Evaluation capabilities on flight simulation systems: fusing measurements of the trainee's psychophysiological state with simulators data.

Franck Mignet, Michel Varkevisser, Gregor Pavlin, Thales Research and Technology, Delft, The Netherlands

I. Introduction

Flight simulators are routinely used nowadays for the training of pilots. Exercising using a flight simulator is recognized as a way to repeatedly expose trainee pilots to a broad range of standard tasks and exceptional situations. Still sessions on a simulator remain quite expensive and instructors make significant efforts to optimize the learning value of sessions. Critical determinants for an accurate evaluation of a trainee include actions on the system, and performance indicators that are related to the operational objectives. For the instructor, understanding the psycho-physiological state of the trainee, for instance his mental workload during a specific situation, or his vigilance, can be very valuable information to assess the performance of the trainee and adapt the training scheme. For instance the information that the trainee was mentally overloaded by the exercise enables the instructor to refine the assessment of a training session: while the trainee succeeded in executing the task, a few repetitions may be required for the exercise to become a routine and the trainee is not ready to perform more complex exercises. Increasingly, the physiological and mental processes and states of the pilot (or crew) are taken into account for evaluation purposes in addition to external observer ratings. Although observer ratings provide valuable insights concerning the trainee's actions, they can only provide qualitative and very uncertain insights concerning mental state and delicate behavioral patterns, such as eye movements. The interpretation of various psycho-physiological measurements is challenging for observers. In addition, accurate interpretation requires putting the measurements in relationship with the trained tasks, a time consuming tasks for the instructor. Systems supporting the fusion of information coming from psycho-physiological measurements and tasks requirements could improve instructor's overall evaluation capability. In this paper, we present the challenges of operationalizing psycho-physiological measurements and exercise requirements and present an approach based on information fusion enabling better decision support for the instructor and discuss the expected improvements. We conclude with an attempt to conceptualize the visualization of fused information in a so called mental state awareness panel.

II. The challenges of using psycho-physiological measurements

Knowledge of psycho-physiological measurements of the trainee opens new perspectives for a instructor. While sensors have been developed in the recent years, in practice, using the psycho-physiological measurements is not a trivial task.

Often, the operationalization of a raw measurement into a conceptual quantity requires knowledge and computation. For instance while it is thought that the cardiac rhythm is an indicator of the mental load, literature specify that the Root Means Square of the Successive Differences (RMSSD) is an appropriate quantity that is related to the mental load (Sauvet et al 2009, Cinaz et al 2010). Still, different models exist and various computation parameters play a role, for instance the size of the averaging window. Building an estimator between the sensor measurement and the psychological quantity is a complex task requiring a multidisciplinary team (medical, psychological, physics, signal processing, statistical) and possibly experimental data. In addition, the obtained estimation will be uncertain, due to inaccuracies in the measurement and imperfect estimators and models (Besson et al 2012). The instructor should be aware of this uncertainty when interpreting the input provided by a sub-system combining sensors and estimating model.

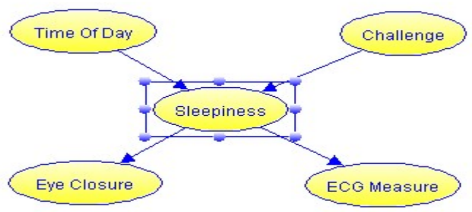
Physiological sensors are generally specialized in one aspect of the physiology (ECG, EEG, conductivity…). However the physical manifestations are not unambiguous. For instance, an accelerated cardiac rhythm can have many causes. Therefore inferring the trainee state based on a single type of physiological effect is fragile. The fragility can partially be alleviated by performing the inference based on several sensors measuring different induced physiological phenomena. Such an inference based on heterogeneous measurement can be seen as a form a fusion. While humans, and therefore instructors, are able to mentally perform such a fusion, the knowledge required makes it unrealistic to expect. Recently the approach of developing models estimating a psychological quantity based on several sensors measuring different induced physiological manifestation has gained interest (Gagnon et al 2014) providing new capabilities to the instructor. However it should be noted that it can be important for the instructor to perform a fusion of the psychological quantities themselves. For instance, a trainee showing sign of sleepiness and low mental load will draw the instructor to conclude that the exercise is too easy, while the same sleepiness, together with a high mental load would lead the instructor to think that the trainee is simply too tired to train new skills. Systems taking advantages of this higher level fusion are not used at the moment to our knowledge.

