Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight

De Rivecourt, M.; Kuperus, M.N.; Post, W.J.; Mulder, L.J.M.

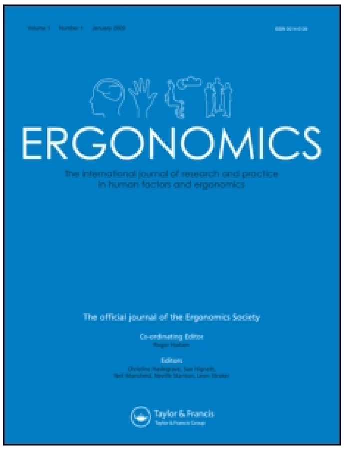
Published in:
Ergonomics

DOI:
10.1080/00140130802120267

Publication date:
2008

Citation for published version (APA):
De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. Ergonomics, 51(9), 1295-1319. DOI: 10.1080/00140130802120267

This article was downloaded by: *[Bibliotheek der Rijksuniversiteit]*
On: *31 March 2009*
Access details: *Access Details: [subscription number 793222190]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Ergonomics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713701117

Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight

M. De Rivecourt [a]; M. N. Kuperus [a]; W. J. Post [b]; L. J. M. Mulder [a]
[a] University of Groningen, The Netherlands [b] University Medical Center Groningen, The Netherlands

Online Publication Date: 01 September 2008

To cite this Article De Rivecourt, M., Kuperus, M. N., Post, W. J. and Mulder, L. J. M.(2008)'Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight',Ergonomics,51:9,1295 — 1319

To link to this Article: DOI: 10.1080/00140130802120267

URL: http://dx.doi.org/10.1080/00140130802120267

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Ergonomics
Vol. 51, No. 9, September 2008, 1295–1319

Taylor & Francis
Taylor & Francis Group

Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight

M. De Rivecourt[a]*, M.N. Kuperus[a], W.J. Post[b] and L.J.M. Mulder[a]

[a]*University of Groningen, The Netherlands;* [b]*University Medical Center Groningen, The Netherlands*

This study examines effects of momentary changes in mental effort on cardiovascular and eye activity measures. A total of 19 male pilots performed an instrument flight task. Task load was manipulated by having the pilots perform flight manoeuvres of varying complexity. Multilevel analyses demonstrated clear effects of momentary changes in mental effort on both the cardiovascular and the eye activity measures. An increase in task load resulted in an increase of heart rate and a decrease in heart rate variability, mean dwell time and fixation duration. Heart rate differentiated between resting period and task execution. Heart rate variability from short data segments provided more insight in intermediate levels of mental effort. The eye activity measures were sensitive to intermediate levels of mental effort as well. Attitude changes resulted in an increase of mean dwell time and mean fixation duration. Task analysis is required to use eye measures as valid indices of mental effort. Having indications of the effects of changing mental demands during daily work of operators is of great importance nowadays. This paper presents an approach to estimate such effects on the basis of heart rate and eye activity measures. In particular, the use of averaged short-term heart rate variability measures is a relatively new aspect.

Keywords: mental workload; mental effort; heart rate; heart rate variability; eye activity; subjective ratings

1. Introduction

1.1. Mental workload and mental effort

Operating an airplane is characterised by large fluctuations in task demands. During approach and landing, task demands are high because a large number of critical decisions have to be taken in a short period of time. During the crossing of the Atlantic, however, task demands are low because the pilot only has to monitor the flight control computer and make an occasional position report. The term ‘mental workload’ is often used to indicate to which degree task demands affect the information-processing capacity of the operator. Gopher and Donchin (1986) defined mental workload as the difference between the available and the required capacity of the information-processing system to perform a task at any given time. When the required capacity exceeds the available capacity, the operator is overloaded and performance will decrease. For example, a pilot might no longer be able to process all critical information during landing with a single

---

*Corresponding author. Email: michel@derivecourt.nl

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140130802120267
http://www.informaworld.com

engine failure at an unfamiliar airport at night. On the other hand, performance may also decrease when the available capacity exceeds the required capacity (Johnson and Proctor 2004). In this case the system is under loaded, which might make it more difficult for the operator to remain vigilant and process all relevant information. A pilot on a long-range navigation flight, for instance, might not notice that the fuel quantity indicator is decreasing at an abnormally high rate. To guarantee an optimal level of performance, the required capacity should match the available capacity of the operator. Therefore, pilot performance could benefit from a system that continuously tries to adapt the task demands to the operator's workload (Endsley and Kaber 1999). For instance, during approach and landing, the information of systems that are operating within limits could be suppressed. In contrast, task demands could be increased during intercontinental flight by increasing the information about the functioning of less-critical systems. However, to be able to optimise the level of mental workload such a system would be dependent on valid and reliable measures of current task demands and operator capacity.

