View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CURVE/open

Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test

Mansikka, HP, Virtanen, K, Harris, D & Simola, P

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Mansikka, HP, Virtanen, K, Harris, D & Simola, P 2016, 'Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test' *Applied Ergonomics*, vol 56, pp. 213-219
https://dx.doi.org/10.1016/j.apergo.2016.04.006

DOI 10.1016/j.apergo.2016.04.006
ISSN 0003-6870
ESSN 1872-9126

Publisher: Elsevier

NOTICE: this is the author's version of a work that was accepted for publication in *Applied Ergonomics*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Applied Ergonomics*, [56, (2016)] DOI: 10.1016/j.apergo.2016.04.006

© 2016, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test

Heikki Mansikka[a], *, Kai Virtanen[b], Don Harris[a], Petteri Simola[c, d]

a Human Systems Integration Group, Faculty of Engineering and Computing, Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom

b Systems Analysis Laboratory, Department of Mathematics and Systems Analysis, School of Science, Aalto University, P.O.Box 11100, FIN 00076 Aalto, Helsinki, Finland

c Finnish Defence Research Agency, Human Performance Division, Rantatie 66, 04401 Jarvenpaa, Tuusula, Finland

d Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland

Abstract

Increased task demand will increase the pilot mental workload (PMWL). When PMWL is increased, mental overload may occur resulting in degraded performance. During pilots' instrument flight rules (IFR) proficiency test, PMWL is typically not measured. Therefore, little is known about workload during the proficiency test and pilots' potential to cope with higher task demands than those experienced during the test. In this study, fighter pilots' performance and PMWL was measured during a real IFR proficiency test in an F/A-18 simulator. PMWL was measured using heart rate (HR) and heart rate variation (HRV). Performance was rated using Finnish Air Force's official rating scales. Results indicated that HR and HRV differentiate varying task demands in situations where variations in performance are insignificant. It was concluded that during a proficiency test, PMWL should be measured together with the task performance measurement.

Introduction

Pilots' instrument flight rules (IFR) performance is an essential contributor to an operational effectiveness and a safety of flight. European Aviation Safety Agency requires pilots to pass an annual revalidation flight, or a check ride, in order to maintain their IFR currencies (https://easa.europa.eu/regulations). During an IFR check ride, the pilots' performance is assessed against the predefined performance criteria with the intent of verifying their proficiency to operate in instrument meteorological conditions (IMC). In military aviation, similar IFR (re-)validation check rides are used (Mavin and Roth, 2014). Modern, high fidelity simulators allow IFR check rides to be flown in a simulated environment, which reduces risk, allows for more precise data logging and performance feedback, and increases aircraft availability (Sarter et al., 2007; Weitzman et al., 1979; Valverde, 1973).

When task demand is increased during an IFR flight, pilots may compensate it by investing more effort which in turn increases the pilot mental workload (PMWL) (Shaw et al., 2013). Once the mental capacity and/or willingness to invest more effort are exceeded, at some point pilots' performance begins to degrade (Young et al., 2015; O'Donnell, Eggemeier, and Thomas, 1986). There is great potential of compromising flight safety and mission success if these conditions occur during live flying. Measuring PMWL during an IFR check ride can give valuable information about the pilots' ability to maintain the desired performance during events of high task demand. Two pilots with an equal task performance during an IFR check ride may have significantly different cognitive spare capacities, which reflects their potential to cope with subsequent task demand increase (O'Donnell, Eggemeier, and Thomas, 1986; Yerkes and Dodson, 1908). PMWL or spare mental capacity is typically not evaluated during an IFR check ride. To the best of the authors' knowledge, no previous PMWL assessments in the open literature have considered fighter pilots' IFR check rides.

Evaluation of the pilots’ spare mental capacity requires measuring of PMWL for which task performance, subjective reports and physiological metrics are typically used (Boff et al., 1994). Subjective measures of PMWL, such as the NASA-Task Load Index (NASA-TLX) and the Modified Cooper Harper (MCH) scale, have been widely used in aviation domain (Hart and Staveland, 1988; Casali and Wierwille, 1983; Wierwille et al., 1985). While the multidimensional scales, such as the NASA-TLX, have a better reliability, diagnosticity and validity than the uni-dimensional scales, these types of subjective reports are too intrusive to be used during flight or simulated flight. Also, it should be noted that the subjective ratings can become dissociated with performance, especially if the task is resource limited (Yeh and Wickens,1988). In addition, the data for these measures are typically collected after the trial making them less capable of identifying sudden changes in PMWL. In the aviation domain even sudden, short term PMWL overload conditions may jeopardize flight safety and need to be therefore identified. The instantaneous self-assessment (ISA) technique was considered as a potential real-time subjective measure of PMWL. However, as the PMWL was measured during a real IFR check ride, the use of ISA had to be discarded due to potential primary task intrusion (Tattersall and Foord, 1996). Furthermore, if PMWL is to be used as an additional criterion for an IFR check ride performance, possible pilot biases could compromise the reliability of the subjective measures.