In the previous example, the critical information used by the instructor to interpret the measurements is the situation: given the exercise (or period of) at hand what is the expected mental load for the trainee? Let's assume that the exercise is seen as mentally demanding, the instructor will probably deem unlikely an exercise-induced boredom and sleepiness of the trainee. Context, such as the task difficulty, influences strongly the state of the trainee. By comparing instructor's expectations of the mental state of the trainee (based on the task at hand) with the measurements by sensors, it is possible to detect anomalies. The instructor can then interpret the anomalies and take action. In many case, the instructor will profit more of the capability to detect anomalies in the psychological quantities estimated than in the knowledge of the value of the quantity itself. Literature presents models involving the relationship between tasks and psycho-physiological state of the trainee (Neerincx 2009) but does not relate it to physiological measurements. Some authors (Lan 2002) have discussed Bayesian networks encompassing context in the large sense and measurements but not relating context to the task at hand. We have found no example in the literature combining task at hand with measurements to estimate the anomaly in the values of the state.

To summarize we have identified the following challenges for a decision support system for the instructor: detect anomalies of the psycho-physiological state, by combining heterogeneous and uncertain information at various level of abstraction, ranging from sensor measurement, interpreted measurements to information about the task to realize.

III. Bayesian network as information fusion approach

Bayesian networks are a well-established way of representing stochastic phenomena in rigorous way. Bayesian networks represent the full joint probability distribution of the stochastic variables involved in the phenomena as a directed acyclic graph (DAG) in which nodes are annotated with conditional probability distributions [Russell 2003]. Each node represents a stochastic variable relevant for the phenomenon. For our purpose we will limit ourselves to the case in which the variables are discrete, the possible values being called states. Directed edges join nodes representing the influences between the variables. Further, each node is annotated with a conditional probability table (CPT) holding the conditional probability distributions corresponding to each possible combination of the states of the parent variables. CPTs express the strength of the influence between the parent nodes and the annotated node. It can be shown that a Bayesian network, if properly designed, is a complete, non-redundant and, in most cases, very compact representation of the phenomenon (more than the full joint distribution). As an illustrative example, Figure 1 presents a simple Bayesian network, representing how the sleepiness of a trainee pilot during an exercise is influenced by the level of challenge of the training exercise and the moment of the day at which the exercise takes place (night or day, assuming a common sleepiness pattern). The sleepiness influences (among other) the eye movement of the trainee as well as his heart rate. The CPT annotating the node Sleepiness shows how the various states of the parent variable (Challenge and Time of Day) affect the probability of the sleepiness: higher if the exercise is not challenging and it is night, lower if the exercise is challenging and it is day.



Time Of Day	Night		Day	
Challenge	Difficult	Easy	Difficult	Easy
High	0.6	0.9	0.15	0.45
Low	0.4	0.1	0.85	0.55

Figure 1: A simple example of a Bayesian network modeling the relationship between task challenge, the time of day, the resulting Sleepiness and the relevant physiological measurement (Eye closure, ECG (RMSSD)).

Efficient algorithms have been developed enabling computation of probability distribution for specific aspects of the phenomenon under study given a set of observation using Bayesian network. For instance it is possible to compute the probability of each state of the variable “sleepiness” knowing that the exercise is challenging, it is day time, his eyes are closed and his ECG Measure (RMSSD) is low. Given our network parameters, the chance of state “high” sleepiness is 0.2411. Different algorithms exists that can compute exact or approximate probabilities (Jensen 2001).

For our type of applications, the Bayesian networks are models encoding uncertain knowledge about the phenomenon we want to study. To reach our goal (the estimation of the state of the trainee), the initial step of constructing the Bayesian network is required. The required knowledge can be obtained in many ways ranging from elicitation from human experts to fully automatic (machine learning) based on data. Mixed approaches are also possible. To understand approaches to model construction, it is useful to consider a Bayesian network as having 2 constituents, (i) the structure, that is, which variables are relevant and directly influence each other and (ii) the parameters, which can be seen as the strength of the influence. The choice of the approach depends on the characteristics of the problem. Using machine learning approaches to learn the structure of the model requires a very large amount of data, limiting the applicability of such approaches. In addition the variable of the resulting model can not be interpreted in terms of real aspects of the phenomenon any more. On the other side, if the expert community already possesses a significant amount of knowledge regarding the dependency between aspects of the phenomenon, the construction of efficient and interpretable models is facilitated. For instance, expert instructors can easily express which part of a training exercise will generate mental effort, stress or conversely boredom. In many cases, the model construction approach uses both the human expertise to construct the structure and machine learning to learn the parameters. In this way one can efficiently construct models that take many variables into account.

IV. Useful patterns for modeling of the trainee's state

The fusion approach using Bayesian networks exposed above shows the capability to fuse information at the required level of abstraction from sensor measurement to task to be performed. Given the complexity and large quantity of data required by structural learning, an approach using expert knowledge to construct the structure of models seems more promising. The long history of training pilots and existing knowledge regarding the skills to be developed means that domain experts can be relied on to assess the expected difficulty of tasks and exercises while medical and human factor literature can provide significant insight on the relationship between psycho-physiological quantity to be estimated and the sensor measurements. To support the modeling task, research in representing knowledge with Bayesian network provides useful elements and design patterns facilitating the development of the required models (Jensen 2001). In this section we will present relevant elements and modeling patterns.

