator, mounted on hydraulic jacks with six degrees of freedom, is an exact replica of the NH90 helicopter. It aims at training pilots rather than doing research; therefore, it provides a highly realistic environment, with low flexibility, however.

Six subjects, all mastering the piloting of the NH90 simulator, took part in this experiment. Two of them (subjects 4 and 5) were professional helicopter pilots with more than 5000 h of flight and a good expertise on the kind of scenarios we proposed (offshore missions).

The subjects had to flight on two offshore missions, taking off from "Marignane airport" (France) and reaching in turn different offshore platforms. The two missions were planned to last about 20 and 30 min, respectively. The subjects were briefed about the missions before to take off.

The first mission aimed at familiarizing the subjects with the experimental conditions and at recording the level of the physiological signals in quiet conditions. The second mission aimed at pushing the subjects to their limits and to make them experience different levels of workload. However, the scenario had to stick to reality for the subjects to stay involved in the task.

For each of the two missions, a guideline scenario had been defined with possible incidents planned at some specific phases of the mission, in order to induce different workload levels. The sequence of workload level variations was the invariant that the experimenter tried to stick to as much as possible. To this end, the pilot reactions and decisions during the flight were analyzed in real time, and elements of the scenarios, such as weather conditions or occurrence of some possible technical incidents, were manipulated online by the experimenter to bring the pilot to the current expected workload level. A graphical representation of the guideline scenario used for the second mission is shown in Fig. 7.

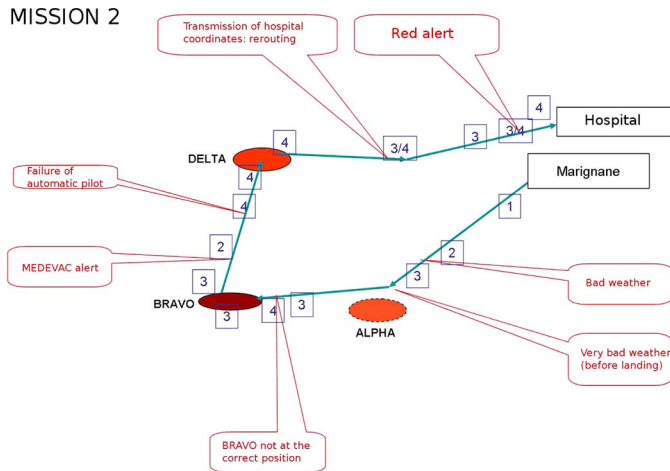


Fig. 7. Graphical representation of the guideline scenario used for the second mission. The subjects took off from Marignane airport and had to reach three offshore platforms (Alpha, Bravo, and Delta). Different incidents could occur during the flight, in order to bring the pilot to different workload levels. Other incidents could be introduced, or, on the contrary, some of the intended incidents canceled, if needed. The numbers 1 to 4 stand for the expected levels of workload at different instants.

During the flights, the physiological measurements SC, RMS1, RMS2, and RMS3 were acquired. SC, RMS1, and RMS2 were the same measurements than recorded during the laboratory experiment (see Section II). RMS3 was the RMS of the left *trapezius descendens* EMG. The more realistic conditions of the experiment made the acquisition conditions more challenging for the simple sensors we used: the motions of the pilots (who were wearing flight suits) led to very noisy ECG and Vt (ventilation) signals, and we have decided not to exploit them in the model. For every t_a time, a recorded voice asked the subjects to verbally rate their workload level (*workload* was defined during the briefing) on a rating scale going from 1 (very low workload) to 4 (very high workload). t_a was set to 3 min for subjects 1 and 2, and to 1 min and 30 s for the other subjects. Some additional spontaneous demands could be made between two automatic demands, with a minimal refractory time of 30 s.

Unfortunately, although the laboratory studies have shown the RT to a secondary task and the subjects' affect to be valuable to the model (see Section IV-B), none of them could be exploited in this simulator experiment. The simulator could only time events to the second, a precision too small for analyzing RT, and online evaluation of the subjects' affects would have improperly interfered with our workload manipulation primary objective. In addition, recording these two features would have somehow decreased the ecological validity level of the study.