Physiological measures do not have the limitations mentioned above. Many physiological measures, however, are not suitable for a check ride use, mainly because they generate unacceptable pilot intrusion, lack pilot acceptance and disturb simulator and aircraft instruments. Heart rate (HR) and heart rate variation (HRV) measures, although somewhat less sophisticated than some of the more recently developed physiological measures, have been widely employed in real and simulated aircraft environments, enjoy high face validity among the pilot population and generate little, if any, pilot intrusion (Ylonen et al., 1997; Lee and Liu, 2003; Hankins and€ Wilson, 1998; Dussault et al., 2004) For these reasons, this study used electrocardiogram (ECG) based measures as a method to measure task demand induced activation of the autonomic nervous system (ANS). From an ECG, the normal-to-normal (NN) interval of the heart rhythm was identified. HR and HRV were derived from the NN interval and used as measures of PMWL. Before this study, HR and HRV have not been measured during a real F/A-18 IFR check ride.

Different components of HRV have been used as measures of ANS modulation. HR, although often associated with reactions to variations in the physical task demands, has also been associated with the changes in the piloting task's mental demands. Table 1 summarizes the products of the NN interval used in this study. Also, Table 1 describes how HR and the components of HRV are affected by the increased PMWL.

TABLE 1 ABOUT HERE

Several studies have shown HRV and HR to be relatively insensitive to changes in task demand, with HRV and HR being able to differentiate the task demand variations only between the task and rest conditions (Veltman and Gaillard, 1996; Jorna, 1992; Wilson, 1992; Fallahi et al., 2016, Wei et al., 2014). In a more recent study, Mansikka et al. (2016) successfully used HR and HRV to identify different levels of task demands during simulated fighter missions when the task demand was intentionally and somewhat artificially varied from very modest to extremely high; the temporal demand of the repeated flying task varied from 6 min to 35 s to 2 min 20 s. In this study, the fighter pilots' performance and PMWL were measured during a real instrument check ride without artificial manipulation of task demand. The instrument check ride was carried out in a high-fidelity simulator and comprised of clearly identifiable mission segments. Each mission segment consisted of different piloting task and thus generated mission segment specific task demands. The pilots' PMWL measured with HR/HRV and performance variations between different mission segments was studied.

It was hypothesized that HR and the HRV components presented in Table 1 could differentiate the task demand differences between the check ride's mission segments. Also, it was theorized that the PMWL measures could identify differences between the mission segments even when there were no significant performance differences between them. That is, even when the pilots could maintain their performance unchanged from mission segment to mission segment, there would be significant differences in their ANS responses to the changing task demands. Such a finding would support the use of both performance and PMWL measures in future check rides; the differences in the values of the PMWL measures could provide valuable insights about the PMWL's relation to performance and about the differences in the pilots' cognitive spare capacities during events of varying task demands. Ultimately, the level of PMWL could at some later stage be used as an additional IFR check ride criterion where the pilot would have to achieve a minimum performance score without exceeding the given level of PMWL. This study was aimed at evaluating if HR and HRV have potential as such measures of PMWL.

Method

Participants

Data from 26 volunteer Finnish Air Force (FinAF) male F/A-18 pilots with a 1st class IFR qualification were collected. The pilots' average flying experience with the F/A-18 was 781 h (SD ¼ 390). Relevant data concerning the pilots' activities for the 12 h before the check ride were recorded. All pilots had passed an extensive aeromedical examination within the last 12 months and were fit to fly at the time of the study. A written, informed consent was obtained from each subject. The study was reviewed and approved in the Coventry University's Ethical Review Process.

Study design

The data collection was undertaken during official F/A-18 1st class IFR check rides. A Boeing built Weapon Tactics and Situational Awareness Trainer (WTSAT) was used for the piloting task. The WTSAT is used at the FinAF's fighter squadrons for basic and advanced F/A-18 pilot training. The WTSAT is a non-motion, high fidelity flying simulator, with a 135 field of view and a fully functional cockpit. The WTSAT replicates the F/A-18 flying characteristics with such a high accuracy that the FinAF F/A-18 pilots can use it to fly their annual instrument check rides. Each pilot's check ride was briefed, controlled, scored and debriefed by a qualified F/ A-18 examiner pilot. Single examiner pilot was responsible for the check rides' scoring. The subjects' official IFR ratings were based on their performance score during the mission. It was therefore assumed that the subjects invested a high degree of mental effort on the task.

The mission comprised of seven recognizable segments: 'Takeoff and Ingress', 'Maneuvering', 'Level Turns', 'Single Engine Maneuvering (SEM)', 'VOR (VHF Omni Directional Radio Range) Approach', ILS (Instrument Landing System) Approach' and 'PAR (Precision Approach Radar) Approach'. The different segments were linked together to form a complete, logical flying mission. The 'Takeoff and Ingress' segment consisted of final checks before the takeoff, IFR takeoff and initial climb, turning climb as well as leveling at the designated altitude, speed and heading. The 'Maneuvering' segment included basic aerobatic maneuvers, recoveries from unusual attitudes and basic fighter maneuvers. The 'Level Turns' segment contained a serial of steep turns with constant bank angle, altitude, load factor and airspeed. The 'SEM' segment included single engine emergency procedures and a simulated single-engine approach followed by a single engine goaround. The approach segments comprised of standard approaches with identifiable phases of initial approach, intermediate approach and final approach. The 'VOR approach' and the 'ILS approach' segments included also the missed approach phase. It was expected that the segments including instrument approaches would have the highest task loading as the pilots were not allowed to use any autopilot functions while the required control accuracy greatly increased as the pilots descended towards their approach specific minimum altitudes. On the other hand, the ‘Maneuvering’ segment was expected to have the lowest task loading as this segment was closest to a ‘free flight’ condition where the pilots had numerous control input options, each providing an acceptable control accuracy.

