Fighter pilots' heart rate, heart rate variation and performance during instrument approaches

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

Original citation & hyperlink:
Mansikka, HP, Simola, P, Virtanen, K, Harris, D & Oksama, L 2016, 'Fighter pilots' heart rate, heart rate variation and performance during instrument approaches' *Ergonomics*, vol 59, no. 10, pp. 1344 -1352
http://dx.doi.org/10.1080/00140139.2015.1136699

DOI 10.1080/00140139.2015.1136699
ISSN 0014-0139
ESSN 1366-5847

Publisher: Taylor and Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in Ergonomics on 4th March 2016, available online: http://www.tandfonline.com/10.1080/00140139.2015.1136699

MANUSCRIPT COVERPAGE, MANSIKKA

Corresponding Author: Heikki Mansikka, Human Systems Integration Group, Faculty of Engineering and Computing, Coventry University, Coventry, United Kingdom; Finnish Defence Research Agency, Riihimäki, Finland

Address: Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom

Mobile: +358 (0)299 800

Email: mansikkh@uni.coventry.ac.uk

Co-author: Petteri Simola, Finnish Defence Research Agency, Tuusula, Finland; Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland

Address: Finnish Defence Research Agency, Human Performance Division, Rantatie 66, 04401 Järvenpää, Finland

Mobile: +358 (0)299 800

Email: petteri.simola@mil.fi

Co-author: Kai Virtanen, Systems Analysis Laboratory, Department of Mathematics and Systems Analysis, School of Science, Aalto University, Helsinki, Finland

Address: P.O.Box 11100, FIN 00076 Aalto, Finland

Mobile: +358 (0)505658418

Email: kai.virtanen@aalto.fi

Co-author: Lauri Oksama, National Defence University, Helsinki, Finland

Address: Kadettikouluntie 7, 00861 Helsinki, Finland

Mobile: +358 (0)299550451

Email: loksama@utu.fi

Co-author: Don Harris, Human Systems Integration Group, Faculty of Engineering and Computing, Coventry University, Coventry, United Kingdom

Address: Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom

Mobile: +44 (0)24 7765 8052

Email: don.harris@coventry.ac.uk

Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during Instrument Approaches

Abstract: Fighter pilots' heart rate (HR), heart rate variation (HRV) and performance during instrument approaches were examined. The subjects were required to fly instrument approaches in a high fidelity simulator under various levels of task demand. The task demand was manipulated by increasing the load on the subjects by reducing the range at which they commenced the approach. HR and the time domain components of HRV were used as measures of pilot mental workload (PMWL). The findings of this study indicate that HR and HRV are sensitive to varying task demands. HR and HRV were able to distinguish the level of PMWL after which the subjects were no longer able to cope with the increasing task demands and their Instrument Landing System (ILS) performance fell to a sub-standard level. The major finding was the HR/HRV's ability to differentiate the sub-standard performance approaches from the high performance approaches.

Keywords: pilot mental workload, heart rate, heart rate variation, performance

Practitioner Summary: This paper examined if HR and HRV were sensitive to varying task demands in a fighter aviation environment and if these measures were related to variations in pilot performance.

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

Introduction

The cockpit of a modern multirole fighter is one of the most cognitively demanding work environments, exposing the pilot to extreme physical and psychological stress and fatigue (Driskell and Salas 1991). Pilots' failure to cope with the task demands may degrade flight safety and compromise mission success with fatal results (O'Hare et al. 1994; O'hare 2000; Shappell and Wiegmann 1997; Sheridan and Simpson 1979). During aircraft and system development, a great deal of effort is placed on managing pilot mental workload (PMWL) through the design of the human-machine interfaces, i.e., fitting the task to the man (Grandjean and Kroemer 1997). However, once the platform is released for operational use, workload management becomes an issue of fitting the man to the task, e.g., through selection and training.

To make sure pilots are competent for their flying duty, air forces conduct mandatory proficiency checks for their flight crews (Mavin and Roth 2014). These proficiency checks or 'check rides' are conducted to assess the pilots' performance against standards with the aim of guaranteeing their acceptable operational performance.

A check ride usually consists of mission critical task elements where pilot performance is evaluated by an instructor pilot or examiner. For the pilot to pass a check ride, s/he needs to score a predefined number of points on a specific grading scale. Based on performance the pilot may be given a certificate to operate a specific platform or piece of equipment, a qualification to operate in certain weather conditions, or the pilot may - or may not - be given an appropriate readiness status.

Even though a live aircraft mission is in some instances the recommended method of conducting a check ride, a high fidelity simulator is often the preferred platform. This is as a result of the lower operating costs of the simulator, the easily adjustable environment and system conditions. Also, missions including critical emergency procedures, use of deadly force or operations with a minimal safety margin are almost impossible to conduct realistically or safely in a live flying environment

While a simulator mission can be designed to be mentally extremely demanding, it will inherently lack the stressors of a real flying mission such as the sense of risk and the fear of collision, injury or death. Consequently a simulator mission is generally less cognitively demanding than a similar mission in a real flying environment (Svensson et al. 1997; Jorna 1993). If PWML is not part of the performance assessment criteria, a check ride conducted in a simulator may provide misleading indications of pilot's performance in real-life situations; a pilot may show acceptable performance in the simulator but executes the same tasks to a sub-standard level in a similar live mission as a result of increased PMWL or stress (Berkun 1964; Lieberman et al. 2005; Young et al. 2014).