B. Model

A preprocessing step similar to the one presented in Section III was performed on the physiological signals (low-pass filtering and normalization of the signals between 0 and 1, using the maximal and minimal values found on the first mission). The two missions were segmented in intervals of size $[t_a - 45, t_a + 45]$, where t_a is the instant of workload evaluation demand, in seconds. Thus, each interval, or segment, is labeled with the workload score given by the subject at t_a .

TABLE I
PERFORMANCE OF THE BN MODELS TAKING THREE PHYSIOLOGICAL NODES AS INPUTS AND TESTED ON THE DATA ACQUIRED ON THE NH90 SIMULATOR, FOR THE SIX SUBJECTS S

Physiological nodes Φ_i	S1	S2	S3	S4	S5	S6	Mean performance
SC;RMS1;RMS2	25%	43%	100%	64%	46%	44%	53.7%
SC;RMS1;RMS3	15%	50%	67%	100%	55%	36%	53.8%
SC;RMS2;RMS3	20%	32%	46%	100%	45%	36%	46.5%
RMS1;RMS2;RMS3	10%	39%	43%	36%	58%	33%	36.5%

These values defined the ground truth. For each segment, Shannon's entropy values were estimated on 15-s-long windows, slided by 5-s steps along the signal. The pdfs of the data were estimated using 100-bin histograms on the range $[0, 1]$. The entropy was then normalized between 0 and 1, using the maximal and minimal values found on the training data set.

Naive BNs with structure 1 [see Fig. 3(a)] have been used, taking three physiological features as inputs. The output of the network, i.e., W , was the subjects' workload for the current time segment, as subjectively self-evaluated. Since the laboratory study had suggested that increasing the number of physiological inputs would lead to better model performance, we also tested a model taking the four physiological features as inputs. However, the addition of a fourth node increases greatly the dimension of the joint pdfs described by the model. The results are then plagued by the curse of dimensionality and should be handled with care.

The pdfs of the BNs were estimated using histograms with the following parameters: for SC, five bins whose edges ranged from $\log(1)$ to $\log(3)$, by $\log(0.5)$ steps; for the RMS, three bins of width 0.3, covering the interval $[0, 0.9]$, with a fourth bin taking on all the values between 0.9 and 1; for W , four bins centered on $\{1, 2, 3, 4\}$.

A cross-validation scheme was used, where the training set was made of all the segments but one, used as the testing set. Each of the segments appeared, in turn, in the testing set. The performance of the model is assessed by looking at the percentage of correctly inferred workload levels for each segment (the number of segments being different for each subject because of different flight durations).

C. Results

The performance achieved by the three physiological node models is presented in Table I. Inputs RMS2 and RMS3 reflect the activity of the two *trapezius descendens* muscles and can then bring very redundant information. It can be observed that the two models, taking as input only one source of information about the *trapezius* muscle activity, achieved better performance than the model including both RMS2 and RMS3 (models SC;RMS2;RMS3 and RMS1;RMS2;RMS3). The four-physiological-node model lead to an average performance of 47.7%. This model exceeded a 50% good prediction rate only for subjects 3 and 5, and achieved very poor performance for subject 1 (5% only), whereas the good prediction rates were around 40% for the three remaining subjects.

D. Discussion of the Simulator Study Results

The objective of this second study, taking place in a helicopter full-flight simulator, was to discuss the difficulty raised by such a challenging situation (very close to real conditions). More precisely, we aimed at investigating to which extent the method and findings from previous laboratory study would stand up against less constrained and more true-to-life experimental conditions (with subjects experiencing stronger sensations and pressure than in laboratory experiment, in particular).

A first limit quickly showed up, as inputs labeled as important from the laboratory experiment (affect and RT on a secondary task, in particular, but also some physiological signals that could not be reliably recorded) could not be used directly and simply in this new challenging context. This is already an indication of the robustness of these features outside laboratory conditions.