The sensor model pattern

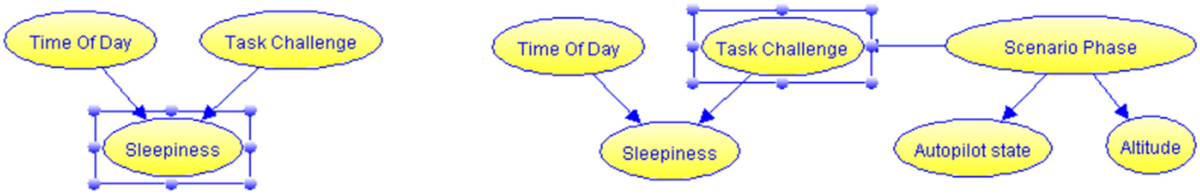
The question of representing the relationship between a quantity of interest and related sensor measurements Bayesian network has been extensively studied (Jensen 2001). A common representation is to use 2 variables, one representing the quantity to estimate and one representing the sensor measurement as a child of the first (Figure 2 (1)). The CPT represents the uncertainty in the measurement: states of the parent variable (the quantity to estimate) result in a large probability mass in the corresponding state in the child, with a small probability of the other states representing the possibility of error.

Figure 2: the sensor pattern with 1 sensor (1) or 2 sensors (2)

Figure 2 (2) shows a possible extensions to the case of multiple sensor used to determine a specific dimension of the trainee state. In this case the model will behave as a naïve Bayesian network for the determination of the Sleepiness. It should be noted that the values of the CPT actually corresponds to the quality measures of the sensor (or a combination sensor and model). For instance the values of top-right entry in the CPT actually correspond to false positive rate of the sensor. Often these values can be obtained from the sensor specification and documentation by the manufacturer so that no specific experiment is required to determine them.

The probabilistic OR pattern to model context

A number of phenomena will affect the value of the psycho-physiological state of the trainee. The primary one is, of course, the task to be executed in the exercise (for instance, its difficulty). But additional aspects can be taken into consideration: for instance the time of the day (or the proximity of a meal) can influence the sleepiness of the trainee. These phenomena have in common that they are causal factors of the psycho-physiological state of the trainee. Collectively they can be seen as the context determining the state of the trainee. While not strictly necessary, respecting the causality while creating Bayesian network model is generally a good practice, facilitating the task of domain experts that could review the model or be requested to provide insights for the parameters of the CPTs. Consequently an appropriate way to model the context is to have variables representing the various independent context phenomena as multiple parents of a single child representing the affected dimension of the psycho-physiological state of the trainee. The CPT of the child is then a form of weighted Probabilistic-OR, where the various possible causes of the state of the trainee contribute positively or negatively, based on their influence independently to the conditional probability of the state.



Time Of Day	Night		Day	
Task Challenge	Difficult	Easy	Difficult	Easy
High	0.6	0.9	0.15	0.45
Low	0.4	0.1	0.85	0.55

(1)

Scenario Phase	cruise	landing
Difficult	0.1	0.95
Easy	0.9	0.5

(2)

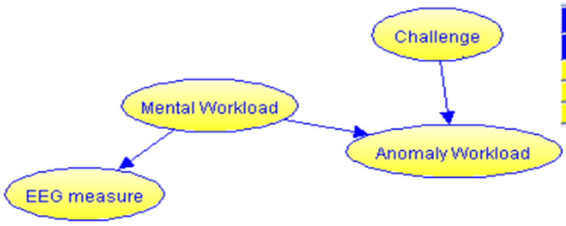
Figure 3: Modeling context (1) a simple probabilistic OR, (2) an extension for the case where the context is not directly known during the training session

Such a CPT could be obtained from proper extraction of knowledge from subject matter experts (expert instructors and pilots). Alternatively, experiments can be setup during which external observers (a Human Factor expert, a medical expert or a psychologist) assess the trainee's state, or even by self-assessment, providing ground truth. The results of the experiment (values of the trainee's state, value of each of modeled phenomena inducing this state) can be used to statistically learn the conditional probability.