Several studies have used laboratory environments to evaluate the relationship between an operator's mental workload and performance (Jorna 1993; Morris and Leung 2006; Vitense, Jacko, and Emery 2003; Zakay and Shub 1998; Iani et al. 2007; Kaber and Endsley 2004). However, these studies provide only a limited understanding of the cognitive demands of real life systems. On the other hand, studies conducted in operational environments provide inadequate insights concerning the level of PMWL leading to pilot's sub-standard performance (Svensson et al. 1997; Magnusson 2002; Veltman 2002; Lahtinen et al. 2007).

There are many sophisticated physiological measures of individual differences in regulated emotional responses available, such as electrodermal activity (EDA) (Collet, Salvia, and Petit-Boulanger 2014), electroencephalography (EEG) (Noel, Bauer Jr, and Lanning 2005), functional near-infrared (fNIR) spectroscopy (Ayaz et al. 2010) and eyelid closure (Mallis and Dinges 2004). These measures are often difficult, if not possible, to implement into a flight simulator environment without unacceptable levels of pilot intrusion and/or disturbances to simulator instrumentation. Other measures such as heart rate (HR) and heart rate variation (HRV), although somewhat less sophisticated and novel, have successfully been applied in a flight simulator environment.

HR and HRV represent the activation of the autonomic nervous system (Stuiver et al. 2014; Hayward et al. 2014; Xhyheri et al. 2012). The time domain methods of HRV analysis involve determining the intervals between successive normal QRS complexes (normal-to-normal, NN). From the NN, other HRV components can be derived and used as measures of mental workload, for example; the mean heart rate (MEANHR) (Pérusse-Lachance et al. 2012; Saperova and Dimitriev 2014; Roman-Liu et al. 2013), the standard deviation of NN intervals (SDNN) (Terkelsen et al. 2005; Tran et al. 2010), the square root of the mean squared differences between successive NN intervals (RMSSD) (Li et al. 2009; Orsila et al. 2008), the number of successive NN interval pairs that differ more than 50 ms (NN50) (Deepak et al. 2014), the NN50 divided by the total number number of all NN intervals (pNN50) (Taelman et al. 2011), the mean of NN intervals (MEANRR) (Terkelsen et al. 2005; Sun et al. 2012) and the triangular index (HRVTRI) (Cinaz et al. 2013), which is the integral of the NN interval density distribution divided by the maximum of the distribution. Several studies have been able to demonstrate the changes in pilots' HR and HRV during different flying mission and phases of missions (Svensson et al. 1997; Aasman, Mulder, and Mulder 1987; Wilson 1993; Veltman and Gaillard 1998; Roscoe 1993; Roscoe 1992). Furthermore, pilots' primary task performance has been successfully linked to PMWL, HR and HRV (Svensson et al. 1997). However, little is known about the relationship between pilot performance and PMWL during a real training mission or a check ride. As a result, practically no attempts have been made to introduce PMWL as an additional performance criterion for a check ride. It is therefore necessary to study the relationship between PMWL and pilot performance in a real, or realistically simulated operating environment, using representative tasks and associated with existing operational performance standards (Jorna 1992; Rasmussen and Jensen 1974). HRV and HR have been proven to be sensitive measures of PMWL at the higher levels of workload. As the present study was focused on the higher end of PMWL, the possible sensitivity limitations of HRV/HR at the lower levels of workload were not an issue.

The aim of this study was to investigate if HR and the selected time domain components of HRV are related to variations in pilot performance during a simulated flying mission. It was hypothesized that pilot performance was associated with HR and the time domain components of HRV. Ultimately the objective of this study was to evaluate if HR and the HRV measures could identify the level of task demands leading to a sub-standard performance. This finding would suggest that the MEANRR could be a useful measure of PMWL if an actual PWML redline could be defined in a fighter simulator environment (Young et al. 2014; Brookhuis, Waard, and Fairclough 2003; Young et al. 2015). To this end, it was necessary to study the dependence between HR, the time domain components of HRV and performance measures. To achieve this, a realistic mission with varying levels of task demands was developed in a high fidelity F/A-18 flight simulator. Operational F/A-18 pilots were recruited as subjects and real air force operational standards were used to assess pilot performance. By utilizing such a test design, pilot performance was measured together with HR and the time domain components of HRV in order to describe the inter-dependence between the pilot performance and PMWL.

Method

Participants

Thirty-five Finnish Air Force (FinAF) F/A-18 pilots participated in the study. The subjects' average flying experience with the F/A-18 was 598 flight hours. Subjects were randomly selected from the fighter squadrons' pilot population. Pilots' backgrounds ranged from wingman to air combat instructor, which resulted in large variation in their flying experience (standard deviation of 445 hours). Written, informed consent was obtained from each subject. A structured proforma was used to collect subjects' background data and information concerning their relevant activities for the 12 hours prior to participating. The proforma was prepared with the assistance of an aeromedical professional (M.D.) from the Satakunta Air Command, Finland. All subjects had gone through an extensive aeromedical testing within the last 12 months and were fit to fly at the time of the study.

Test Design

As a result of the time consuming, repetitive nature of the test design used, it was not possible to utilize an actual instrument proficiency test. Instead, an instrument landing system (ILS) approach, one component of the proficiency test, was used for the task demand manipulation. The subjects completed 12 full test procedures each consisting of an ILS approach with different level of task demand. The task demand was manipulated by increasing the temporal demand on pilots by reducing the range at which they commenced the trial. The trial order was randomized and balanced between subjects.

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 degree 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.