Let us then focus on the robustness of the previous findings, with regard to the entropy-based physiological features. The best results, in terms of both accuracy and diversity, are obtained when the two possible sources of information about the *trapezius* muscle activity (RMS2 and RMS3) are not both inputted to the model. When one of them is solely present, as a complementary input with SC and RMS1 (the EMG-based feature of the *flexor digitorum*), a similar average performance is reached (models SC;RMS1;RMS2 and SC;RMS1;RMS3, 54% of correct estimation in average). The performance is also higher than with the four-physiological-node model (whose results must be analyzed with care, however, due to the curse of dimensionality).

These results complete the laboratory study findings in an interesting way, as they draw attention to the question of optimizing the model's inputs. Laboratory study analysis suggested that increasing the number of physiological nodes would increase the model performance. However, this also increases the model complexity, raising the question of the size of the required training sample and of the model's predictive power. In addition, it is worth noticing that the added features should be carefully selected in order to carry valuable (i.e., complementary) information. These observations strongly advocate the addition of a feature selection, or extraction, step in the modeling process. Such a step could help in reducing the signal-to-noise ratio, by extracting the valuable information from the different inputs while decreasing the dimensionality of these features [38] (a multimodal approach can be possibly used [39]). As already mentioned in Section IV-B, it could also help in distinguishing between the physiological variations related to emotion rather than workload.

It can be observed that the performance is very different from one subject to the other. The best scores are reached for subjects 3 and 4, and the worse for subject 1, this tendency remaining for the different input combinations that were tested. Discussing the models' performance subject per subject, particularly in light of posttest psychological analyses (semidirective interviews carried out with each subject), brings out the question of the ground truth. This question becomes of primary importance as the pressure put on the subject, hence, the sensations they experiment, becomes more real. Indeed, the two most expert pilots with simulator, real helicopter piloting,

and mission types were pilots 4 and 5. As professional pilots, they are trained to avoid overload as much as possible (it was very hard to push them to workload level 4). Subject 3 also had some basic to intermediate knowledge in both simulator and helicopter piloting and in offshore missions. Interestingly, subject 4 is an instructor, and as such, he might be better in evaluating correctly his current workload level. Subjects 1 and 2, on the contrary, neither had experience on helicopter piloting nor on offshore missions, with subject 1 being the less experimented. As a result, this subject had to put a lot of effort in mastering the simulator by itself: contrary to the other subjects, he estimated his workload to be globally higher in the first mission than in the second one. In our opinion, these learning effects not only impacted the subject's piloting performance but also the subject's self-evaluation of his workload level, particularly in this study, where no validated questionnaire, such as the NASA TLX, was used. Since we only get access to subjective evaluations of the "true" workload level, the quality of the ground truth is largely dependent on the ability of the subjects to self-assess their workload level. This makes it difficult to distinguish whether the failures come from the modeling approach by itself or from a poor quality of the ground truth. Then, it is also very difficult to make direct comparisons between different methods proposed in the literature [40].

VI. CONCLUSION

The question of estimating in real time the workload of operators involved in complex multitask operations, such as helicopter piloting, remains a big issue. This work intends to take a few steps forward in this challenging problem by addressing the question of the effectiveness of physiological, cognitive, and psychological features for estimating helicopter pilots' workload, using BNs.

By using a BN approach to workload estimation, we aimed at being able to discuss some of the main issues faced in such undertaking. First, the effectiveness of different kinds and numbers of model's inputs was addressed in constrained laboratory conditions. Second, the method and findings were challenged in a more ecological context, using an NH90 full-flight simulator.