The whole mission was flown in IMC. The cloud base was adjusted below the 1st class decision height (DH) for the ILS approach and below the 1st class minimum descent altitude for the VOR approach, thus forcing the pilots to commence go-arounds after reaching their approach specific descent minimum. For the PAR approach, the cloud base was set at DH (60 m/200 ft) thus allowing a full stop landing. A moderate, variable and gusty wind was set for the mission. A typical IFR check ride lasted just over an hour from an engine start to the final landing.

Procedure

Each mission segment was scored by the examiner pilot. For the purposes of analysis, the performance scores were retrieved and calculated as percentages of the maximum scores. For a pilot to achieve a 1st class IFR rating, s/he has to score at least 60% of the maximum score in each segment. Both the control accuracy and the smoothness of the aircraft control were assessed. To minimize the effects of the inter-rater variability, only the control accuracy scores were used for the analysis conducted in this study. The scoring of the ILS approach was based on deviations from the target airspeed, glide slope and localizer. The VOR approaches were scored based on deviations from the target speed, step down fixes and the final approach course. A mission playback was used to increase the scoring accuracy of the approaches; the playback was stopped at every 0.5 NM (0.9 km) during the approaches. While stopped, the deviations were recorded and scored. The scoring of the PAR approach was not used for the analysis as different malfunctions were activated during the PAR approaches making them inconsistent between the pilots. To achieve a 100% ILS performance score, the maximum control error at 5 NM (9.3 km) was 60 ft (18.3 m) for the glideslope, 300 ft (91.4 m) for the localizer and 5 kt (9.3 km/h) for the airspeed. As the control accuracy requirement is increased towards the approach minima, the maximum allowable control error at DH was 10 ft (3.0 m) for the glideslope, 20 ft (6.1 m) for the localizer and 5 kt (9.3 km/h) for the airspeed. The scoring of the VOR segment was similar to that of the ILS segment. As the VOR approach is a non-precision approach, its precision requirement was not as tight as it was for the ILS segment. The scoring of the other segments was based on variations of target flight parameters defined for different maneuvers and reflected the control accuracy requirements of the ILS approach.

The ECG recording, manipulation and interpretation were done in accordance with the guidance in Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology (Camm et al., 1996). Before the mission, the subjects were equipped with Mind Media Nexus-10 MKII system for the ECG recording. Three electrodes were placed below the left (negative) and right (ground) clavicle and the left costal cartilage (positive), respectively. ECG data were collected continuously during the whole mission. Five minute ECG samples were retrieved from each mission segment for further analysis. Data were first recorded using Biotrace þ software (version V2012C) from where the samples were exported to Kubios HRV 2.2 software for further analysis and NN interval artifact removal. A sampling rate of 1024 Hz was used for all samples and a 256 s window width with a 50% overlap was used for the fast Fourier transformation. Piecewise cubic spline interpolation was used to support artifact corrections; on preliminary inspection, all inter beat intervals 0.35 s longer or shorter than the local average, at HR of 60 beats per minute, were considered as artefacts. However, the artefacts were ultimately carefully edited using beat to beat visual checks and manual corrections (Tarvainen et al., 2014; Camm et al.,1996). Noisy data were excluded from the analysis. The values of HR and the following components of HRV were analyzed from each subject: MEANRR, SDNN, RMSSD, NN50, pNN50, HRVTRI, LFnu, HFnu, and LF/HF.

Results

Treatment of data

Each pilots' performance data from every mission segment were retrieved. In a similar fashion, the values of the pilots’ HR and HRV components were collected from each mission segment. In general, ECG data were uncluttered with very few artefacts. However, ECG data from one subject were lost due to a software error. In addition, ECG data from one subject were corrupted and thus excluded from the analysis. As a result, the findings of this study were based on data from 24 subjects.

Data were analyzed using IBM SPSS software (version 22). Normality of the distributions of the performance scores as well as the HR and the HRV components' values in each mission segment were verified using the Shapiro-Wilk test. The performance scores and the HR/HRV components’ values were first analyzed with the repeated measures ANOVA. Only after the ANOVA results proved to be significant, the results were further analyzed using paired t-test for the subsequent pairwise comparisons.

Analysis

The pilots were able to maintain high performance levels across all the mission segments; the ‘SEM’ segment had the highest mean performance score of 97.3% (SD = 4.0) whereas the ‘Maneuvering’ had the lowest mean performance score of 89.8% (SD = 5.5). Table 2 presents the descriptive statistics for the performance scores.