For the study the wind was set to 320 degrees, 10 knots (5.14 m/s) with moderate gusts. Clouds were set to overcast with the cloud top at 30,000 ft (9,144 m) and the cloud base at 200 ft (60 m) from the ground level. Instrument meteorological conditions (IMC) visibility was set to 0 ft (0 m) and the runway visual range was set to 700 m (2,296 ft). Light conditions were set to mirror the average light at the Tampere – Pirkkala airport (International Civil Aviation

Organization code: EFTP) in Finland on 1st June at 12:00 o´clock local time. Runway was dry and the braking action was good. The arresting cable and the net barrier were not available.

Before commencing the trials a baseline ILS mission was flown. For this mission, the simulator was initialized to 2,000 ft (607 m) above ground level , 9.5 NM (17.6 km) from the touchdown point, minimum approach speed, straight and level flight as well as 0 ft (0 m) azimuth and heading error for the standard ILS approach. The cockpit settings were, however, set incorrectly for the approach and landing; for example, the radios were set to wrong frequencies, the altimeter setting was incorrect and the platform was not configured for landing. For the baseline ILS mission, the objective was to fly a simple, undisturbed ILS approach.

Each trial consisted of an ILS task and additional flying related sub tasks. The ILS task was a standard ILS approach to EFTP runway 24. The pilots were tasked to fly the ILS approach using a platform specific minimum approach speed and a flight profile established for the approach in the official instrument approach chart (IAC). The ILS task started at the glideslope (GS) intercept range and ended at 0.5 NM (0.9 km) from the touchdown point, which was the range at which the standard GS met the ILS decision height (DH). The platform specific DH was the same as the cloud base, thus allowing subjects to land after the successful ILS approach. The sub tasks comprised of carefully selected activities relevant to F/A-18 operations. These included tasks such as setting up the cockpit instruments for the specified approach and landing, flying from the DH to touchdown, communicating with the air traffic controller (ATC) and reacting to in-flight emergencies requiring immediate pilot actions. The sub tasks and the different components of the ILS task used in the study are listed in Table 1 where the ILS task components are marked with a shaded background. To force the subjects to study the instrument approach chart, during each trial a different instrument approach chart was used (only the ILS flight profile, runway altitude and ILS localizer (LLZ) frequency were kept identical). The subjects had to copy three altitudes from the instrument approach chart to their knee pads and to study the instrument approach chart’s frequencies as well as to tune six radio presets accordingly. With the exception of the instrument approach chart, the pilots were highly familiar with the sub tasks so they had no need to refer to other check lists or supporting documentation to undertake them. Although the sub tasks were standard procedures for any F/A-18 pilot, the subjects practiced each sub task before the trials.

[Table 1 near here]

For the trials the simulator was initialized to same parameters (with the exception of the starting range) as for the baseline ILS mission. As the task demand was manipulated by varying the starting ranges of the trials, each trial started from a trial specific starting range. The starting ranges were measured as a horizontal distance from the touchdown point. The starting ranges varied from 5.5 NM (10.2 km) to 15.5 NM (28.7 km) at 1 NM (1.9 km) increments. The minimum value of the starting range variable (5.5 NM) equaled the ILS glideslope intercept range, i.e., the starting point of the ILS task. The maximum value of the starting range variable (15.5 NM) was defined during the pretesting as a maximum range it took for the pilots to perform the sub tasks when done undisturbed and at pilots' own pace. Subjects were tasked to fly each approach at a constant minimum approach speed. As a result the time pressure for the ILS task and the sub tasks varied from 6 minutes and 35 seconds (15.5 NM starting range) to 2 minutes and 20 seconds (5.5 NM starting range). Each trial ended at touchdown or attempted touchdown. A 1 NM decrease in the starting range reduced the available time to conduct the ILS task and the sub tasks by 25.5 seconds.

Triggering times for the sub tasks, except for the landing itself, were randomized between the trials and could potentially occur anywhere between the start of the trial and the landing. Within a trial the sub task triggering times were same for each subject. Tasks including radio transmissions were prepared as audio files and activated based on the elapsed time from the start of the trial. The audio files were played through the pilot's headset. When two or more ATC radio transmissions were to be triggered simultaneously, they were manually separated during the audio file preparation. The manual deconfliction of the sub task triggering times was limited to ATC radio transmissions only. Figure 1 illustrates how the starting range variable, the ILS task and the sub tasks were related. In Figure 1, the 5.5 NM and the 15.5 NM starting ranges are shown to highlight the difference between the trials of highest and the lowest temporal demand.

[Figure 1 near here]

Each trial was separated by a rest period lasting approximately three minutes. During the rest period the simulator was re-initialized for the next approach. The flying mission used for the study was treated as a flight curriculum's training sortie and the subjects prepared for the mission accordingly.

Procedure

The ratio between the time needed for task completion within a trial and the time available to complete them, or the time pressure, was used as an independent variable. To increase the sense of authenticity of the flying mission the subjects were free to select their individual piloting and problem solving strategies.

ECG (electrocardiograph) was recorded with Mind Media NeXus-10 MKII system supported by Biotrace+ software (version V2012C). Three electrodes were placed below the left (negative) and right (ground) clavicle and the left costal cartilage (positive) respectively. The Biotrace+ samples were exported to Kubios HRV 2.2 software for further analysis and RR interval artifact removal. All artifacts were detected and removed manually and noisy data was excluded from the further analysis. A specialist of internal medicine was consulted when necessary. ECG measuring, 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). After the last trial the subjects were asked about the level of intrusion caused by the NeXus-10 MKII system. None of the pilots reported intrusion of any kind.