Although we believe our method can apply to any type of vehicle piloting (cars, aircrafts, etc.), we do not pretend to propose, at this stage, a definite and universal method for real-time workload estimation. The take-home message is that extrapolation from laboratory to real operation (or, as here, top-level simulation) conditions is not straightforward. On the one hand, real (or realistic) conditions are much more demanding and generate more variable and complex behaviors than laboratory conditions. On the other hand, even highly realistic simulations face the specific difficulty of *presence* (belief of being involved in a real mission) that impacts subjects' performance and feelings. In addition, the results point out the needs for a careful optimization of the input features, using multimodal feature extraction algorithms, in order to increase the signal-to-noise ratio before the classification step itself. Specifically, this feature extraction step should try to take the utmost advantage of the complementary and redundant information in physiological inputs but should also distinguish, in these signals, the

affect-related variations from the workload-related variations. The problem of the subjects' specificity, particularly in terms of the sensations experienced and ground truth elaboration, is also discussed as an element that should not be overlooked.

These studies should be viewed as preliminary and replicated using larger sample sets notably. Their primary goal, however, was to give us some clues about the added value of different features to models' performance. We hope these results can be used as valuable guidelines for future works addressing the problem of real-time workload estimation.

ACKNOWLEDGMENT

The authors would like to thank C. Goulon and M. Huet, for their help in building the graphical dynamic scene of the laboratory experiment; C. Valot and J.-Y. Grau, for fruitful discussions; M. Gillet and the Helisim staff, for their help in defining and running the simulator experiment; the students who have collaborated on the project; and all the subjects who took part in the experiment.