TABLE 2 ABOUT HERE

The repeated measures ANOVA revealed significant differences in the performance scores between the mission segments; $F(5,115) = 4.9$, $p < 0.05$, partial $h^2 = 0.176$. In the pairwise comparisons, seven mission segment pairs had significant performance differences between them. The results of the pairwise comparisons are summarized in Table 3.

While the pilots' performance remained relatively stable between the different mission segments, there were changes in HR and in the components of HRV. The descriptive statistics of the HR values and the HRV components’ values for different mission segments are presented in Table 4.

TABLES 3 and 4 ABOUT HERE

Sphericity was assumed only for some HR/HRV measures. As a result, the degrees of freedom in ANOVAs vary between different HR/HRV measures. The repeated measures ANOVA revealed significant differences across the mission segments for: MEANRR $F(5,115) = 3.15$, $p < 0.05$, partial $h^2 = 0.120$; MEANHR $F(5,115)$ ¼ 2.78, $p < 0.05$, partial $h^2 = 0.108$; SDNN $F(3,71) = 3.51$, $p < 0.05$, partial $h^2 = 0.132$; HRVTRI $F(5,115) = 7.79$, $p < 0.05$, partial $h^2 = 0.253$; LF/HF$(5115) = 3.16$, $p < 0.05$, partial $h^2 = 0.121$. ANOVA did not reveal significant differences for: LFnu $F(3,79) = 2.33$, $p > 0.05$, partial $h^2 = 0.092$; HFnu $F(3,79) = 2.32$, $p > 0.05$, partial $h^2 = 0.092$; RMSSD $F(3,74) = 1.26$, $p > 0.05$, partial $h^2 = 0.052$; NN50 $F(3,75) = 1.65$, $p > 0.05$, partial $h^2 = 0.067$; pNN50 $F(3,73) = 2.06$, $p > 0.05$, partial $h^2 = 0.082$. The measures with the significant ANOVA differences were further analyzed with pairwise comparisons. These results are summarized in Table 5.

TABLE 5 ABOUT HERE

All mission segments that were differentiated by the pilots' performance scores were also differentiable by their HR/HRV responses. The ‘Takeoff and Ingress’ and ‘SEM’ mission segment pair and the “ILS Approach” and “VOR Approach” segment pair were neither differentiated by the performance scores nor by the HR/ HRV responses. All other mission segments were differentiated by some of the HR/HRV measures. HR and the HRV components were able to differentiate six mission segment pairs that had nonsignificant performance differences. The mission segment pairs with the significant performance score differences and/or with the significant differences in the HR values or in the HRV components' values are summarized in Table 6.

TABLE 6 ABOUT HERE

Discussion

IFR check rides, along with other safety mechanisms, are in place to ensure pilots' ability to cope with the task demands of their flying duties (Mavin and Roth, 2014). When flight safety and the mission success are being considered, it is critical that the pilots' mental capacity is not exceeded during a flying mission. The unexpected events during a live flying mission can exceed the task demands experienced during a simulator IFR check ride. If a pilot is already at the upper limit of his/her cognitive capacity during a simulator check ride, an increased task demand during a live flying mission have a potential to exceed the pilot's mental capacity and impair his/her performance. Measuring pilots' PMWL and performance during an IFR check ride can give valuable insights about pilots' ability to cope with the high task demands. This study successfully utilized HR and HRV, as measures of PMWL, during real F/ A-18 IFR check rides.

As shown by this simulator study, the experienced F/A-18 pilots were able to maintain high and mostly equal performance across all the segments of the IFR check ride. At the same time, the HR values and the HRV components' values indicated that their PMWL between the different mission segments was not equal. With a slightly modified test design, it would be possible to study if the pilots could be differentiated by their performance and PMWL; insignificant and non-significant differences in pilots' performance coupled with significant differences in their PMWL could be used to reflect the pilots’ different mental spare capacities.

This study was able to replicate the findings of the earlier mental workload related HR and HRV studies (Roscoe, 1975, 1993; Wilson, 2002; Terkelsen et al., 2005; Tran et al., 2010; Li et al., 2009; Orsila et al., 2008; Deepak et al., 2014; Taelman et al., 2011; Sun et al., 2012; Cinaz et al., 2013; Svensson and Wilson, 2002; Vuksanovic and Gal, 2007). Unlike some earlier studies using HR and HRV (Veltman and Gaillard, 1996; Wilson, 1992; Jorna, 1992; Fallahi et al., 2016, Wei et al., 2014), this study was able to differentiate ANS response variations measured with HR/HRV between mission segments instead of differentiating just the rest and trial conditions. In addition, whereas Mansikka et al. (2016) successfully used HR and HRV to differentiate large task demand changes during a simulated flight, this study replicated these results with smaller task demand variations obtained during a realistic, simulated flying task. The flying mission consisted of separate mission segments. Each mission segment exposed pilots to different task demands as each mission segment tested different aspects of pilots' IMC flying abilities. Both the performance and the ECG data were retrieved from each mission segment. As a result, HR and HRV data provided an adjunct and sensitive measure of PMWL and could, at some later stage, be used to support the evaluation of the pilots' spare mental capacities. The measured differences in the pilots' spare mental capacities can give valuable information about the pilots’ ability to maintain the desired performance during events of high task demand.

