



Psycho-Physiological Evaluation of the Pilot: A Study Conducted with Pilots of the French Air Force

Vincent Ferrari¹, Jean-François Gagnon^{2(✉)}, Cyril Camachon¹, and Maëlle Kopf³

¹ Air Force Research Center, Salon-de-Provence, France
vincent.ferrari@ecole-air.fr

² Thales Canada Inc. Research and Technology, Québec, Canada
jean-francois.gagnon@ca.thalesgroup.com

³ Thales AVS, Osny, France

Abstract. The present paper concerns the individualization of the training of aircraft pilots. Specifically, it presents the data collection, and modeling efforts carried out to assess trainees' transition from a controlled, effortful piloting experience (i.e., System 2), to an automatic, effortless process (i.e., System 1). It is argued that cardiovascular activity can be associated with deployment of effort, and therefore be used for assessing the transition across systems. Heart rate, respiration rate, and heart rate variability were sampled on 11 pilots (5 students "novice" and 6 instructors "experts"), performing 6 one-hour flights (5 flights in tandem: one student and one instructor, the 6th flight with an instructor flying alone). These data were used for the development of a prediction model computing the probability of a pilot being an expert or a novice. After a "leave-one-tandem-out" validation, the accuracy of the model was 86.86%. The results are discussed in terms of effortful processes and skill acquisition. Further work will consist in implementing contextual parameters in the model in order to improve the prediction. Such a model could be used by instructors and trainees as a supporting tool for tracking progress of the training at the individual level.

Keywords: Psycho-physiology · Heart rate · Heart rate variability · Ease of flight Training · Quantification of learning

1 Introduction

In a complex and dynamic environment such as piloting a fighter plane, a pilot must constantly be prepared to react to unexpected situations, and engage additional cognitive resources for carrying out his mission. Hence, an important part of military pilot training is to appreciate the unpredictable nature of the mission (Fornette et al. 2015). This educational approach is all the more demanding since it begins early in the formation of the trainee and should allow him to perform well in spite of the uncertainty of the operating environment.

For the instructor, identifying the specific moment to go from an expected to an unexpected situation is crucial, as the change must only be operated once the trainee has acquired the fundamentals of the flight, that is to say a set of knowledge devoted to maintaining the aircraft in a safe area (i.e. maintain altitude, heading, etc.). If unexpected

changes are brought too early in the training, the trainee will not be able to fully integrate the basis, and will not reach the optimal and target state which is referred to as “ease in flight”. Indeed, excessive demand on resources imposed by the attended task(s) typically results in performance degradation (Nourbakhsh et al. 2013; Stanton et al. 2005). On the contrary, a delayed addition of unexpected situations will have no benefit and inefficiently extend the training time. Currently, instructors exclusively rely on their experience and subjective observations to detect this key moment.

From a cognitive point of view, ease of flight could be associated with automatic processes (i.e., System 1) as opposed to controlled processes (i.e., System 2). Over the last decades, this multiple system theory of decision making has been widely studied and has accumulated a large body of evidence (see Sanfey and Chang 2008 for a brief review). System 1 has been described as fast, effortless, and unconscious whereas system 2 has been depicted as slow, effortful, and conscious.

Skill acquisition can be viewed as a shift from system 2 to system 1. Kahneman (2003) has even linked System 1 to “intuition”, frequently associated with how experts make decisions (e.g., Dreyfus 2014 [in Zsombok and Klein]).

Automated processes require very little cognitive resources, as opposed to controlled processes. An expert pilot has automated the majority of recurrent piloting tasks and procedures (e.g., take-off), which, for him, do not require an important engagement of cognitive resources. In comparison, a novice who has not fully automated the procedures will have to spend more energy to reach a similar performance. In a systemic view of the phenomenon, this difference in terms of energetic cost is expected to have physiological corollaries, in particular cardiorespiratory, which can be used as indicators of energetic spending. Indeed, several physiological corollaries of cognitive efforts have been identified over the last decades, including in piloting tasks (Roscoe 1992).