A five minute pre-trial rest period was used to record the rest baseline HR/HRV and a three minute sample was taken from it for further analysis. During the rest baseline recording the subjects sat undisturbed in the simulator. As there are great differences in the individual cardiac activity, the subjects' cardiac responses to varying task demands were compared within each subject and not across subjects (Roscoe 1993).

A three minute sample was taken from the end of the baseline ILS and each trial's ECG recording. The values of the HR/HRV components recorded during each trial were compared to the ECG data from the subject's other trials, the baseline rest condition and the baseline ILS mission. HR/HRV components used for measuring PMWL are listed in Table 2.

[Table 2 near here]

The ILS task performance was rated between the ILS GS intercept range and the ILS DH using an official FinAF instrument check ride rating scale. The rating was based on a deviation from the target speed along with the LLZ and GS errors. The values of the rating scale ranged from 5 (best performance) to 0 (worst performance). The ILS scoring was conducted by using the simulator's mission replay. Between the GS interception range and the DH the mission playback was stopped at every 0.5NM (0.9km). While stopped, the deviations from the GS, LLZ and target speed were recorded and scored. The mean of the scores was used as an ILS task performance score. The ILS task performance was rated by a qualified F/A-18 examiner pilot. The examiner pilot's ILS performance scoring was based solely on the deviations from the target flight parameters (deviations from target speed, GS and LLZ). More subjectively rated aspects of performance (such as smoothness of aircraft handling) were not scored. The ILS task performance score and the values of the HR/HRV components were used as dependent variables.

Results

Treatment of Data

For a pilot to achieve a 1st class instrument rating on a real instrument check ride, s/he needs to achieve at least 60% of the ILS maximum score. In this study the threshold for the sub-standard performance was set to 60% of the absolute maximum ILS score, which mirrored the Finnish air force standards for the official instrument check ride.

The ILS scores were used to form three different performance categories. A *high performance* category was formed by selecting each pilot's ILS performance score from the baseline ILS mission. For the formulation of the *sub-standard performance* category only the trials with the sub-standard performance were considered. Out of these trials, the trial with the highest ILS performance score was selected for each subject. A *low performance* category was formed by selecting each pilot's trial that had the lowest ILS performance score. The ILS performance scores and the respective values of the HR and HRV components were plotted for each trial.

Data were analyzed using IBM™ SPSS™ software (version 22). The 5.5 NM (10.2 km) and the 6.5 NM (12 km) trials were left out from the final analysis as their durations were too short for a reliable HRV analysis. The 7.5 NM (13.9 km) trial was the most frequent candidate for the low performance category. It also had the most missing data, which reduced the number of subjects to 23. To increase the sample size, the 7.5 NM (13.9 km) trial was excluded from the analysis and replaced with the 8.5 NM (15.7 km) trial resulting in a sample size of 28 subjects.

Values of each subjects' HR/HRV components were retrieved for the analysis from four measurement points. The measurement points comprised of the last three minutes of the baseline rest, the baseline ILS mission (i.e., the high performance category), the trial with the highest sub-standard ILS performance score (i.e., the sub-standard performance category) and the trial that had the weakest ILS performance score (i.e., the low performance category).

The HR/HRV components' values were analyzed using the repeated measures MANOVA. Post-hoc pairwise comparisons were carried out with the paired t-test. Violation of sphericity and homoscedasticity was handled with the Greenhouse-Geisser correction when necessary.

Analysis

Table 3 presents the descriptive statistics of the HR/HRV components for each measurement point. There were statistically significant overall HR/HRV differences between performance categories; $F(7,21)=3.9$, $p<0.05$, $\eta^2_p=0.94$. Significant HR/HRV differences between performance categories were found on: MEANRR $F(3,81)=47.1$, $p<0.001$, $\eta^2_p = 0.64$; SDNN $F(3,81)=6.5$, $p<0.01$, $\eta^2_p=0.19$; MEANHR $F(3,81)=31.6$, $p<0.01$, $\eta^2_p=0.54$; NN50 $F(3,81)=18.1$, $p<0.001$, $\eta^2_p=0.40$; pNN50 $F(3,81)=8.4$, $p<0.01$, $\eta^2_p=0.24$; HRVTRI $F(3,81)=17.2$, $p<0.001$, $\eta^2_p=0.38$.

[Table 3 near here]

The results of the pairwise comparisons are summarized in Table 4. All HR/HRV components showed significant difference between the baseline rest and the high performance category ($p<0.05$). MEANRR ($p<0.01$) and MEANHR ($p<0.01$) were able to differentiate the high performance category from the sub-standard performance category. Figure 2 illustrates the MEANRR values across the measurement points. The task demand and the ILS performance changes between the sub-standard performance category and the low performance category were not differentiated by the HR/HRV components used.

[Table 4 near here]

[Figure 2 near here]

Discussion

This study extended the findings of earlier studies by investigating the associations between the PMWL and performance. The results can be used to support the identification of the PMWL redline in a realistic, or realistic simulated environment. To this end, the association between the HR/HRV components and the pilot's ILS performance was studied. With careful selection of PMWL metrics and thorough task analysis, it was possible to replicate the findings of the earlier mental workload related HR/HRV studies (see, e.g., Perusse-Lachance et al., 2012; Saperova and Dimitriev, 2014; Terkelsen et al., 2005; Tran et al., 2010; Li et al., 2009; Orsila et al., 2008; Monge, Gomez, and Molina, 2014; Taelman et al., 2011; Sun et al., 2012).