As summarized in Table 6, the ‘Takeoff and Ingress’ and ‘SEM’ mission segment pair and the “ILS Approach” and “VOR Approach” segment pair were neither differentiated by the performance scores nor by the HR/HRV responses. Thus, it can be concluded that the task demand and the resulting PMWL of these mission segment pairs were very similar. There were seven mission segment pairs (‘Takeoff and Ingress’ and ‘Maneuvering’; ‘Takeoff and Ingress’ and ‘ILS Approach’; ‘Maneuvering’ and ‘Level Turns’; ‘Maneuvering’ and ‘SEM’; ‘Maneuvering’ and ‘VOR Approach’; ‘Maneuvering’ and ‘ILS Approach’; ‘SEM’ and ‘ILS Approach’) that were differentiated both by the performance scores and by one or more of the HR/HRV measures. It was concluded that these segments were different in their task demands and also generated different ANS responses as revealed by the changes in HR/HRV.

Out of the total of 15 mission segment pairs analyzed, there were eight mission segment pairs which were not differentiated by the performance scores. However, of these eight mission segment pairs six were differentiated by one or more of the HR/HRV measures. In other words, there were PMWL differences in 75% of those mission segment pairs that could not reveal differences in performance, i.e., PMWL measured with HR/HRV was more sensitive than the performance score when the mission segments and their task demands were differentiated. Although the overt, traditional performance measures suggest otherwise, the subjects’ average potential to operate effectively and safely varied between the mission segments. HR/HRV proved to be potential measures of PMWL should the individual PMWL and performance differences between the pilots be evaluated and used as an additional IFR check ride criterion where the pilot would have to achieve a minimum performance score without exceeding the given level of PMWL.

Conclusions and recommendations

It is concluded that the differences in the pilots' PMWL between the check ride's mission segments can be differentiated by HR and HRV. HR and HRV were also capable of identifying the differences between the mission segments even when there were no significant performance differences between them. The utilization of HR and HRV as measures of PMWL can improve awareness of the pilots' mental potential to respond to high task demands and may support the assessment of the differences between their spare capacities. Assuming that a continuous increase of PMWL will - at some point - degrade pilot performance, the evaluation of individual differences could reveal if some pilots are closer to the threshold of impaired performance than others with a similar performance.

This can give valuable insights about the pilots' spare mental capacities during events of high task load, which in turn could be used to improve both the flight safety and the operational effectiveness.

For the highly experienced pilots, an instrument check ride is a routine mission with a few, if any, unexpected events. Therefore, an instrument check ride does not challenge the pilots' perception capacities with an excessive information load or the higher order mental processing capacities with a constantly changing operating environment. Nor does it challenge the pilots' ability to project the future actions of the other entities relative to their ownship. Future studies should stress these issues as they are essential elements of any tactical fighter mission. However, such studies should be cautious when trying to explain the association between PMWL measured with HR/HRV and performance; it would be an oversimplification to expect that PMWL alone could explain the performance differences. During the tactical flying missions, the pilots' ability to build and maintain their situational awareness is greatly stressed. As the situational awareness, PMWL and the pilot performance are interlinked, any future attempts to explain the pilots’ mental potential during a highly complex flying mission should somehow include all these three measures.

This study shows promising results. Both HR and HRV predominantly followed the expected pattern. The association between the performance and HR/HRV was not completely monotonic and thus requires further research - especially in a more complex fighter aviation scenario. However, this study together with the earlier findings provides an encouraging basis to extend the testing of HR/HRV's sensitivity to varying task demands in more tactical flying environments.

Disclosure statement

The study was funded by the Finnish Defence Research Agency. The authors have no financial interest or benefit arising from applications of this research.

Acknowledgments

This study was ordered by the Finnish Air Force and was conducted jointly by the Finnish Defence Research Agency, Aalto University and Coventry University. The study was coordinated and funded by the Finnish Defence Research Agency.

References

Boff, K.R., Kaufman, L., Thomas, J.P., 1994. Handbook of perception and human performance. In: Cognitive Processes and Performance, vol. 2. John Wiley and Sons Inc., New York.

Camm, A.J., Malik, M., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L., Kleiger, R.E., 1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the Eur. Soc. of Cardiol. the North Am. Soc. of Pacing Electrophysiol. Circ. 93 (5), 1043e1065.

Casali, J.G., Wierwille, W.W., 1983. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Hum. Factors 25 (6), 623e641.

Cinaz, B., Arnrich, B., La Marca, R., Tro€ster, G., 2013. Monitoring of mental workload levels during an everyday life office-work scenario. Pers. Ubiquit. Comput. 17 (2), 229e239.

Deepak, A., Deepak, A.N., Nallulwar, S., Vitthal, K., 2014. Time domain measures of heart rate variability during acute mental stress in type 2 diabetics-a case control study. Natl. J. Physiol. Pharm. Pharmacol. 4 (1), 34e38.

Dussault, C., Jouanin, J.-C., Guezennec, Y.-C., 2004. EEG and ECG changes during selected flight sequences. Aviat. Space Environ. Med. 75 (10), 889e897.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A.R., Miyake, S., 2016. Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. J. Appl. Ergon. 52 (2016), 95e103.