In the model of Figure 3 (1), the values of the variable representing the phenomena related to the task at hand need to be provided to the fusion system. While it is possible to envision a system in which the instructor would provide in advanced such values for the entire duration of the exercise, that approach would lack robustness if the exercise allows for timing deviation during the session. To overcome this issue, one can represent the difficulty of the task at hand as a hidden variable influenced by a variable representing a specific phase of the flight that can be inferred from observation from the flight simulator (for instance, altitude, plane motion…) (Figure 3 (2)). For instance, in an exercise simulating a flight with an airliner, the flying phase "cruise" could be inferred from the altitude and autopilot state. Adding such a model fragment would not diminish the ability of domain experts to assess the conditional probability of the trainee's state for the various state of the variable phase: for instance, sleepiness is more likely during a "cruise" phase of an exercise than during the "landing" phase of an exercise.

A derivation of the XOR pattern to model anomaly

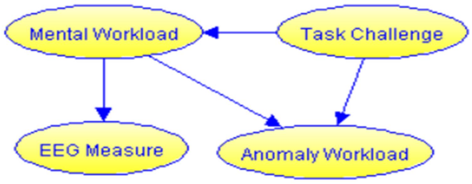
As suggested in section 2, the instructor will more often base his decision on his perception of an anomaly of the trainee state, than on the trainee state itself. Therefore it is useful to include in the model a variable expressing the anomaly: is the inferred value of the trainee state normal or not, given the task at hand? While numerous problems require the estimation of the degree of anomaly of a given quantity, there is no agreement on a pattern to represent it. Instead, existing approaches use specific measures to estimate the inconsistencies (Jensen 2001). In this section we introduce a pattern to represent anomalies. An anomaly can be formulated as a disagreement between an expected value and an actual value. In our example, the Task Challenge represents the expected value for the workload, whereas the Mental Workload is the actual value (inferred from the sensor measurement). Following the concepts behind the Probabilistic-OR pattern, the literature proposes the Probabilistic-XOR pattern (Jurgelenaite 2005). The common logical function XOR (or EXOR, or parity function) expresses the notions of agreement / disagreement: it returns value 1 (or true) when both inputs are identical (“agreement”), value 0 (or false) when inputs are different (“disagreement”). The Probabilistic-OR is a variant thereof. In our case, we will want to use the state “normal” as the state resulting from an agreement between the parent nodes and the state “Anomaly” as the state resulting from a disagreement. For the richness of interpretation (explained further in section VI) we introduce two anomalous states, both expressing disagreement: “Anomaly Low” represent the situation where the actual workload is lower than expected and “Anomaly High” the situation where the actual workload is higher than expected. Applying the XOR pattern to our problem would give us the scheme presented in the table in Figure 4:



Task Challenge	Difficult		Easy	
Mental Workload	High	Low	High	Low
Anomaly_High	0.15	0.05	0.75	0.05
normal	0.8	0.15	0.2	0.75
Anomaly_Low	0.05	0.8	0.05	0.2

Figure 4: Probabilistic XOR gate to model anomaly

The entry for the state “normal” of the CPT of such a probabilistic XOR gate represents a high probability mass when the expected and actual value agrees and low when they disagree. However a careful analysis of the influences between variables in the above network shows a problematic aspect: since the variable representing our expectations is supposed to summarize (or be influence by) the objective context and the task at hand, it seems natural to think that it will influence the actual state of the trainee. In the case of Mental Workload, it seems natural that the context will influence the cognitive task load which in turn will influence the state of the trainee. The contrary would lead to the bizarre conclusion that the task at hand does not influence the Mental Workload of the trainee. Consequently, there should be a link from the variable representing the expected value to the inferred value. The structure becomes such as presented in Figure 5.



Task Challenge	Difficult		Easy	
Mental Workload	High	Low	High	Low
Anomaly_High	0.15	0.05	0.75	0.05
normal	0.8	0.15	0.2	0.75
Anomaly_Low	0.05	0.8	0.05	0.2

Task Challenge	Difficult	Easy
High	0.75	0.3
Low	0.25	0.7

Mental Workload	High	Low
High	0.96	0.1
Low	0.04	0.9

Figure 5: Anomaly pattern derived from the XOR gate, respecting the influence of Task difficulty on Mental Workload

Complete model

The patterns discussed in the previous sections can be combined to form a Bayesian network representing the phenomenon. Figure 6 presents such a model enabling the estimation of the mental workload of the trainee and its degree of anomaly given the task to realize.

Figure 6: A possible model for the estimation of the Mental Workload of the trainee, including sensor model, context and anomaly pattern.

The mental workload of the trainee is represented by the variable “Mental Workload” while its degree of anomaly is represented by the variable “Anomaly Workload”. The bottom left part of the model uses the sensor model pattern for the fusion of several sensor measurements of the mental workload enabling more robust inference for the estimation of the mental workload. The right part represents the inclusion of the context, “scenario phases” that can be inferred using information from the Flight Management system. Similar models have been developed for other dimension of the trainee’s mental state, such as the vigilance.