For instance, increase in heart rate (HR) has been associated with effort, cognitive (e.g., Kennedy and Scholey 2000) and physical. Notably, it was used by Dahlstrom and Nahlinder (2009) to estimate mental workload for pilots in simulators and in-flight. It has great potential for in-flight mental workload estimation because it is easily obtained, and less subject to noise than other typically used measures, like electro-encephalogram. HR variability (HRV) refers to the regularity of consecutive R-R intervals of the QRS complex as measured by an electro-cardiogram (ECG). Although not as intuitive as HR, HRV is one of the most frequently used metric associated with mental effort, both in fundamental and applied research. For instance, HRV was associated with mental overload in a simulated piloting task (Durantin et al. 2014), and with several fundamental neuro-cognitive tasks (Gagnon et al. 2016). Finally, respiration rate (RR) has been linked with energetic spending, has been considered a measure of task demands (Overbeek et al. 2014) and was also associated with negative valance and arousal (Masa et al. 2003).

In the context of air force pilot training, we hypothesize that during identical flights, the trainees will have to deploy a greater amount of mental effort than instructors for reaching similar performances. Therefore, trainees should exhibit a specific pattern of physiological parameters: HR and RR should be higher, and HRV should be lower than for instructors. Based on this premise, we hypothesize that it is possible to predict the role of the pilot (trainee or instructor) using physiological measures.

Moreover, the use of physiological measures could allow the identification of “expected pattern” among experienced pilots, which would be used as references when considering the same metrics among trainees in identical situations. The variation of the difference between the expected pattern (expert) and the observation (trainee) could be interpreted as a consequence of the levels of cognitive automation of the processes in the given situation for a trainee. Hence, this paper considers the possibility of quantifying learning by comparing his metrics to the reference measured on his instructor.

1.1 Objectives

The main goal of the present paper is to open the way towards an objective measure of “ease in flight”, which would assist instructors and their students during training. Such an objective measure would be a key element in the process of individualizing the training of pilots. As learning skills have a great variability between trainees, objectively quantifying to which degree a student easily performs a task could allow a great improvement in the training. Specifically, this paper is organized around two objectives, described below.

Objective 1

The first objective is to assess the impact of roles and flight phases on physiological measures. Specifically, three variation of the main hypothesis are formulated:

- H1. Mean physiological values will differ across roles
 - H1a. Mean heart rate will be higher for trainees when compared with instructors
 - H1b. Mean heart rate variability will be lower for trainees when compared with instructors
 - H1c. Breathing rate should be higher for trainees when compared with instructors

Objective 2

The second objective is to develop a model for predicting the level of expertise based on physiological measures. The physiological predictors will be comprised of statistically significant predictors that varied across roles. The model will be applied to the physiological measures and predicted expertise will be assessed. This model assumes that instructors have greater expertise than trainees.

The goal is to evaluate if such a model could help dynamically (1) quantify the progression of training and (2) identify periods of time where the instructor might not be fully in control of the flight.

2 Method

Eleven pilot participants were equipped with a Zephyr Bio Harness 3.0 chest strap measuring the electrical activity of the heart (ECG), RR, and accelerations on 3-axis. They were also equipped with an Android mobile phone on which the Sensor Hub (Gagnon et al. 2016) application was installed. The application integrates all generated data, processes HR, HRV (frequency and temporal domain), accelerations (3-axis), respiration rate, and global positioning system coordinates.

Participants were organized in tandems consisting of an instructor (assumed expert) and a trainee (assumed novice). The data were collected on five comparable aerobatic flights with trainees of approximately the same skill level. One of the flights was performed by an instructor flying alone. Each flight was broken down into five phases: pre-flight (briefing), take-off, flight, landing, and post-flight (debriefing).

During the flight, instructors performed specific maneuvers that the trainees had to perform immediately after, therefore transferring the control of the plane from one to another. Instructors were responsible for take-off and landing.

3 Results

Results are described in two sub-sections, aligned with the objectives. First the statistical significance tests are reported to evaluate the impact of the key factors (role and phases) on individual physiological measures. Second, a classifier of expertise is developed and described.