Hankins, T.C., Wilson, G.F., 1998. A comparison of heart rate, eye activity, EEG and Subjective measures of pilot mental workload during flight. Aviat. Space Environ. Med. 69 (4), 360e367.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52, 139e183.

Jorna, P.G.A.M., 1992. Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. Biol. Psychol. 34 (2), 237e257.

Lee, Y.-H., Liu, B.-S., 2003. Inflight workload assessment: comparison of subjective and physiological measurements. Aviat. Space Environ. Med. 74 (10), 1078e1084.

Li, Z., Snieder, H., Su, S., Ding, X., Thayer, J.F., Treiber, F.A., Wang, X., 2009. A longitudinal study in youth of heart rate variability at rest and in response to stress. Int. J. Psychopysiol. 73 (3), 212e217.

Mansikka, H., Simola, P., Virtanen, K., Harris, D., Oksama, L., 2016. Fighter pilots' heart rate, heart rate variation and performance during instrument approaches. Ergonomics 1e9.

Mavin, T.J., Roth, W.-M., 2014. A holistic view of cockpit performance: an analysis of the assessment discourse of flight examiners. Int. J. Aviat. Psychol. 24 (3), 210e227.

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., Yamamura, T., 2009. Physiological responses to workload change. a test/retest examination. J. Appl. Ergon. 40 (6), 987e996.

O'Donnell, R.D., Eggemeier, F.T., 1986. Workload assessment methodology. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance, Cognitive Processes and Performance, vol. 2. John Wiley and Sons, Inc., New York, 42-1-42-49.

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., Savinainen, M., Nygard, C.-H., 2008. Perceived mental stress and reactions in heart rate variabilityea pilot study among employees of an electronics company. Int. J. Occup. Saf. Ergon. 14 (3), 275e283.

Roscoe, A.H., 1993. Heart rate as a psychophysiological measure for in-flight workload assessment. Ergonomics 36 (9), 1055e1062.

Roscoe, A.H., 1975. Heart rate monitoring of pilots during steep-gradient approaches. Aviat. Space Environ. Med. 46 (11), 1410e1413.

Sarter, N.B., Mumaw, R.J., Wickens, C.D., 2007. Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. Hum. Factors 49 (3), 347e357.

Shaw, T., Satterfield, K., Ramirez, R., Finomore, V., 2013. Using cerebral hemovelocity to measure workload during a spatialised auditory vigilance task in novice and experienced observers. Ergonomics 56 (8), 1251e1263.

Skibniewski, F.W., Dziuda, L., Baran, P., Krej, M., Guzowski, S., Piotrowski, M., Truszczynski, O., 2015. Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. Aerosp. Med. Hum. Perform. 86 (6), 518e523.

Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P., Griss, M., 2012. Activityaware mental stress detection using physiological sensors. In: Uhler, D., Mehta, K. (Eds.), Mobile Computing, Applications, and Services. Springer, Berlin, pp. 211e230.

Svensson, E., Wilson, G.F., 2002. Psychological and psychophysiological models of pilot performance for systems development and mission evaluation. Int. J. Aviat. Psychol. 12 (1), 95e110.

Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A., Van Huffel, S., 2011. Instantaneous changes in heart rate regulation due to mental load in simulated office work. Eur. J. Appl. Physiol. 111 (7), 1497e1505.

Tarvainen, M.P., Niskanen, J.-P., Lipponen, J.A., Ranta-Aho, P.O., Karjalainen, P.A., 2014. Kubios HRV e heart rate variability analysis software. Comput. Methods Programs Biomed. 113 (1), 210e220.

Tattersall, A.J., Foord, P.S., 1996. An experimental evaluation of instantaneous selfassessment as a measure of workload. Ergonomics 39 (5), 740e748.

Terkelsen, A.J., Mølgaard, H., Hansen, J., Kæseler, A.O., Jensen, T.S., 2005. Acute pain increases heart rate: differential mechanisms during rest and mental stress. Auton. Neurosci. 121 (1e2), 101e109.

Tran, B.W., Papoiu, A.D.P., Russoniello, C.V., Wang, H., Patel, T.S., Chan, Y.-H., Yosipovitch, G., 2010. Effect of itch, scratching and mental stress on autonomic nervous system function in atopic dermatitis. Acta Derm. Venereol. 90 (4), 354e361.

Valverde, H.H., 1973. A review of flight simulator transfer of training studies. Hum. Factors 15 (6), 510e522.

Veltman, J.A., Gaillard, A.W.K., 1996. Physiological indices of workload in a simulated flight task. Biol. Psychol. 42 (3), 323e342.

Vuksanovic, V., Gal, V., 2007. Heart rate variability in mental stress aloud. Med. Eng. Phys. 29 (3), 344e349.

Wei, Z., Damin, Z., Xiaoru, W., Chen, L., Huan, Z., 2014. A model for discrimination and prediction of mental workload of aircraft cockpit display interface. Chin. J. of Aeronautics 27 (5), 1070e1077.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., Compton, G.L., 1979. Proficiency maintenance and assessment in an instrument flight simulator. Hum. Factors J. of the Hum. Factors Ergonomics Soc. 21 (6), 701e710.