In general, a more demanding task will require higher mental effort to perform such a task adequately. However, there are many task situations in which overall task performance is only minimally affected when less effort is invested (Hockey 1997). This gives large flexibility in human performance during daily life. Experienced operators only expend the highest mental effort when it is really necessary.

Mental effort can be used as an indirect measure of operator capacity. An increase in mental effort, for instance, will automatically result in a decrease of the remaining capacity of the operator. In aviation research, mental effort has been studied mostly by varying task demands in a simulated flight task. In such an approach it is assumed that an increase in task demands is accompanied by an increase in mental effort (Mulder 1986, Vicente *et al.* 1987, Cnossen *et al.* 2004). Therefore, it is important that motivated participants are used, who are willing to exert effort during task performance (Veltman 2002).

In some studies, task demands are varied by the introduction of a secondary task. Examples are the use of a verbal loading task by Tole *et al.* (1982) and the use of a computational task by Callan (1998). In other studies, no additional tasks are used. Participants have to perform a realistic flight task and the resulting variations in task demands are assessed by means of subjective ratings or task analyses. In a study by Veltman (2002), for example, task demands are varied by having the participants fly six scenarios with increasing levels of difficulty.

1.2. Physiological measures

Mental effort can be measured by performance, self-report (subjective) and physiological measures. However, subjective measures have drawbacks. According to Veltman and Gaillard (1996a), participants are having difficulties distinguishing task demands from invested effort. Physiological measures are the alternative. Several physiological measures have shown to be sensitive to mental effort. Heart rate (HR), for instance, was found to increase with an increase in mental effort (e.g. Roscoe 1992, Wilson 2001, Svensson and Wilson 2002). Another widely used cardiovascular measure is HR variability (HRV). The HR spectrum (representing all HRV) can be divided into three frequency bands (Mulder 1985), which are associated with biological control mechanisms (Kramer 1991). The low-frequency band (0.02–0.06 Hz) is mostly related to the regulation of body temperature (Kitney 1980) and gradual changes in task characteristics (Mulder and Mulder 1987). The mid-frequency band (0.07–0.14 Hz) is mainly related to fluctuations in short-term blood pressure control (baroreflex) and strongly related to momentary changes in blood pressure (van Roon *et al*. 2004). Finally, the high-frequency band (0.15–0.40 Hz) is primarily related to respiratory fluctuations (Kramer 1991). Laboratory studies have shown that the mid-frequency band is most consistently related to mental effort (Mulder 1985); however, both the mid- and high-frequency bands have proved to be sensitive to mental effort. In general, HRV in these two bands was found to decrease with an increase in mental effort (e.g. Jorna 1993, Veltman and Gaillard 1993, Althaus *et al*. 1999, Boucsein and Backs 2000).

In laboratory studies that use mental loading tasks, HRV tends to be more sensitive to variations in mental effort than HR (Aasman *et al*. 1987). In aviation research, however, HR often provides insight in intermediate variations in mental effort (Veltman and Gaillard 1996b, Wilson 2002), while HRV can only be used to distinguish the resting period from task execution (Jorna 1992, Wilson 1992). These differences in research results might be explained by the relatively high workload conditions in these studies.

Several eye activity-related measures were also found to provide useful information about mental effort. In most studies an increase in mental effort was associated with an increase in pupil diameter and a decrease in the number of eye blinks (for a review, see May *et al*. 1990). In addition, mean dwell time (total time looking in a specific region) and fixation duration (time between saccades) tend to decrease with an increase in task demands and visual complexity (Duchowski 2002). Secondary mental tasks, however, generally lead to an increase of mean fixation duration and dwell time on the relevant displays (Hoogeboom and Hilburn 2001). Aviation research provides similar results. Svensson and Wilson (2002) found that mean fixation duration decreases as mental workload increases. In contrast, when task demands were increased by secondary mental tasks (Tole *et al*. 1982, Callan 1998) mean fixation duration was found to increase. It appears that the effect of mental effort on eye activity measures is highly dependent on specific task characteristics. Therefore, it seems necessary to perform an extended task analysis before eye activity measures can be used as indices for mental effort.

1.3. Momentary changes in mental effort

In most mental effort studies, the effect of varying task demands is assessed over relatively long time periods (minutes as opposed to seconds). As a result, it is largely unclear to which degree physiological measures can provide a valid indication of momentary changes in invested effort. Information about these momentary changes is of crucial importance for a system that aims to maintain mental workload at an optimal level during human performance. Most human operations are characterised by unexpected situations, which will require the operator to make critical decisions in a matter of seconds. A fire warning during flight, for instance, will lead to an instantaneous peak in invested effort by a pilot. On the other hand, the invested effort might decrease rapidly as soon as the pilot finds out that it is only a false alarm.