The results of this study clearly indicate that HR and HRV are sensitive to varying ratios between the time available and the time required for completing the tasks in the fighter aviation environment. Also, ECG monitoring was a relatively cheap method to assess HR and HRV, it has a high face validity and it did not generate intrusion of any kind. As this study had a within-subject, repeated measures design, most of the issues related to variations in skill and experience and the idiosyncratic heart rate responses were avoided. Furthermore, because of the way the three categories of the ILS performance were defined, every pilot effectively set their own datum.

The sub-standard performance category and the low performance category were not differentiated by the HR/HRV components used. Some subjects may have found the low performance category trial impossible and have eventually given up, i.e., they have invested less effort for the low performance category trial than they did for the sub-standard performance category trial. As a result the ILS performance was extremely poor while the HR/HRV response to increased task demand suggested lower PMWL.

A single measure is typically not enough to capture all the dimensions of the mental workload. When PMWL is considered, the number of usable measures of pilots' psychological responses is highly limited, mainly because of the limitations and restrictions a simulator environment puts on the measuring instrumentation. A real flying environment adds more restrictions to the available measures, as the instrumentation has to be flight-worthy and it may not compromise flight safety and mission success. If pilot responses to varying (mental) task demands are measured during a real flight, the physical demands of fighter aviation generate remarkable source of error. It is fortunate that there are certain check rides that are flown in a simulator and can therefore be considered as 'operational missions'. But even if multiple physiological measures could be utilized in a simulator environment, they may still fail to fully explain the relation of mental task demand, PWML and performance, as there are other constructs, such as SA, that may influence PMWL and performance (Durso and Alexander, 2010). These concepts, however, were not the emphasis of this study. While the ergonomics community waits for the more sophisticated objective measures of task demand, PMWL and SA to become available also for the flying environment, it has to settle to those - maybe less sophisticated - measures accepted (on a case by case basis) by the aviation authorities. In that sense, HR and HRV have both justified their place among the usable measures of task demand in aviation domain (see for example: Magnusson 2002, Ylönen et al. 1997, Jorna 1993). While the multidimensional nature of PMWL is acknowledged, there are many studies that have been able to utilize HR and HRV as measures of PMWL (Svensson et al. 1997; Magnusson 2002; Veltman 2002; Lahtinen et al. 2007). This study showed that the MEANRR component of the HRV is a strong candidate when different measures of PMWL are being considered for a fighter aviation environment.

The main finding of this study was the MEANRR's ability to differentiate the high performance ILS approaches from the sub-standard performance approaches. In this context, such a finding has not been previously reported. The results of this study provide encouraging basis for testing HR/HRV components' sensitivity on more realistic and complex fighter missions. The MEANRR's sensitivity to varying task demands must be tested on a real instrument check ride where the task demands and the pilot's level of mental effort are realistic. If these results are equally positive, the HR/HRV measures must then be tested against the pilot performance on a realistic, simulated combat mission, as combat missions represent the ultimate task demands a fighter pilot has to cope with.

Acknowledgments

The authors are thankful to the two anonymous reviewers who provided valuable comments on an earlier version of this paper.

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

Aasman, J., G. Mulder and L.J. Mulder.1987. "Operator Effort and the Measurement of Heart-Rate Variability." *Human Factors* 29 (2): 161-170.

Anupama, D., A. N. Deepak, S. Nallulwar and V. Khode. 2014. "Time Domain Measures of Heart Rate Variability during Acute Mental Stress in Type 2 Diabetics-A Case Control Study." *National Journal of Physiology, Pharmacy and Pharmacology* 4 (1): 34-38.

Ayaz, H., B. Willems, S. Bunce, P. A. Shewokis, K. Izzetoglu, S. Hah, A. Deshmukh, B. Onaral. 2010. "Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors." In *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*, edited by T. Marek, W. Karwowski and V. Rice, 21-32. London: Taylor & Francis.

Berkun, M. M. 1964. "Performance Decrement Under Psychological Stress." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 6 (1): 21-30.

Brookhuis, K. A., D. De Waard and S. H. Fairclough. 2003. "Criteria for Driver Impairment." *Ergonomics* 46 (5): 433-445.

Burcu, C., B. Arnrich, R. La Marca, and G. Tröster. 2013. "Monitoring of Mental Workload Levels during an Everyday Life Office-Work Scenario." *Personal and Ubiquitous Computing* 17 (2): 229-239.

Camm, A. J., M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy and R. E. Kleiger. 1996. "Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology." *Circulation* 93 (5): 1043-1065.

Collet, C., E. Salvia and C. Petit-Boulanger. 2014. "Measuring Workload with Electrodermal Activity during Common Braking Actions." *Ergonomics* 57 (6): 886-896.

Driskell, J. E. and E. Salas. 1991. "Group Decision Making Under Stress." *Journal of Applied Psychology* 76 (3): 473.

Grandjean, E. and K. H. E. Kroemer. 1997. *Fitting the Task to the Human: A Textbook of Occupational Ergonomics*. New York, NY: Taylor & Francis.

Hayward, C., H. Patel, S. Manohar, J. Waton, L. Middleton, A. Lyon, R. Sutton and S. Rosen. 2014. "39 Short Time Window Analysis of Heart Rate Variability during Tilt Table Testing Provides Insights in to the Timing of Changes in the Autonomic Nervous System." *Europace* 16 (suppl 3): iii16-iii16.

Iani, C., D. Gopher, A. J. Grunwald and P. Lavie. 2007. "Peripheral Arterial Tone as an on-Line Measure of Load in a Simulated Flight Task." *Ergonomics* 50 (7): 1026-1035.

Jorna, PGAM. 1992. "Spectral Analysis of Heart Rate and Psychological State: A Review of its Validity as a Workload Index." *Biological Psychology* 34 (2): 237-257.