V. Experimental results using the sensor model and context patterns

During the European FP7 project ACROSS, a model similar to the one presented in the previous section (Figure 6) was developed to model the vigilance of the trainee using the sensor model pattern (Section IV, “The sensor model pattern”) and the context pattern (Section IV, “The probabilistic OR pattern to model context”), but not the anomaly pattern. The model was tested during experiments with pilots in a simulator. While the number of experiments was not large enough to provide statistically significant quality measure of the model, the robustness features expected from the information fusion were observed in a number of cases. Figure 7 presents an experiment during which the sensor dedicated to drowsiness stays fixed to a value indicating no drowsiness, a possible malfunction of the sensor. However the fusion of this measurement with cardiac rhythms measurements, and infra-red encephalographic measurements, both indirect indicators of sleepiness together with the requirements of the task (a long period of high altitude cruise mode flying) resulted in the system indicating some period of lower vigilance. The occurrence of period of lower vigilance during that phase of the experiment was confirmed a-posteriori by the trainee. The system here showed its capability to improve the true positive detection rate.

Figure 7: example of improved detection of true positive

During another experiment (Figure 8), the sensor indicated a number of possible drowsiness event. Another sensor indicated significant mental activity. The task at hand, a difficult go-around including partial failure of flying devices, required significant mental activity and attention. The system displayed a desirable behavior of indicating normal vigilance thus of reducing the false positive rate.

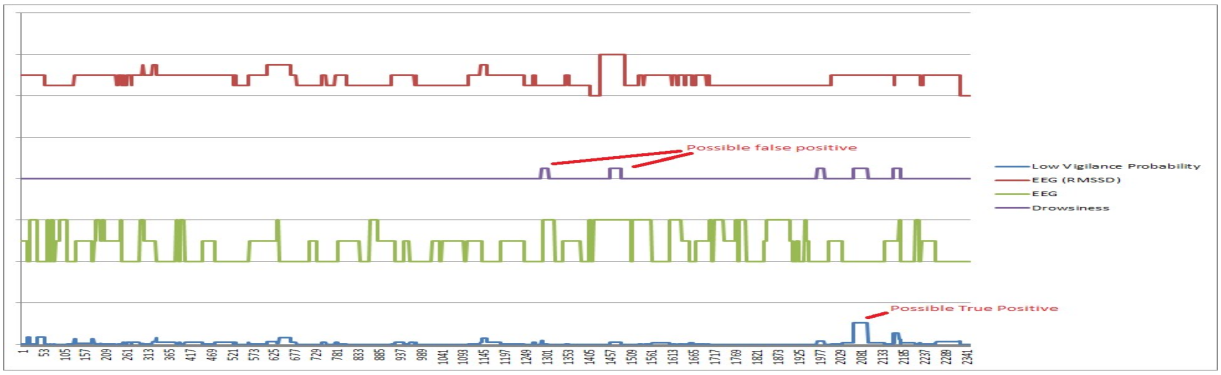


Figure 8: Example of true detection of false positive

VI. Synthetic examples using the anomaly pattern

Synthetic experiments can help understanding the behavior of pattern used to model anomaly. For the experiment below, the model Figure 5 can be used, representing some of the phenomena influencing the mental workload of the trainee. It should be noted that the behavior will strongly depend on the strength of the relationships between the task at hand (context) and the (not observed) mental workload as well as between the (not observed) mental workload and the sensor measurement (which is linked to the sensor quality). Figure 9 represents the probabilities of the states of the Anomaly variable for different phases of the training and simulating different measurements from the sensor. In the first phase, the exercise is easy and the EEG sensor measures a low workload. The state “normal” of the anomaly variable indicates a very high chance. In the following phase the exercise becomes difficult and the sensor measures high workload. Again, no anomaly is detected by the system. In the following phase the task is still deemed difficult but for a short period of time the sensor indicates a low mental activity. As expected, the system indicates a high chance for the state "Anomaly Low". Based on this the instructor attention should be triggered. Several interpretations are possible. One is that the sensor is malfunctioning. Another is that the training task is actually easier than the instructor indicated to the system (or, if it is inferred from FMS indication, possibly that the trainee is using a non-standard approach to solve the problem). Finally the mental workload of the trainee could be low because he hasn't grasped the seriousness of the problem he has to solve (inexact situation awareness), likely annunciating mistakes. After a long phase of difficult tasks, the training continues with an easy phase. However, let's assume the sensor is still indicating high mental workload of the trainee. The system will attract the attention of the instructor with a significant chance for the state "anomaly high". The instructor should attempt to interpret the anomaly: malfunctioning sensor, inaccurately estimated difficulty of the task, or possibly the trainee is too tired to cope with even an easy task. The instructor could try to confirm or infirm one of the above hypothesis by complementary observation before possibly deciding to end the session. The example above shows how the instructor's attention can be triggered and how the system supports the instructor in assessing and optimizing his session.