Wierwille, W.W., Rahimi, M., Casali, J.G., 1985. Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. Hum. Factors 27 (5), 489e502.

Wilson, G.F., 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int. J. Aviat. Psychol. 12 (1), 3e18.

Wilson, G.F., 1992. Applied use of cardiac and respiration measures: practical considerations and precautions. Biol. Psychol. 34 (2), 163e178.

Wu, B., Hou, F., Yao, Z., Niu, J., Huang, W., 2011. Using physiological parameters to evaluate operator's workload in manual controlled rendezvous and docking (RVD). In: Duffy, V.G. (Ed.), Digital Human Modeling. Springer, Berlin, pp. 426e435.

Ylo€nen, H., Lyytinen, H., Leino, T., Lepp€aluoto, J., Kuronen, P., 1997. Heart rate responses to real and simulated BA hawk MK 51 flight. Aviat. Space Environ. Med. 68 (7), 601e605.

Yeh, Y.-Y., Wickens, C.D., 1988. Dissociation of performance and subjective measures of workload. Hum. Factors 30 (1), 111e120.

Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58 (1), 1e17.

Web References

http://easa.europa.eu/system/files/dfu/AMC%20and%20GM%20to%20Part-FCL.pdf.

(accessed 26.08.15.).

Table 1

HR and HRV components and their expected change due to increased PMWL.

MeasureUnit	Description	Expected change	References
MEANHR [1/ min]	The mean heart rate.	Increase	Roscoe 1975; Wilson 2002; Roscoe 1993; Vuksanovic and Gal, 2007a
MEANRR [ms]	The mean of NN intervals.	Decrease	Terkelsen et al., 2005; Sun et al., 2012
SDNN [ms]	The standard deviation of NN intervals.	Decrease	Terkelsen et al., 2005; Tran et al., 2010
RMSSD [ms]	The square root of the mean squared differences between successive NN intervals.	Increase	Li et al., 2009; Orsila et al., 2008
NN50 [count}	The number of successive NN interval pairs that differ more than 50 ms	Decrease	Deepak et al., 2014
pNN50 [%]	The NN50 divided by the total number of NN intervals.	Decrease	Taelman et al., 2011
HRVTRI [-]	The integral of the NN interval density distribution divided by the maximum of the distribution.	Decrease	Cinaz et al., 2013
LFnu [-]	The normalized low frequency (0.04e0.15 Hz) component of HRV.	Increase	Wu et al., 2011; Miyake et al., 2009
HFnu [-]	The normalized high frequency (0.15e0.4 Hz) component of HRV.	Decrease	Wilson 2002
LF/HF [-]	The ratio between the power of low frequency (LF) and high frequency (HF) components of HRV.	Increase	Skibniewski et al., 2015

Table 2

Means (M), standard deviations (SD), maxima (Max) and minima (Min) of the mission segments' performance scores (N = 24). SEM = Single Engine Maneuvering, VOR = VHF Omni Directional Radio Range, ILS = Instrument Landing System.

Mission segment	**Performance scores (% from maximum)**			
	M	**SD**	**Max**	**Min**
Takeoff and Ingress	96.3	3.1	100.0	90.0
Maneuvering	89.8	5.5	98.3	78.3
Level Turns	95.0	9.6	100.0	58.5
SEM	97.3	4.0	100.0	83.3
VOR Approach	94.4	6.3	100.0	73.6
ILS Approach	93.8	4.3	100.0	86.0

Table 3

Pairwise means (M) and standard errors (SE) of the performance scores as well as the corresponding test statistics (t) in the pairwise comparisons between the mission segments.

Mission Segment Pairs		M	SE	t
Take off and Ingress	Maneuvering	6.4	1.2	5.248***
	Level Turns	1.3	2.2	0.577
	SEM	-1.0	1.0	-0.957
	VOR Approach	1.9	1.4	1.376
	ILS Approach	2.5	1.0	2.427*
Maneuvering	Level Turns	-5.2	2.1	-2.419*
	SEM	-7.4	1.3	-5.657***
	VOR Approach	-4.5	1.5	-3.082**
	ILS Approach	-4.0	1.3	-3.150**
Level Turns	SEM	-2.2	2.2	-0.988
	VOR Approach	0.7	2.4	0.273
	ILS Approach	1.2	1.9	0.629
SEM	VOR Approach	2.9	1.5	1.906
	ILS Approach	3.4	1.1	3.085**
VOR Approach	ILS Approach	0.6	1.5	0.389

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ (N = 24)

Table 4

Means (M) and standard deviations (SD) of the HR values and the HRV components' values for the mission segments (N = 24).