3.1 Factors Influencing Physiological Parameters

Three mixed ANOVAs were carried out to test the effect of the role (Trainee vs Instructor), phase (repeated 5 levels), and their interaction on (1) mean HR in bpm, (2) mean HRV in ms, and (3) mean RR in bpm.

Hypothesis H1a

Results show that both role $F(1,6) = 9.27, p < .05$ and phase $F(4,30) = 14.51, p < .001$ had a statistically significant impact on mean HR in bpm. Interaction of role and phase was not statistically significant $F(4,30) = 1.81, N.S$. In line with hypothesis H1a, mean HR in bpm is statistically higher in the trainee condition (mean = 113.56, sd = 22.73) when compared with the instructor condition (mean = 74.82, sd = 11.21). Results are presented in Fig. 1.

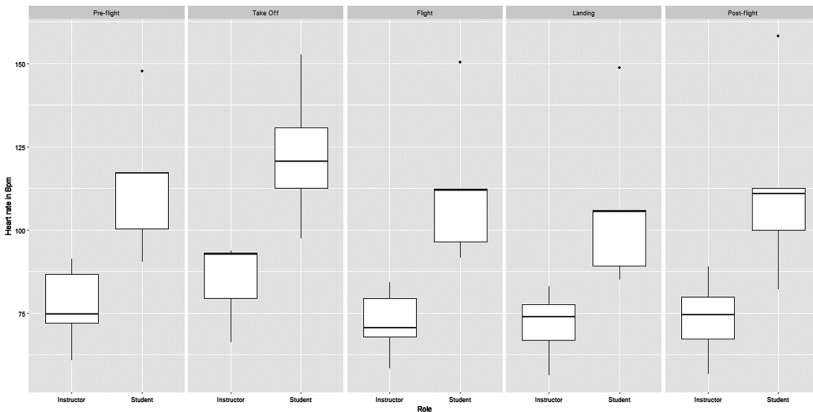


Fig. 1. Mean HR in bpm by role and phase.

Hypothesis H1b

Results show that both role $F(1,6) = 11.30, p < .05$ and phase $F(4,30) = 3.50, p < .05$ had a statistically significant impact on mean HRV in ms. Interaction of role and phase was not statistically significant $F(4,30) = 1.77, N.S.$ In line with hypothesis H1b, mean HRV in ms is statistically lower in the student condition (mean = 36.58, sd = 16.93) when compared with the instructor condition (mean = 65.66, sd = 17.80). Results are presented in Fig. 2.

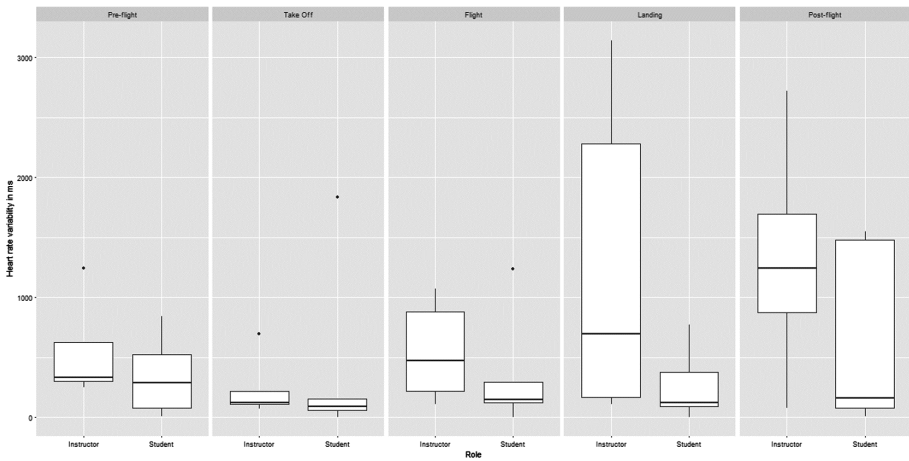


Fig. 2. Mean HRV by role and phase.