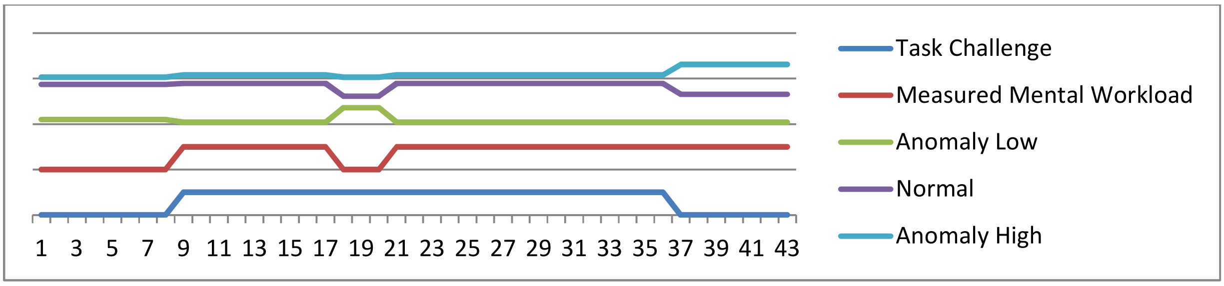


Figure 9: Synthetic experiment on the anomaly pattern applied to mental workload

VII. Visualization

The instructors' main task is to provide pilots with flight instructions in compliance with regulations, policies, procedures, and techniques. Learning analytics is an approach that can support in improving the overall learning experience by introducing tailored technologies to support instructors and trainees (Martinez-Maldonado et al., 2010). Through technology enhanced learning, pilot instructors are provided with detailed briefings on all phases of flight during simulator training, determining and reporting on trainee progress and proficiency. If a trainee fails in part of the training, specialized instructions and counseling are required. In this section we introduce visualizations for instructors to become aware of the pilot's mental state. As discussed in previous sections, this is determined by mental (over)load of trainees in combination with their situational awareness and poor subsequent decision-making (Vidulich, et al., 2010). The quality of decisions follows from the pilot's actions, procedures, and communication, which are beyond the scope of the current paper. In essence, the instructor should have a quick overview of the pilot's current situational awareness (SA), and mental state, given the flight phase he is in. To visualize these, we introduce the Mental State Awareness Panel (MSAP). The visualizations are based on the fusion algorithm outcomes as discussed in the previous sections.

According to the learning analytics principles, instructors have to deal with constraints and contingencies in the learning environment (Martinez-Maldonado et al., 2010). Thus, in complex training environments technologies should be tailored towards optimized awareness and decision support of the instructor. Probabilities as introduced in the previous sections are likely to be too abstract for instructors and need to be translated to meaningful information elements for the instructor. The MSAP concept adheres to the following design principles: (i) easy to use, (ii) easy to interpret the results, (iii) providing decision support, (iv) suitable for the task/learning environment, (v) self-descriptive. The global layout of the MSAP modules is depicted in Figure 10. The top part of the screen is dedicated to flight details, real-time/replay functions. The notifications are the direct result of the 'fusion outcomes and the 'Graph" module shows the fusion outcomes as a function over time to detect trends. The 'View + Notes' module depicts a view of the pilot (e.g. camera or eye tracking). Additionally, the instructor can add notes (free text or predefined) at any time, or attach notes to a specific notification.

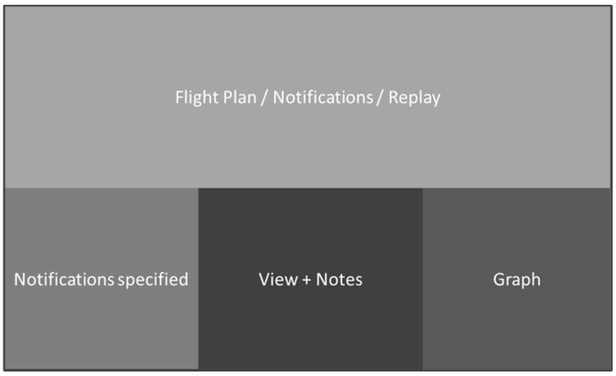


Figure 10: Global layout of MSAP.