		Takeoff and ingress		**Maneuvering**		**Level turns**		**SEM**		**VOR approach**		**ILS approach**	
		M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
MEANRR	**[ms]**	677.6	115.3	676.0	107.3	689.4	126.8	686.7	114.0	661.8	103.6	666.2	105.1
MEANHR	**[1/min]**	91.7	14.1	91.7	13.2	90.2	15.4	90.2	13.7	93.5	13.5	92.7	13.4
SDNN	**[ms]**	64.7	27.0	67.8	33.3	53.4	19.6	61.2	22.5	64.4	25.3	58.0	20.5
RMSSD	**[ms]**	27.4	11.3	26.8	16.1	25.0	11.2	27.6	13.6	25.8	10.1	23.9	9.6
NN50	**[count]**	29.5	26.6	27.8	29.4	24.5	26.4	28.7	32.0	24.4	23.0	22.5	22.2
pNN50	**[%]**	7.2	7.1	7.0	8.2	6.4	7.6	7.3	8.8	5.7	6.0	5.4	5.9
HRVTRI	**[-]**	14.2	4.7	15.6	5.4	11.7	3.8	13.8	5.0	12.9	4.3	12.9	4.8
LFnu	**[-]**	77.0	10.0	78.3	12.7	79.8	11.1	80.2	9.3	83.0	9.2	82.0	9.4
HFnu	**[-]**	22.9	10.0	21.6	12.7	20.1	10.9	19.7	9.3	16.9	9.2	18.0	9.3
LF/HF	**[-]**	4.2	2.3	4.9	2.8	5.5	3.7	5.4	3.3	6.6	3.9	6.0	3.2

Table 5

Pairwise means (M) and standard errors (SE) of HR and the HRV components as well as the corresponding test statistics (t) in the pairwise comparisons between the mission segments

		MEANRR			MEANHR			SDNN			HRVTRI			LF/HF		
		M	SE	t	M	SE	t	M	SE	t	M	SE	t	M	SE	t
Takeoff and Ingress	Maneuvering	1.6	6.4	0.256	0.0	0.8	0.056	-3.1	3.2	-0.974	-1.4	0.5	-2.655*	-0.7	0.6	-1.198
	Level Turns	-11.8	7.7	-1.528	1.4	1.1	1.279	11.4	2.7	4.225***	2.5	0.5	4.572***	-1.3	0.7	-1.954
	SEM	9.1	7.5	1.209	1.4	1.0	1.425	3.6	3.2	1.116	0.3	0.7	0.460	-1.2	0.6	-2.066
	VOR Approach	15.8	8.6	1.831	-1.9	1.2	-1.641	0.3	4.7	0.072	1.3	0.5	2.319*	-2.4	0.8	-3.062**
	ILS Approach	11.4	9.8	1.162	-1.1	1.2	-0.905	6.7	3.5	1.933	1.3	0.6	2.148*	-1.7	0.6	-2.751*
Maneuvering	Level Turns	-13.4	8.4	-1.587	1.5	1.2	1.224	14.4	4.4	3.316**	3.9	0.8	5.226***	-0.6	0.6	-0.971
	SEM	-10.7	7.9	-1.357	1.5	1.1	1.343	6.6	4.2	1.600	1.7	0.7	2.351*	-0.5	0.7	-0.796
	VOR Approach	14.2	9.2	1.539	-1.9	1.3	-1.456	3.4	5.3	0.648	2.7	0.8	3.404**	-1.7	0.8	-2.195*
	ILS Approach	9.8	9.8	0.996	-1.1	1.3	-0.791	9.8	5.3	1.847	2.7	0.8	3.403**	-1.0	0.6	-1.818
Level Turns	SEM	2.7	6.9	0.386	0.0	0.9	0.150	-7.8	2.3	-3.325**	-2.2	0.8	-2.644*	0.1	0.6	0.134 -
	VOR Approach	27.6	10.9	2.534*	3.3	1.3	2.538*	-11.0	4.8	-2.320	-1.2	0.7	-1.848	-1.1	0.9	1.311
	ILS Approach	23.2	10.9	2.126*	-2.5	1.3	-1.904	-4.6	2.8	-1.637*	-1.2	0.7	-1.875	-0.4	0.7	-0.638
SEM	VOR Approach	24.9	9.3	2.684*	-3.3	1.1	-2.972**	-3.2	4.0	-0.818	1.0	0.6	1.521	-1.2	0.6	-2.004
	ILS Approach	20.5	8.3	2.464*	-2.5	1.0	-2.450*	3.1	3.0	1.042	0.9	0.8	1.139	-0.5	0.7	-0.784
VOR	ILS Approach	-4.4	7.0	-0.636	0.8	0.9	0.905	6.4	4.4	1.443	0.0	0.5	0.042	0.7	0.6	1.225

***p < 0.001; **p < 0.01; *p < 0.05 (N = 24).

Table 6

The mission segment pairs with the significant differences of the performance score and/or with the significant differences of HR and the HRV components (N = 24). The mission segments pairs with the significant performance score differences are denoted by a shaded background.

	Take off and ingress	**Maneuvering**	**Level Turns**	**SEM**	**VOR Approach**
Maneuvering	HRVTRI				
Level Turns	SDNN, HRVTRI	SDNN, HRVTRI			
SEM	-	HRVTRI	SDNN, HRVTRI		
VOR Approach	HRVTRI, LF/HF	HRVTRI, LF/HF	MEANRR, MEANHR, SDNN	MEANRR, MEANHR	
ILS Approach	HRVTRI, LF/HF	HRVTRI	MEANRR	MEANRR, MEANHR	-