Hypothesis H1c

Results show that phase $F(4,30) = 5.48, p < .001$ had a statistically significant impact on mean RR in bpm. Role did not have a significant impact $F(1,6) = 1.11, N.S.$ However interaction of role and phase was statistically significant $F(4,30) = 4.39, p < .01$. Unsupportive of hypothesis H1c, mean RR in bpm is not statistically higher in the student condition (mean = 19.80, sd = 2.21) when compared with the instructor condition (mean = 18.23, sd = 3.26), but there is a significant interaction of the two factors $F(4,30) = 4.66, p < .01$ on respiration rate. Results are presented in Fig. 3.

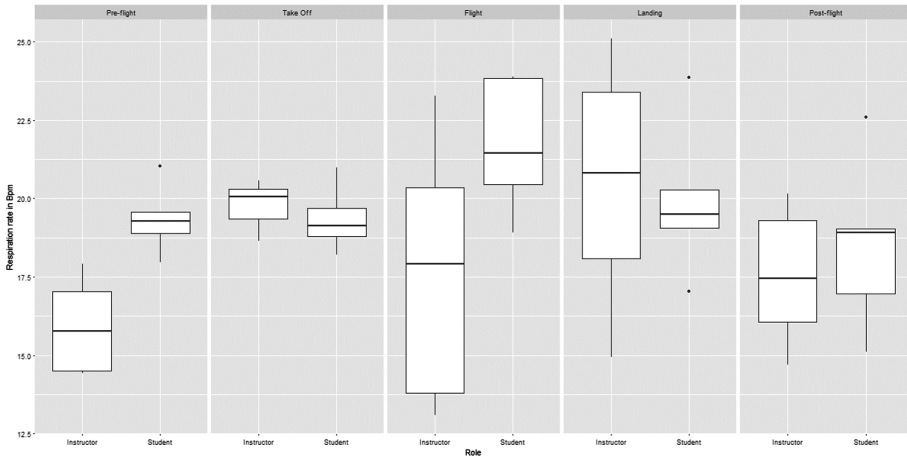


Fig. 3. Mean respiration rate in bpm by role and phase.

3.2 Modeling Effort Linked with Expertise

In addition to statistical significance tests, an integrated model was developed to predict the role of the participant based on HR, HRV, and RR as predictors. In an attempt to remain parsimonious and explainable, the generalized linear model (GLM) was employed. However, rather than using phases as temporal separations, equal non-overlapping bins of 10 min were created. For each of these bins, mean HR in bpm, mean HRV in ms and mean RR in bpm were calculated. The model was developed using this data. The reason for the creation of bins is that the flight phases are highly variable in terms of length and would therefore induce a bias in the statistical representativeness of metrics within the model. For instance, a very short phase of two minutes would have the same statistical weight than a phase lasting 45 min.

The model was validated for generalization using a “leave-one-tandem-out” procedure. The final model used for predictions was retrained on all the data.

Results show that the model achieved an accuracy of 86.86% (95% confidence interval = 80.03–92.02), $\chi = .74$. The predictors (and associated betas β) are represented in order of relative influence in Table 1.

Table 1. Model predictors and associated β .

Predictor	β
Heart rate in bpm	-2.6211
Heart rate variability in ms	1.1584
Breathing rate in bpm	0.4858
Intercept	-0.1214

The model was then applied to each individual data to see how the predictions unfold in time during a flight. The numeric prediction represents the probability that the

observed physiological pattern (composed of HR, HRV, and RR) is the one of an instructor. Hence, when the probability exceeds 50%, the point is classified as “instructor”, and conversely when below 50%. By showing the predicted probability, we can track changes in the progression of each individual. We plotted the predictions of two tandems that were deemed interesting for discussion. Tandem 1 model predictions were plotted in Fig. 4, and tandem 5 in Fig. 7. Alongside the predictions of the model, we plotted the most influencing factor of the model (i.e., heart rate in bpm), and altitude in meters to provide some context.