With regard to HR measures, Mulder (1992) suggests the use of so-called ‘spectral profiles’ as indicators of the spectral power in a specified frequency band as a function of time. Using this technique, momentary changes in HR and HRV can be related to changes in task demands and performance measures.

1.4. *Purpose of the study*

The main aim of the present study is to determine to which degree HR and eye activity measures reflect momentary changes in mental effort. To this end, the participants will perform an instrument flight task. The resulting variations in task demands will be estimated by means of subjective ratings and a task analysis. Subsequently, the effect of these varying task demands (and mental effort) on the physiological measures will be determined.

Three cardiovascular and two eye activity measures were selected. The cardiovascular measures consist of both HR and HRV (mid- and high-frequency bands), because these measures have repeatedly shown to be sensitive to changes in mental effort. In line with previous research results, an increase in HR and a decrease in HRV (mid- and high-frequency bands) are expected to indicate an increase in mental effort (e.g. Mulder 1992). The eye activity measures consist of mean dwell time and fixation duration. Both eye activity measures are selected because they seem to be most suited to provide information about changes in mental effort during an instrument flight task. An instrument flight task can be regarded as a tracking task. Participants scan their flight instruments to detect deviations from the prescribed flight parameters and make adjustments accordingly. Mean dwell time and fixation duration are directly related to this scanning behaviour. With increasing task demands, participants are expected to increase their effort to scan for the resulting deviations from the prescribed parameters. This should lead to an increase in scanning rate and a resulting decrease in mean dwell time and fixation duration. This expectation is in line with a study of van Orden *et al.* (2001), who found a decrease in mean fixation duration with an increase in tracking error on a visual compensatory tracking task.

In this study, a two-step process will be followed. First, the effects of an increase in task demands for relatively long time periods will be assessed by comparing six consecutive flight segments. When the results are in line with the stated expectations (and previous research results), the feasibility of the proposed measures to detect momentary changes in task demands during these segments will be evaluated.

2. Methods

2.1. *Participants*

In total, 19 male participants in the range of 17 to 27 years were recruited from the Royal Netherlands Air Force pilot selection and training programme. Since the pilot training provides a competitive environment, the participants were expected to be motivated to attain maximum task performance.

The participants had prior experience in a simulator, but had not yet started their military flight training. Although a few participants had received prior civil flight training, they all had less than 100 flight hours and were relatively inexperienced.

2.2. *Simulator*

The study took place at the Center for Man and Aviation in Soesterberg, the Netherlands. The instrument flight task was executed in the ALSIM AL 100 Flight Trainer (Figure 1). This is a fixed-base flight simulator, which is intended to help pilots become familiar with the basics of instrument flying. During flight, instrument meteorological conditions were simulated. The visual display was deactivated and the participants could use only the computerised flight instruments.

Figure 1. The ALSIM AL 100 Flight Trainer.

2.3. *Subjective ratings*

The rating scale of mental effort (RSME) was used to measure subjective effort. This scale ranges from 0 to 150 and has nine descriptive labels, such as 'not effortful' (score 2), 'fairly effortful (score 57) and 'awfully effortful' (score 113). Participants could indicate their invested effort by marking this scale at the appropriate point.

2.4. *Flight task*

The experimental task consisted of an instrument flight profile of 28.5 min, during which the participants had to execute flight manoeuvres at a specific time. The instructions for the execution of the instrument flight task are included in Table 1. In summary, the flight profile consisted of a take-off to the north, a left horizontal turn, a right climbing turn, a left descending turn, a straight climb directly followed by a left climbing turn and a left descending turn directly followed by a right climbing turn. Data recording started once the participants had established straight and level flight at 1000 feet with an airspeed of 95 knots. A graphic representation of the vertical (top) and horizontal (bottom) flight profile is depicted in Figure 2.

The flight profile was divided into 37 flight elements. Each of these elements can be described by a specific set of parameters (or a transition to this set of parameters). The first

Table 1. Instructions for the execution of the instrument flight task.