REFERENCES

- [1] R. Nikolova, V. Padev, and M. Vukov, "Functional determination of the operator state in the interaction of humans with automated systems," Nat. Center of Hygiene, Med. Ecol. and Nutrition Dept. of Occup. Med. Psycho-Physiol. Lab., Sofia, Bulgaria, Tech. Rep. RTO-MP-088, Oct. 2003, RTO/NATO, St. Joseph Ottawa/Hull.
- [2] A. Norcio and J. Stanley, "Adaptive human-computer interfaces: A literature survey and perspective," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 2, pp. 399–408, Mar./Apr. 1989.
- [3] P. A. Hancock and R. Parasuraman, "Human factors and safety in the design of intelligent vehicle-highway systems (IVHS)," *J. Safety Res.*, vol. 23, no. 4, pp. 181–198, Winter 1992.
- [4] W. Liao, W. Zhang, Z. Zhu, Q. Ji, and W. D. Gray, "Toward a decision-theoretic framework for affect recognition and user assistance," *Int. J. Hum.-Comput. Studies*, vol. 64, no. 9, pp. 847–873, Sep. 2006.
- [5] C. Wu, Y. Liu, and B. Lin, "A queuing model based intelligent human-machine task allocator," *IEEE Trans. Intell. Transp.*, vol. 13, no. 3, pp. 1125–1137, Sep. 2012.
- [6] R. O'Donnell and T. F. Eggemeier, "Workload assessment methodology," in *Handbook of Perception and Human Performance*. New York, NY, USA: Wiley, 1986, pp. 42.1–42.49.
- [7] P. Hancock, *Human Mental Workload*. New York, NY, USA: Elsevier, 1988.
- [8] B. Cain, "A review of the mental workload literature," Defence Res. and Develop. Canada Toronto Human Syst. Integr. Section, Toronto, ON, Canada, Tech. Rep. RTO-TR-HFM-121-Part-II, Jul. 2007.
- [9] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals," *Amer. Psychol.*, vol. 45, no. 1, pp. 16–28, Jan. 1990.
- [10] P. Philip, P. Sagaspe, J. Taillard, C. Valtat, N. Moore, T. Akerstedt, A. Charles, and B. Bioulac, "Fatigue, sleepiness, and performance in simulated versus real driving conditions," *Sleep*, vol. 28, no. 12, pp. 1511–1516, Dec. 2005.
- [11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 1988.
- [12] D. Heckerman, "A tutorial on learning with Bayesian networks," Microsoft Research, Redmond, WA, USA, Tech. Rep., Mar. 1995.
- [13] G. Rigas, Y. Goletsis, and D. I. Fotiadis, "Real-time driver's stress event detection," *IEEE Trans. Intell. Transp.*, vol. 13, no. 1, pp. 221–234, Mar. 2012.
- [14] F. Di Nocera, M. Camilli, and M. Terenzi, "Using the distribution of eye fixations to assess pilots' mental workload," *Proc. Human Factors Ergonom. Soc. Annu. Meet.*, vol. 50, no. 1, pp. 63–65, Oct. 2006.
- [15] Y. Zhang, Y. Owechko, and J. Zhang, "Learning-based driver workload estimation," in *Computational Intelligence in Automotive Applications*, D. Prokhorov, Ed. Berlin/Heidelberg, Germany: Springer-Verlag, 2008, ser. Studies in Computational Intelligence, pp. 1–24.
- [16] F. Tango and M. Botta, "Evaluation of distraction in a driver-vehicle-environment framework: An application of different data-mining techniques," in *ICDM*, P. Perner, Ed. Berlin/Heidelberg, Germany: Springer-Verlag, 2009, pp. 176–190.
- [17] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
- [18] G. F. Wilson and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Hum. Factors*, vol. 45, no. 4, pp. 635–643, Winter 2003.
- [19] J. A. Cannon, P. A. Krokmal, R. V. Lenth, and R. Murphey, "An algorithm for online detection of temporal changes in operator cognitive state using real-time psychophysiological data," *Biomed. Signal Process.*, vol. 5, no. 3, pp. 229–236, Jul. 2010.
- [20] F. Putze, J.-P. Jarvis, and T. Schultz, "Multimodal recognition of cognitive workload for multitasking in the car," in *Proc. IEEE ICPR*, Aug. 2010, pp. 3748–3751.
- [21] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.
- [22] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," *Pattern Anal. Appl.*, vol. 9, no. 1, pp. 58–69, May 2006.
- [23] J. D. Mayer, P. Salovey, and D. R. Caruso, "Emotional intelligence: Theory, findings, and implications," *Psychol. Inquiry*, vol. 15, no. 3, pp. 197–215, Jul. 2004.
- [24] A. M. Isen, "An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications," *J. Consum. Psychol.*, vol. 11, no. 2, pp. 75–85, Jan. 2001.
- [25] N. H. Fridja, "Varieties of affect: Emotions and episodes, moods and sentiments," in *The Nature of Emotion*. New York, NY, USA: Oxford Univ. Press, 1994, pp. 59–67.
- [26] R. S. Lazarus and S. Folkman, *Stress, Appraisal, and Coping*. New York, NY, USA: Springer-Verlag, 1984.
- [27] D. Watson, L. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Pers. Soc. Psychol.*, vol. 54, no. 6, pp. 1063–1070, Jun. 1988.
- [28] P. Gaudreau, X. Sanchez, and J.-P. Blondin, "Positive and negative affective states in a performance-related setting: Testing the factorial structure of the PANAS across two samples of French-Canadian participants," *Eur. J. Psychol. Assess.*, vol. 22, no. 4, pp. 240–249, 2006.
- [29] L. Li, D. Wen, N.-N. Zheng, and L.-C. Shen, "Cognitive cars: A new frontier for ADAS research," *IEEE Trans. Intell. Transp.*, vol. 13, no. 1, pp. 395–407, Mar. 2012.
- [30] P. Besson, E. Dousset, C. Bourdin, L. Bringoux, T. Marquete, D. R. Mestre, and J.-L. Vercher, "Bayesian network classifiers inferring workload from physiological features: Compared performance," in *Proc. IEEE IV*, Alcalá de Henares, Spain, Jun. 2012, pp. 282–287.
- [31] P. Besson, C. Maïano, C. Bringoux, T. Marquete, D. R. Mestre, C. Bourdin, E. Dousset, M. Durand, and J.-L. Vercher, "Cognitive workload and affective state: A computational study using Bayesian networks," in *Proc. IEEE IS*, Sofia, Bulgaria, Sep. 2012, pp. 140–145.
- [32] S. G. Hart and L. E. Staveland, "NASA Task Load Index (TLX)," Human Performance Res. Grp. NASA Ames Res. Center, Moffett Field, CA, USA, Computerized Version v1.0.
- [33] J. L. Andreassi, *Human Behavior and Physiological Response*, 4th ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 2000.
- [34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, D. L. Schilling, Ed. Hoboken, NJ, USA: Wiley, 1991.
- [35] P. Ekman, *Basic Emotions*. Sussex, U.K.: Wiley, 1999, ch. 3, pp. 45–58.
- [36] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning," Ph.D. dissertation, Comput. Sci. Div., Univ. of California, Berkeley, CA, USA, 2002.
- [37] M. E. Dawson, A. M. Schell, D. L. Filion, J. T. Cacioppo, L. G. Tassinary, and G. L. Berntson, "The electrodermal system," in *Handbook of Psychophysiology*. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 200–223.
- [38] I. Guyon, *Feature Extraction: Foundations and Applications*. New York, NY, USA: Springer-Verlag, 2006, ser. Studies in fuzziness and soft computing.
- [39] T. Butz and J.-P. Thiran, "From error probability to information theoretic (multi-modal) signal processing," *Signal Process.*, vol. 85, no. 5, pp. 875–902, May 2005.
- [40] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