Jorna, PGAM. 1993. "Heart Rate and Workload Variations in Actual and Simulated Flight." *Ergonomics* 36 (9): 1043-1054.

Kaber, D. B. and M. R. Endsley. 2004. "The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task." *Theoretical Issues in Ergonomics Science* 5 (2): 113-153.

Lahtinen, T., J. P. Koskelo, T. Laitinen and T. K. Leino. 2007. "Heart Rate and Performance during Combat Missions in a Flight Simulator." *Aviation, Space, and Environmental Medicine* 78 (4): 387-391.

Lieberman, H. R., G. P. Bathalon, C. M. Falco, C. A. Morgan, P. J. Niro and W. J. Tharion. 2005. "The Fog of War: Decrements in Cognitive Performance and Mood Associated with Combat-Like Stress." *Aviation, Space, and Environmental Medicine* 76 (Supplement 1): C7-C14.

Magnusson, S. 2002. "Similarities and Differences in Psychophysiological Reactions between Simulated and Real Air-to-Ground Missions." *The International Journal of Aviation Psychology* 12 (1): 49-61.

Mallis, M. M., and D. F. Dinges. 2004. “Monitoring Alertness by Eyelid Closure.” In *Handbook of Ergonomics and Human Factors Methods*, edited by N. Stanton, A. Hedge, H. W. Hendrick, K. A. Brookhuis, and E. Salas. Chap. 25. London: Taylor & Francis.

Mavin, T. J. and W-M. Roth. 2014. "A Holistic View of Cockpit Performance: An Analysis of the Assessment Discourse of Flight Examiners." *The International Journal of Aviation Psychology* 24 (3): 210-227.

Morris, C. H. and Y. K. Leung. 2006. "Pilot Mental Workload: How Well do Pilots really Perform?" *Ergonomics* 49 (15): 1581-1596.

Noel, J. B., K. W. Bauer Jr, and J. W. Lanning. 2005. "Improving Pilot Mental Workload Classification through Feature Exploitation and Combination: A Feasibility Study." *Computers & Operations Research* 32 (10): 2713-2730.

O'hare, D. 2000. "The ‘Wheel of Misfortune’: A Taxonomic Approach to Human Factors in Accident Investigation and Analysis in Aviation and Other Complex Systems." *Ergonomics* 43 (12): 2001-2019.

O'Hare, D., M. Wiggins, R. Batt, and D. Morrison. 1994. "Cognitive Failure Analysis for Aircraft Accident Investigation." *Ergonomics* 37 (11): 1855-1869.

Orsila, R., M. Virtanen, T. Luukkaala, M. Tarvainen, P. Karjalainen, J. Viik, M. Savinainen, and C-H. Nygard. 2008. "Perceived Mental Stress and Reactions in Heart Rate variability– a Pilot Study among Employees of an Electronics Company." *International Journal of Occupational Safety and Ergonomics (JOSE)* 14 (3): 275-283.

Pérusse-Lachance, E., A. Tremblay, J-P. Chaput, P. Poirier, N. Teasdale, V. Drapeau, C. Sénécal and P. Brassard. 2012. "Mental Work Stimulates Cardiovascular Responses through a Reduction in Cardiac Parasympathetic Modulation in Men and Women." *Bioenergetics* S1 001: 1-6.

Rasmussen, J. and A. Jensen. 1974. "Mental Procedures in Real-Life Tasks: A Case Study of Electronic Trouble Shooting." *Ergonomics* 17 (3): 293-307.

Roman-Liu, D., I. Grabarek, P. Bartuzi, and W. Choromański. 2013. "The Influence of Mental Load on Muscle Tension." *Ergonomics* 56 (7): 1125-1133.

Roscoe, A. H. 1992. "Assessing Pilot Workload. Why Measure Heart Rate, HRV and Respiration?" *Biological Psychology* 34 (2): 259-287.

Roscoe, A. H. 1993. "Heart Rate as a Psychophysiological Measure for in-Flight Workload Assessment." *Ergonomics* 36 (9): 1055-1062.

Saperova, E. and D. Dimitriev. 2014. "The Effects of Mood State and Psychological Well-being on Heart Rate Variability during Mental Stress." *International Journal of Psychophysiology* 94 (2): 232.

Shappell, S. A. and D. A. Wiegmann. 1997. "A Human Error Approach to Accident Investigation: The Taxonomy of Unsafe Operations." *The International Journal of Aviation Psychology* 7 (4): 269-291.

Sheridan, T. B. and R. W. Simpson. 1979.*Toward the Definition and Measurement of the Mental Workload of Transport Pilots*.

Stuiver, A., K. A. Brookhuis, D. de Waard and B. Mulder. 2014. "Short-Term Cardiovascular Measures for Driver Support: Increasing Sensitivity for Detecting Changes in Mental Workload." *International Journal of Psychophysiology* 92 (1): 35-41.

Sun, F-T., C. Kuo, H-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. 2012. "Activity-Aware Mental Stress Detection using Physiological Sensors." In *Mobile Computing, Applications, and Services*, 211-230. Santa Clara, CA: Springer.

Svensson, E., M. Angelborg-Thanderez, L. Sjöberg, and S. Olsson. 1997. "Information Complexity-Mental Workload and Performance in Combat Aircraft." *Ergonomics* 40 (3): 362-380.

Taelman, J., S. Vandeput, E. Vlemincx, A. Spaepen, and S. Van Huffel. 2011. "Instantaneous Changes in Heart Rate Regulation due to Mental Load in Simulated Office Work." *European Journal of Applied Physiology* 111 (7): 1497-1505.