Time	Manoeuvre
00:00	Takeoff heading North and climb with airspeed 75 knots Level off at 1000 feet with 95 knots (START DATA RECORDING)
04:00	Climb with 500 feet per min and an airspeed of 75 knots Level off at 2000 feet with airspeed 95 knots
08:00	Decelerate to airspeed 80 knots
09:00	Descend with airspeed 80 knots and a descent rate of 500 feet per min Level off at 1000 feet with airspeed 95 knots
13:00	Turn 270° over left (turn rate 3° per s) and rollout on heading East
16:30	Turn 360° over right (turn rate 3° per s) while climbing with 500 feet per min Maintain airspeed 75 knots and a maximum bank angle of 15° Level off at 2000 feet with airspeed 105 knots on heading East
20:30	Decelerate to airspeed 80 knots
21:00	Turn 270° over left (turn rate 3° per s) while descending with 500 feet per min Maintain airspeed 80 knots Level off at 1250 feet with an airspeed of 90 knots on heading South
23:00	Climb with 500 feet per min and an airspeed 75 knots Level off at 2250 feet and airspeed 80 knots
23:30	Turn 270° over left (turn rate 3° per s) while climbing 500 feet per min Maintain airspeed 75 knots and a maximum bank angle of 15° Level off at 2250 feet and airspeed 80 knots on heading West
25:30	Turn 90° over left (turn rate 3° per s) while descending with 500 feet per min Maintain airspeed 80 knots Level off at 2000 ft with airspeed 80 knots on heading South
26:00	Turn 90° over right (turn rate 3° per s) while climbing with 500 feet per min Maintain airspeed 75 knots and a maximum bank angle of 15° Level off at 2250 feet with airspeed 95 knots on heading West
28:30	END DATA RECORDING

flight element, for example, is straight and level flight at 1000 feet with an airspeed of 95 knots (see Table 1: time 00:00, START DATA RECORDING). This flight element is followed at time 04:00 by the transition to a climb (element 2) and a climb with 500 feet per min and an airspeed of 75 knots (element 3). Finally, the last two flight elements are the transition to straight and level flight (element 36) and straight and level flight at 2250 feet with an airspeed 95 knots on heading west (element 37). The duration of the flight elements is relatively short and ranges from 100 down to 5 s. Distinguishing these elements is relevant for determining the feasibility of physiological measures as indices of momentary changes in mental effort.

The RSME was presented after flight to avoid interference with the execution of the task. Therefore, invested effort had to be rated from memory. This made a reliable rating for each of the 37 flight elements unlikely. Consequently, only task analysis was used to estimate momentary variations in task demands. During the task analysis, all 37 flight elements were assigned to one of four task demand levels (with increasing level indicating more demanding elements). The criterion for the assignment to a specific level was the number of variables that had to be checked against time by the participants. In straight and level flight, the heading, altitude and airspeed are constant

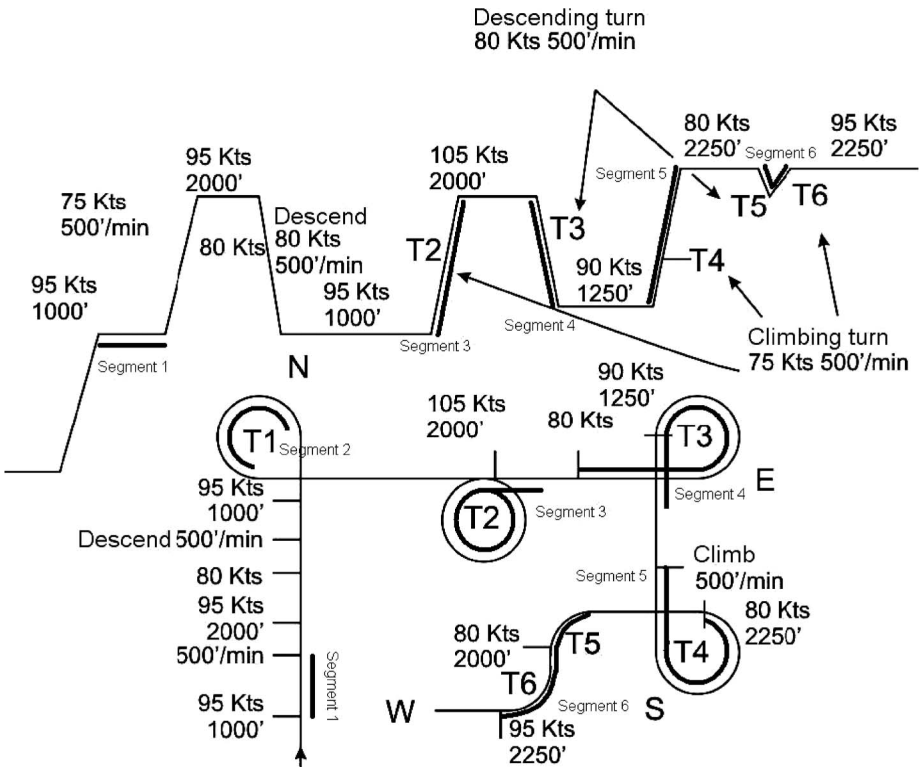


Figure 2. Graphic representation of the flight profile. The climbing and descending patterns are depicted in the top picture, while the pattern of horizontal changes is depicted in the bottom picture.