Patricia Besson was born in France, in 1977. She received the M.Sc. degree in biomedical engineering in 2001 from the University of Lyon, Lyon, France, spending the 2000–2001 academic year as an Exchange Student at the University of Montreal, Montreal, QC, Canada, and the Ph.D. degree from the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in June 2007.

Since 2007, she has been a Researcher with the Institute of Movement Sciences, Aix-Marseille University, Marseille, France. Her research focuses on understanding and modeling how human beings interact with their environment, in order to develop intelligent systems facilitating this interaction.

Christophe Bourdin was born in France, in February 1970. He received the Ph.D. degree from the Université Joseph Fourier, Grenoble, France, in October 1997.

In 1998, he joined the Institute of Movement Sciences (ISM), Aix-Marseille University, Marseille, France, working on human movement controls. Since 2010, he has been a Vice-Dean of the Faculty of Sport Sciences, Aix-Marseille University. He is also the scientific responsible for the Automotive Motion Laboratory (partnership between ISM and Peugeot-Citroen Automobile), which is dedicated to the study of self-movement perception in the automotive domain. Recently, he has developed a line of research related to the multisensory integration processes. His current research project focuses on the sensorimotor adaptation of human movements, with a main interest in studying the internal representations of the environment and of the body (internal models) constructed and updated by the central nervous system.



Lionel Bringoux was born (to be alive) in France, in 1974. He received the Ph.D. degree in sport sciences from the Université Joseph Fourier, Grenoble, France, in 2001.

He was a Postdoctoral Fellow with the “Balance and Spatial Orientation Group,” Imperial College London, London, U.K., under the supervision of Prof. A. M. Bronstein and Prof. M. A. Gresty, conducting studies about the perception of spatial reference frames under various conditions of body tilt with normal and labyrinthine-defective subjects.

Since 2003, he has been an Assistant Professor with the Faculty of Sport Sciences, Aix-Marseille University, Marseille, France, recently joining the team “Perceptual-Motor Behaviours” for studying spatial orientation and objects localization.

Erick Dousset was born in the western south of France. He received the Ph.D. degree from the University Claude Bernard Lyon I, Lyon, France, while working in the Respiratory Physiopathology Department, University of the Mediterranean, Marseille, France.

He is currently an Assistant Professor with the Institute of Movement Sciences, Aix-Marseille University, Marseille. His research topics concern neuromuscular adaptation and regulation during exercise in normal condition or after functional rehabilitation program.



Christophe Maïano was born in Cannes, France, in 1974. He received the M.Sc. degree in human movement sciences from the Université d'Aix-Marseille II, Marseille, France, in June 1999 and the Ph.D. degree in human movement sciences from the Montpellier I University, Montpellier, France, in December 2003.