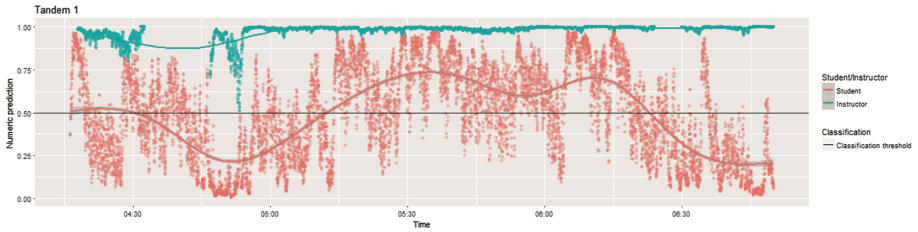


Fig. 4. Tandem 1 - Probability of being an instructor according to the model, by role for the whole flight. The classification threshold corresponds to the point where the most probable classification changes from one role to another. The points over the horizontal line (Classification threshold, 0.5) represent the data which were classified as being those of an instructor.

Tandem 1 (Figs. 4, 5 and 6) shows that the instructor was classified as an instructor all the time. Interestingly, results show that the student was above the 50% threshold (so classified as an instructor) for a long period of the flight, but still had punctual states corresponding to the typical state of a “trainee”.

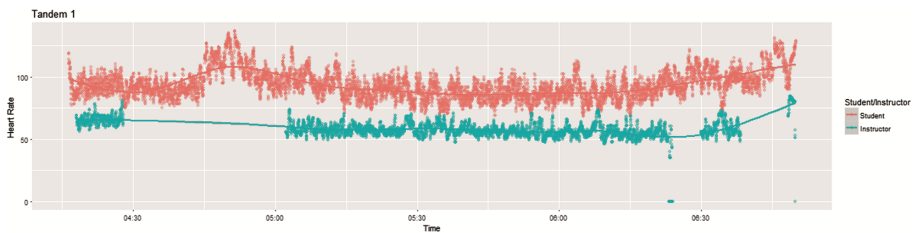


Fig. 5. Tandem 1 - HR in bpm sampling values through the flight. The predictions made by the model are largely based on this metric.

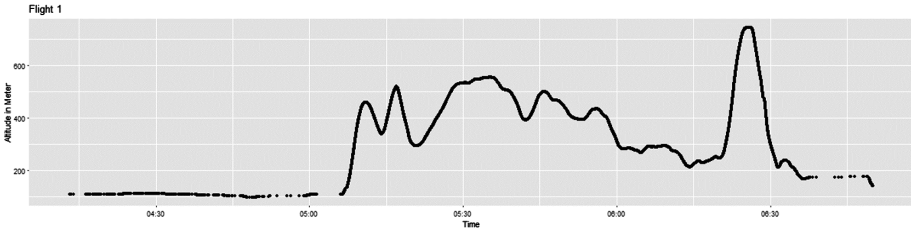


Fig. 6. Tandem 1 - Altitude in meters of the aircraft.

Tandem 5 (Figs. 7, 8 and 9) resulted in a much different pattern than the previous tandem. First, it is observed that the instructor is not classified with as much confidence as instructor from Tandem 1. Punctually, the probability of being an instructor even falls below the 50% threshold. On the other hand, the data shows a progression of the trainee from trainee to instructor as the flight progresses.

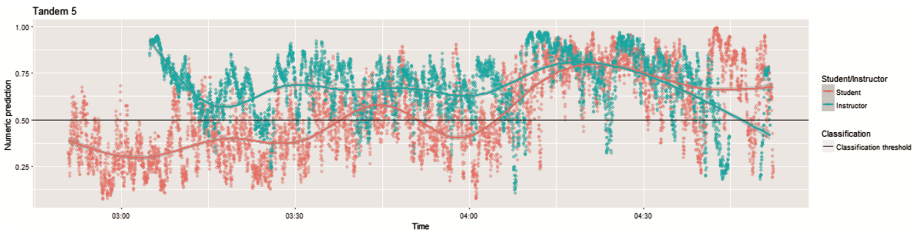


Fig. 7. Tandem 5 - Probability of being an instructor according to the model, by role for the whole flight. The classification threshold corresponds to the point where the most probable classification changes from one role to another. The points over the horizontal line (Classification threshold, 0.5) represent the data which were classified as being those of an instructor.