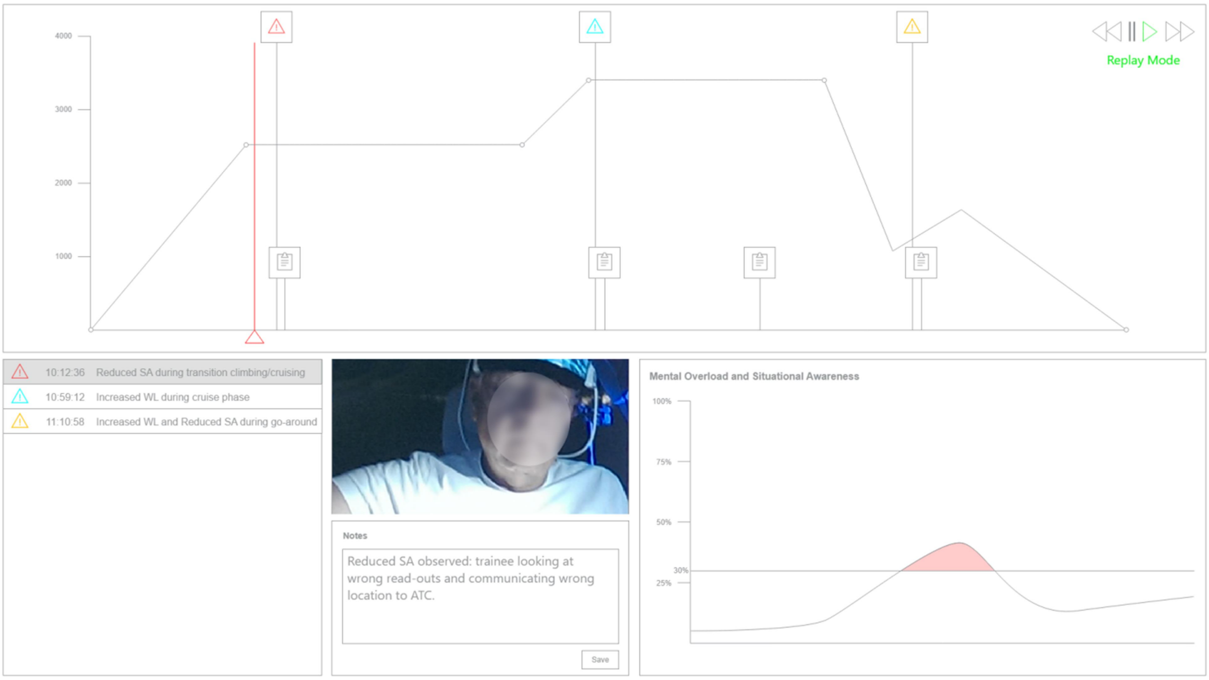


Figure 11: The MSAP concept showing the flight plan and current location of replay (red line). The indicator of the current location can be dragged to any specific location in replay mode. During simulated flight it indicates the 'now' of the simulation. The notifications are depicted on the flight plan and are specified in the lower left box (time + specified). Three examples are given for red, cyan, and amber notifications respectively. In the lower middle box the camera (or eye tracker, etc.) can be viewed and the instructor can add notes to the various moments in time: each time he saves notes, an indicator is shown on the flight plan. The lower right box depicts the mental overload and SA probabilities over time.

For the HMI concept (see Figure 11), the following design aspects were taken into account: (i) Results related to flight phase, (ii) real-time awareness of pilot's state, (iii) trends in mental workload, (iv) replay of flight. The notifications are depicted in cyan (notification), amber (warning), and red (alert), and are compliant to operational standards, mapping to the operational 'severity' and subsequent procedural steps. This provides the instructor with an estimate of the level of operational mental workload. Cyan notifications are allowed to be passed during training, yet will focus the attention of the instructor on situational awareness and decisions made by the trainee. Amber and red notifications would serve as thresholds that pilots during training should typically never reach. In case of passing these thresholds, the instructor might decide to train them on coping with task-related stress, given the simulated circumstances, e.g. a certain flight phase. Note that currently such operational thresholds do not exist for pilots' mental workload. We assume that these will be established in the near future. Note also that a similar approach can be taken for vigilance, with the colored notifications indicating loss of vigilance. In fact, both fusion models could be integrated in MSAP. The instructor would then have the possibility to train both highly complex scenarios and very boring ones to monitor the trainee's responses to these (very common) situations.