From 2006 to 2011, he was an Associate Professor with the Faculty of Sports Sciences, Aix-Marseille University, Marseille, where he was a Researcher with the Institute of Movement Sciences. Since 2011,

he has been an Associate Professor with the Department of Psychoeducation and Psychology, Université du Québec en Outaouais, Gatineau, QC, Canada, where he is a Researcher with the Laboratoire de Cyberpsychologie. His research interests are centered on health (physical and mental), physical activity/exercise, and virtual reality. He is currently studying how virtual reality could be used to improve the rehabilitation of individuals with health problems or disabilities.



Tanguy Marqueste was born in Marseille, France, in 1974. He received the M.Sc. degree and the Ph.D. degree in human movement sciences from the Aix-Marseille University, Marseille, in 2000 and 2003, respectively.

He started teaching biotechnology in the late 1990s. He was a Postdoctoral Researcher with the Spinal Cord Research Center, University of Manitoba, Winnipeg, MB, Canada, and also with the Center for Magnetic Resonance in Biology and Medicine, University of the Mediterranean,

Marseille. He is currently with the Aix-Marseille University, where he is a Researcher with the Human Movement Sciences Institute in the team of neuromuscular plasticity. His current research focuses on the alteration of sensorimotor functions after neuromuscular injuries or diseases and on the use of physical activity, drugs, or biomaterials to enhance their recovery.



Daniel R. Mestre received the Ph.D. degree in psychology from Aix-Marseille University, Marseille, France, in 1987.

He is a Senior Researcher with the Institute of Movement Sciences, Aix-Marseille University, Marseille, France, where he is the Head of the Mediterranean Virtual Reality Center. His major research interests concern active visual motion perception and psychophysical and neuropsychological approaches to sensory information processing. Virtual reality is used as an experimental tool to understand our presence to the world, in both fundamental and applied domains.

Dr. Mestre is a Founding Member of the Association for Virtual Reality in France.

Sophie Gaetan was born in Ollioules, France, in February 1984. She received the Master's degree in clinical psychology and the Ph.D. degree from Aix-Marseille University, Marseille, France, in 2009 and 2012, respectively.

Since October 2011, she has been a Teaching Assistant with the Faculty of Sport Sciences, Aix-Marseille University, where she joined the Laboratory of Clinical Psychopathology: Language and Subjectivity in June 2010 and the Institute of Movement Sciences in October 2011. Her research topics concern the issue of body, emotion, and personality in a contemporary approach about behaviors, questioning not only the boundaries between the normal and the pathological (such as addiction) but also the limits of reality (such as the virtual).

Jean-Pierre Baudry received the degree in psychology from Aix-en-Provence, France, in 1995.

In 1998, he joined Eurocopter (a European aeronautic defence and space company), Marignane, France, where he is currently heading the Human Factors and Cockpit Design Department (ETZWHF).



Jean-Louis Vercher was born in 1958. He received the M.S. degree and the Ph.D. degree in physiology and neurophysiology from the Aix-Marseille University III, Marseille, France, in 1981 and 1984, respectively.

In 1987, he joined the National Centre for Scientific Research (CNRS) as an Assistant Researcher. Since 2000, he has been a Research Director with the CNRS, where he got his habilitation for research direction in 1999. The main topic of his work relates to the study of sensorimotor systems control, coordination, and plasticity. He studied this type of control, as well as adaptive mechanisms, on men and in monkeys, in aiming-tracking tasks and in gaze-orienting tasks (vestibulo-ocular reflex, saccadic and smooth-pursuit eye movements). These two modes of control were also studied in ergonomic situations. His current research relates to the processes of multisensorial integration in a context of orientation, space navigation, and visuo-guided pointing.

He studied this type of control, as well as adaptive mechanisms, on men and in monkeys, in aiming-tracking tasks and in gaze-orienting tasks (vestibulo-ocular reflex, saccadic and smooth-pursuit eye movements). These two modes of control were also studied in ergonomic situations. His current research relates to the processes of multisensorial integration in a context of orientation, space navigation, and visuo-guided pointing.