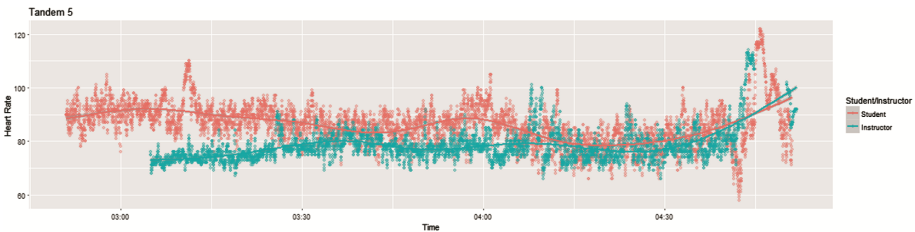


Fig. 8. Tandem 5 – HR in bpm sampling values through the flight. The predictions made by the model are largely based on this metric.

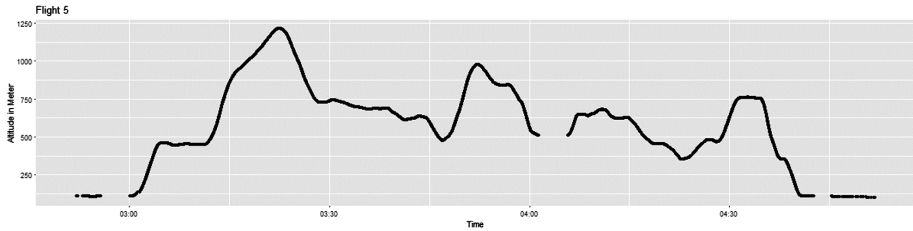


Fig. 9. Tandem 5 - Altitude in meters of the aircraft.

4 Discussion

Results regarding HR and HRV supported both hypotheses concerning the relationship between physiological parameters and roles (H1a, H1b). Indeed, as expected, mean HRV in ms was lower for the trainees when compared with instructors, and conversely for mean HR in bpm. These findings support the assumption that expertise is associated with effortless processes. This is not surprising, but not trivial either as the effects of flight dynamics (especially in aerobic flight) on physiological parameters are still largely unknown. Because aerobatic maneuvers probably require a greater deployment of physical effort when compared with regular flights, the effects associated with aerobatic flight might have prevented the effects associated with cognitive effort deployment from being observed. Fortunately, the results show that roles had a statistically significant impact on physiological parameters.

Results regarding all three variables suggested that flight phases have a significant effect on physiological parameters. The results obtained also highlighted a significant effect of the interaction of the role and phase on the RR in bpm. These results, again, were expected if we consider that different flight phases induce different levels of cognitive effort, depending on the difficulty of each phase.

Effect of flight phases can be considered as a reflection of the differences induced notably by the different procedures associated within each phase, and the variation of expertise of each pilot on these specific situations. By extension, these results raise the importance of taking into account the context of the mission and several associated external parameters, when modeling cognitive efforts and similar concepts. However, the current model of mental effort does not capture flight phases or procedures, and more generally does not take avionic parameters into account. A next step will be to link physiology-based predictions with the context of the mission. The use of avionic and contextual parameters will also allow the consolidation of the “expected good behavior” of a pilot, depending on the situation and the mission which must be performed, and hence improve the accuracy of the model. Such behavioral measures and context aware systems are deemed essential for real-world application of mental effort models and similar concepts (Elkin-Frankston et al. 2017, Bracken et al. 2016, 2017).

We argue that the model developed presented in this paper is linked with effort of mental processes, and that it can be used to quantify learning associated with a given procedure. Indeed, it can be argued that the only difference between the “role” of the

pilots (either instructor or trainee) is expertise since they were measured in tandem on similar flights. Expertise itself cannot be measured directly with physiology without context. Given the nature of the physiological data, and the support to hypotheses in a context where expertise plays a great role, it can be stated that we measured variations in physiological parameters associated with effort. Such a model is interesting because it could allow the identification of procedures which are not yet fully acquired by the trainee. If we consider the example of Tandem 5, presented in Fig. 7, the predictions made by the model do not allow the differentiation of the student from the instructor during the second part of the flight (end of flight, landing, and post-flight). This can be explained by the fact that the physiological pattern of the trainee was similar to the one of an instructor, as captured by the model. Given this information, the instructor could, if the decision of the model matches with his personal appreciation, make the decision of spending more time on other, less automated exercises, and thus individualizing the training. Such individualization lies at the heart of optimal training, especially for combat aviation population (Meland et al. 2015).