Terkelsen, A. J., H. Mølgaard, J. Hansen, O. K. Andersen and T. S. Jensen. 2005. "Acute Pain Increases Heart Rate: Differential Mechanisms during Rest and Mental Stress." *Autonomic Neuroscience* 121 (1–2): 101-109.

Tran, B. W., A. Papoiu, C. V. Russoniello, H. Wang, T. S. Patel, Y-H. Chan and G. Yosipovitch. 2010. "Effect of Itch, Scratching and Mental Stress on Autonomic Nervous System Function in Atopic Dermatitis." *Acta Dermato-Venereologica* 90 (4): 354-361.

Veltman, J. A. 2002. "A Comparative Study of Psychophysiological Reactions during Simulator and Real Flight." *The International Journal of Aviation Psychology* 12 (1): 33-48.

Veltman, J. A. and A. W. K. Gaillard. 1998. "Physiological Workload Reactions to Increasing Levels of Task Difficulty." *Ergonomics* 41 (5): 656-669.

Vitense, H. S., J. A. Jacko and V. K. Emery. 2003. "Multimodal Feedback: An Assessment of Performance and Mental Workload." *Ergonomics* 46 (1-3): 68-87.

Wilson, G. F. 1993. "Air-to-Ground Training Missions: A Psychophysiological Workload Analysis." *Ergonomics* 36 (9): 1071-1087.

Xhyheri, B., O. Manfrini, M. Mazzolini, C. Pizzi and R. Bugiardini. 2012. "Heart Rate Variability Today." *Progress in Cardiovascular Diseases* 55 (3): 321-331.

Ylönen, H., H. Lyytinen, T. Leino, J. Leppäluoto and P. Kuronen. 1997. "Heart rate responses to real and simulated BA Hawk MK 51 flight". *Aviation, space and environmental medicine* 66 (7): 601-605.

Young, M. S., K. A. Brookhuis, C. D. Wickens and P. A. Hancock. 2015. "State of Science: Mental Workload in Ergonomics." *Ergonomics* 58 (1): 1-17.

Zakay, D. and J. Shub. 1998. "Concurrent Duration Production as a Workload Measure." *Ergonomics* 41 (8): 1115-1128.

Zhibin, L., H. Snieder, S. Su, X. Ding, J. F. Thayer, F. A. Treiber and X. Wang. 2009. "A Longitudinal Study in Youth of Heart Rate Variability at Rest and in Response to Stress." *International Journal of Psychophysiology* 73 (3): 212-217.

Task	Required Pilot Action
ATC clearance 1	Read back the clearance. Copy the clearance on a knee pad.
ATC clearances 2-3	Read back the clearance and swith to an indicated frequency
ATC clearance 4	Read back the clearance.
ATC inquiries 1-3	Check the requested flight parameter and report it to the ATC
ATC directives 1-3	Set the cockpit instrument to a directed value
Engine warnings 1-4	Initiate a related emergency procedure
Icing warning	Select an anti-ice switch to 'ON'
Flight control system warning	Initiate a related emergency procedure
Environmental control system warning	Initiate a related emergency procedure
Fuel level warning	Reset a fuel level warning
Mental task	Calculate the landing speed in km/h based on an indicated fuel state
Display failure	Swith to use an alternate display
IAC radio frequencies 1-6	Tune a radio preset to a frequency indicated in the IAC
IAC parameters 1-3	Copy a value indicated in the IAC to a knee pad
Land	Fly from DH to touchdown point and make a full stop landing
ILS localizer	Maintain an approach course in accordance with the ILS localizer
ILS glidepath	Maintain a glidepath in accordance with the ILS glidepath
Approach speed	Maintain a minimum approach speed

Table 1. The sub tasks and the components of the ILS task. The ILS task components have been shaded.

Table 1

Measure	Unit	Description	Expected change due to PMWL increase
MEANRR	[ms]	The mean of NN intervals	Decrease
SDNN	[ms]	The standard deviation of NN intervals	Increase
MEANHR	[1/min]	The mean heart rate	Increase
RMSSD	[ms]	The square root of the mean squared differences between successive NN intervals	Increase
NN50	[count]	The number of successive NN interval pairs that differ more than 50ms	Decrease
pNN50	[%]	The NN50 divided by the total number of NN intervals	Decrease
HRVTRI	[-]	The integral of the NN interval density distribution divided by the maximum of the distribution.	Decrease

Table 2

1

Table 3. Means and standard deviations of the HR/HRV components at the measurement points (N=28).

		Baseline rest		High performance category		Satisfactory performance category		Low performance category	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
MEANRR	[ms]	842.8	168.3	724.7	148.5	660.3	124.8	662.1	124.0
SDNN	[ms]	70.6	22.6	50.9	26.8	49.7	28.3	50.0	26.8
MEANHR	[1/min]	74.6	15.3	86.6	17.3	94.6	17.2	94.6	17.6
RMSSD	[ms]	45.1	22.2	30.9	27.3	31.4	31.9	33.1	34.8
NN50	[count]	81.1	60.6	20.7	30.9	24.6	51.0	27.4	52.7
pNN50	[%]	21.3	16.0	9.2	12.6	9.2	16.2	9.9	16.3
HRVTRI	[-]	15.9	5.0	10.5	4.5	10.6	5.0	10.6	4.2

Table 3

Table 4 The values of the test statistics and changes in pairwise comparisons between measurement points; ***p< 0.001; **p< 0.01; *p< 0.05 (N=28).