VIII. Conclusion

In this paper we discussed the challenges associated with the use of psycho-physiological measurements to improve the quality and efficiency of the assessment of trainee pilot during their exercises on simulators. We show how systems including information fusion approaches such as Bayesian networks, can support the instructor in dealing with some of these difficulties, namely difficulties associated with the uncertainty related to the sensor measurements and the interpretation of the measurements in relationship with the training objectives. Bayesian networks can model the phenomena influencing the trainee's state as well as the sensor inaccuracies when using the right patterns. The introduction of task related information in the model is particularly critical for improvement of the detection of anomalous state. It enables better detection even if the system can rely on only 1 type of sensor or in approaches where sensors measuring several physiological aspects are aggregated in an earlier stage. In addition, an advanced pattern presented in this paper can support the detection of anomalies in the trainee's state given information about the exercise. Bayesian networks supports construction approaches mixing expert knowledge and learning from data. This means that the large body of knowledge from instructors, pilot and medical doctors can be leveraged for the construction of models, limiting the amount of data that has to be collected. Still some parameters may be difficult to obtain from experts in which case well known machine learning algorithms are available to learn parts of the model. Such models were used in the context of the European FP7 project ACROSS. While the amount of data collected does not allow for statistically significant evaluation, observed behavior of the model showed the expected improvements in detecting anomalous state of the trainee while keeping false positive low. Finally we presented a visualization of the outputs of the information fusion component capable of representing the states and processed sensor outputs in a meaningful way to the observer/instructor. Visualization of fusion information is by no means trivial. Only experts in the fields of fusion and, in this case psycho-physiology, could make sense of what the fused outcomes actually mean. Thus, for each domain of application a translation needs to be made for the operators that have to understand the data and have to act upon it. In this case, the pilot instructor would require extensive training to understand the fused data. In complex and expensive simulation environments the outcomes should be easy to interpret and easy to use. Note that the principles for visualization of the fused information outcomes should be integrated in a larger instructor tool or console to support the instructor's main task of providing flight instructions, i.e. also including performance and procedures indicators. In addition to training, such a fusion approach could be integrated in future Crew Monitoring Systems (CMS). These could be integrated in planes in order to adapt cockpits to the state and behavior of pilots during their flight and contribute to safety improvements as well as management of the pilots' mental state. The latter could entail counteracting drowsiness as well as reducing mental overload. Experiments including the fusion approach described in the session have been conducted during the FP7 ACROSS project, aiming at developing cockpit applications and Human-Machine interfaces covering all safety related crew duties. The introduction of such practices during training could constitute a way to gather large amounts of data in a consistent way over time, including intra- and inter-personal variability and enabling the construction of more reliable models required by in-flight applications.

References

Besson, P.,Dousset, E.,Bourdin C.,Bringoux L., Marqueste, T.,Mestre, D. R., and Vercher, J.-L. (2012) “Bayesian Network classifiers inferring workload from physiological features: compared performance”, in Proceeding Intelligent Vehicles Symposium, IEEE IV, Alcalá de Henares, Spain, Jun. 2012, pp. 282–287

Cinaz, B., La Marca, R., Arnrich, B. and Tröster, G. (2010) “Monitoring Of Mental Workload Levels”, IADIS International Conference e-Health 2010.

Gagnon, J-F., Lafond, D., and Rivest, M. (2014) "Sensor-Hub: A Real-Time Data Integration and Processing Nexus for Adaptive C2 Systems." (2014): 63-67.

Jensen, F. V. (2001) “Bayesian Networks and Decision Graphs”. Springer-Verlag, New York, 2001.

Jurgelenaite , R., Lucas, P. J. F. (2005) “Exploiting causal independence in large Bayesian networks”, Knowledge-Based Systems 18 (2005) 153–162

Lan, P., Qiang, J., Looney, G., (2002) “Information Fusion with Bayesian Networks for Monitoring Human Fatigue”. ISIF 2002, pp.535-541

Neerincx M.A., Kennedie S., Grootjen M. and Grootjen F. (2009) “Modeling the Cognitive Task Load and Performance of Naval Operators” In: Schmorrow D.D., Estabrooke I.V., Grootjen M. (eds) Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. FAC 2009. Lecture Notes in Computer Science, vol 5638. Springer, Berlin, Heidelberg

Martinez-Maldonado, R, Pardo, A., Mirriahi, N., Yacef, K., Kay, J., and Clayphan, A. (2010). The LATUX Workflow: Designing and Deploying Awareness Tools in Technology-Enabled Learning Settings. Proceedings of the 5th Int. Conf. on Learning Analytics And Knowledge, Poughkeepsie, New York.

Russel, S. and Norvig, P. (2003) “Artificial Intelligence: A modern approach (2nd Edition)”, Prentice Hall Series in Artificial Intelligence, Prentice Hall, Englewood Cliffs, NJ, 2003

Sauvet, F., Jouanin , J. C., Langrume, C., Van Beers, P., Papelier Y., and Dussault, C. (2009) “Heart Rate Variability in Novice Pilots During and After a Multi-Leg Cross-Country Flight”, Aviation, Space, and Environmental Medicine Vol. 80, No. 10, October 2009

Vidulich, M.A., Wickens, C.D., Tsang, P.S., and Flach, J.M. (2010). Information Processing in aviation. In: Human Factors in Aviation, 2nd edition, Salas, E., and Maurino, D. (Eds.). Elsevier, Academic Press, pp.175-215.