Future work will focus on the development of a feedback mechanism to the instructors and trainees, and quantification of the benefits – in terms of learning – associated with the use of this tool.

References

- Bracken, B., Palmon, N., Kellogg, L., Elkin-Frankston, S., Farry, M.A.: Cross-domain approach to designing an unobtrusive system to assess human state and predict upcoming performance deficits. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 707–711. Sage, Los Angeles (2016)
- Bracken, B., Palmon, N., Koelle, D., Elkin-Frankston, S., Farry, M.A.: Toolkit to assist researchers to more efficiently conduct experiments assessing human state. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 2032–2036, Sage, Los Angeles (2017)
- Dahlstrom, N., Nahlinder, S.: Mental workload in aircraft and simulator during basic civil aviation training. *Int. J. Aviat. Psychol.* **19**(4), 309–325 (2009)
- Dreyfus, H.L.: Intuitive, deliberative, and calculative models of expert performance. In: Zsombok, C.E., Klein, G. (eds.) *Naturalistic Decision Making*. Psychology Press, Hove (2014)
- Durantini, G., Gagnon, J.-F., Tremblay, S., Dehais, F.: Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* **259**, 16–23 (2014)
- Elkin-Frankston, S., Bracken, B.K., Irvin, S., Jenkins, M.: Are behavioral measures useful for detecting cognitive workload during human-computer interaction? In: Ahram, T.Z., Karwowski, W. (eds.) *Advances in The Human Side of Service Engineering*. AISC, vol. 494, pp. 127–137. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-41947-3_13
- Fornette, M.P., Darses, F., Bourgy, M.: How to improve training programs for the management of complex and unforeseen situations. In: Proceedings of the Human Factors and Ergonomics Society Europe, pp. 217–224 (2015)
- Gagnon, O., Lafond, D., Gagnon, J.F., Parizeau, M.: Comparing methods for assessing operator functional state. In: Proceedings of the 2016 IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support, San Diego, CA, USA (2016)
- Kahneman, D.: Maps of bounded rationality: Psychology for behavioral economics. *Am. Econ. Rev.* **93**(5), 1449–1475 (2003)

- Kennedy, D.O., Scholey, A.B.: Glucose administration, heart rate and cognitive performance: effects of increasing mental effort. *Psychopharmacology* **149**(1), 63–71 (2000)
- Masa, J.F., Corral, J., Martin, M.J., Riesco, J.A., Sojo, A., Hernández, M., Douglas, N.J.: Assessment of thoracoabdominal bands to detect respiratory effort-related arousal. *Eur. Respir. J.* **22**(4), 661–667 (2003)
- Meland, A., Fonne, V., Wagstaff, A., Pensgaard, A.M.: Mindfulness-based mental training in a high-performance combat aviation population: a one-year intervention study and two-year follow-up. *Int. J. Aviat. Psychol.* **25**(1), 48–61 (2015)
- Nourbakhsh, N., Wang, Y., Chen, F.: GSR and blink features for cognitive load classification. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) *INTERACT 2013. LNCS*, vol. 8117, pp. 159–166. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40483-2_11
- Overbeek, T.J., van Boxtel, A., Westerink, J.H.: Respiratory sinus arrhythmia responses to cognitive tasks: effects of task factors and RSA indices. *Biol. Psychol.* **99**, 1–14 (2014)
- Roscoe, A.H.: Assessing pilot workload. why measure heart rate, HRV and respiration? *Biol. Psychol.* **34**(2–3), 259–287 (1992)
- Sanfey, A.G., Chang, L.J.: Multiple systems in decision making. *Ann. N. Y. Acad. Sci.* **1128**(1), 53–62 (2008)
- Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.: *Human Factors Methods: A Practical Guide for Engineering and Design*. Ashgate, Farnham (2005)