		Baseline rest - High performance category			High performance category - Satisfactory performance category			Satisfactory performance category - Low performance category		
		Mean Diff.	Std. Error	Sig.	Mean Diff.	Std. Error	Sig.	Mean Diff.	Std. Error	Sig.
MEANRR	[ms]	103.0	11.1	0.000***	58.358	16.2	0.001***	3.0	6.4	0.647
SDNN	[ms]	15.9	4.7	0.002**	2.4	4.9	0.628	-0.7	3.0	0.813
MEANHR	[1/min]	-11.4	1.5	0.000***	-6.8	2.1	0.004**	-1.0	1.2	0.4
RMSSD	[ms]	9.4	4.2	0.035*	1.1	4.6	0.818	-2.9	2.0	0.160
NN50	[count]	51.5	8.2	0.000***	-1.7	5.8	0.768	-4.6	3.3	0.183
pNN50	[%]	9.4	2.1	0.000***	0.9	2.0	0.661	-1.1	1.0	0.250
HRVTRI	[-]	4.8	0.8	0.000***	0.1	0.8	0.851	0.0	0.6	0.957

Table 4

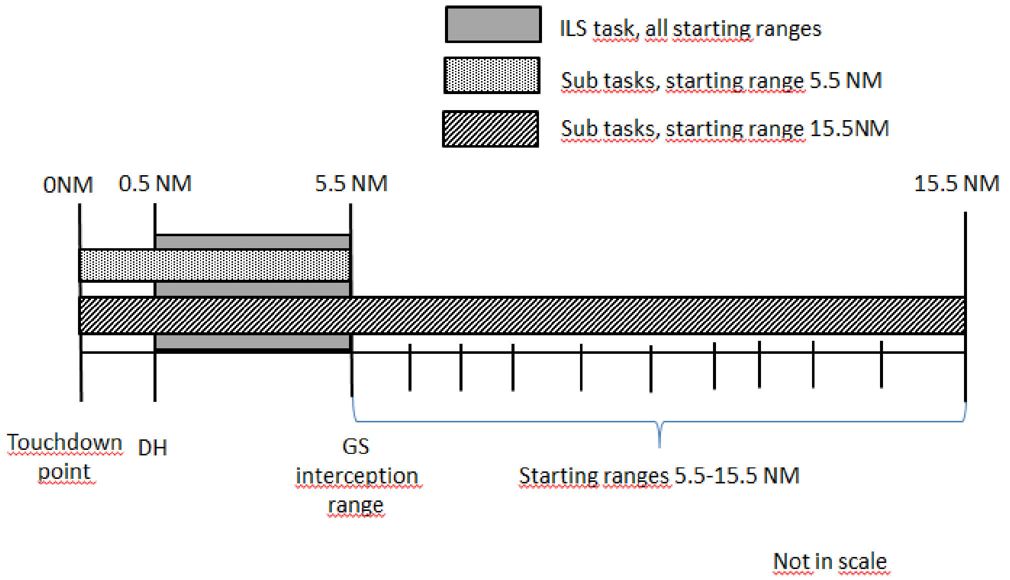
ILS task, all starting ranges
Sub tasks, starting range 5.5 NM
Sub tasks, starting range 15.5NM
0NM
0.5 NM
5.5 NM
15.5 NM
Touchdown point
DH
GS interception range
Starting ranges 5.5-15.5 NM
Not in scale

Figure 1

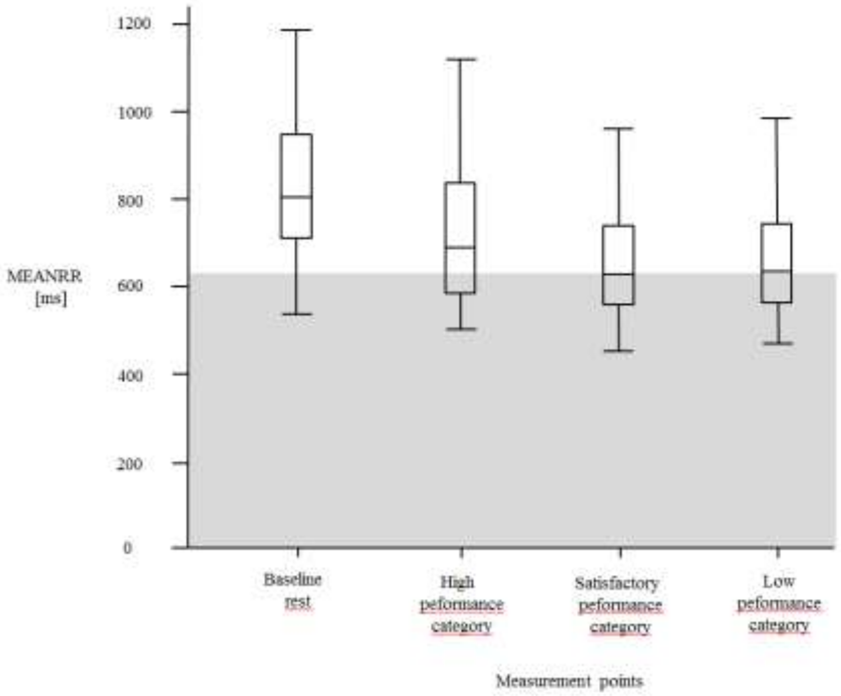
1200
1000
800
600
400
200
0
MEANRR
[ms]
Baseline
rest
High
peformance
category
Satisfactory
peformance
category
Low
peformance
category
Measurement points

Figure 2