and the rate of change of the number of variables that has to be checked against time is zero. Therefore, each straight and level flight element was assigned to task demand level 1 (the least difficult level). During climbs, descents and heading changes the number of variables that has to be checked against time is one. For instance, when the participants had to climb from 1000 feet to 2000 feet at minute 04:00 (see Table 1) with a prescribed climb rate of 500 feet per min, they had to check if they passed 1500 feet at minute 05:00 and adjust their rate of climb to reach 2000 feet at minute 06:00. Therefore, all flight elements that consisted of climbs, descents or heading changes were assigned to task demand level 2. During climbing and descending turns the number of variables that has to be checked against time is two (heading and altitude). Therefore, all flight elements that consisted of climbing and descending turns were assigned to task demand level 3. The last flight elements that have to be assigned to a specific task demand level are the transitions between the previously mentioned flight elements. Here, the earlier described criterion was not used, because the characteristics of a transition are fundamentally different from all other flight elements. During the previously described flight elements, the aircraft generally is in a state of equilibrium; the aircraft is stable and only small control forces are necessary to maintain the desired parameters. During transitions, however, the equilibrium is lost and has to be re-established. This is a relatively difficult task, which requires continuous adjustments. For instance, in the transition to a climb, the aircraft is decelerating from cruise airspeed to climb airspeed. As a result, the pitch angle of the aircraft has to be continually adjusted to maintain the prescribed climb rate. Equilibrium can only be re-established once the airspeed has stabilised at the prescribed climb airspeed. Therefore, transitions were considered to be the most difficult flight elements and were assigned to task demand level 4. It is important to note that task demand level is an ordinal variable. As a result, task demands are expected to increase with increasing task demand level. The differences in task demands between the respective levels, however, are not necessarily equal.

For the assessment of the gradual changes in task demands, six flight segments were composed out of the 37 flight elements. Here, both the results of the task analysis and the RSME were used to estimate the variations in task demands. Because the participants had to be able to rate their invested effort from memory, six distinct segments of the flight were selected (see Figure 2): straight and level flight (segment 1); the left horizontal turn (segment 2); a right climbing turn (segment 3); a deceleration followed by a left descending turn (segment 4); a climb followed by a left climbing turn (segment 5); a left descending turn followed by a right climbing turn (segment 6). The instrument flight task was designed to produce increasing task demands during flight. With the exception of segment 5, this design was supported by the weighted averages of the task demand level during each consecutive flight segment (see Table 2). However, task demand level is an ordinal variable and can only be used as an estimation of the average task demand. Therefore, the results of the subjective ratings were decisive in determining the average task demands during the flight segments.

Table 2. Average task demand level per flight segment, determined by weighing the duration of the flight elements (1 = low demand, 4 = high demand).

Segment	Flight element	Description of flight elements (duration in parentheses)	Task demand level	Weighted average
1	1	Straight and level flight (85 s)	1	1.0
2	12	Transition (10 s)	4	2.4
	13	Horizontal turn (75 s)	2	
	14	Transition (20 s)	4	
3	16	Transition (15 s)	4	3.3
	17	Climbing turn (100 s)	3	
	18	Transition (30 s)	4	
4	20	Transition (25 s)	4	3.4
	21	Straight and level flight (5 s)	1	
	22	Transition (15 s)	4	
	23	Descending turn (70 s)	3	
	24	Transition (25 s)	4	
5	26	Transition (15 s)	4	3.2
	27	Climb (15 s)	2	
	28	Transition (10 s)	4	
	29	Climbing turn (75 s)	3	
	30	Transition (20 s)	4	
6	32	Transition (15 s)	4	3.8
	33	Descending turn (10 s)	3	
	34	Transition (20 s)	4	
	35	Climbing turn (10 s)	3	
	36	Transition (30 s)	4	

2.5. Physiological measures

2.5.1. Cardiovascular measures

Electrocardiographic (ECG) data were collected from electrodes placed over the sternum (ground electrode) and between the first and second floating rib on both sides of the chest. The ECG data were recorded with a sampling rate of 250 Hz using a Twente Medical Systems International amplifier/data-acquisition combination (Porti, Enschede, The Netherlands). The signal was analysed using the software package Carspan (Mulder 1992). R-waves of the ECG data were located and artefacts were corrected conservatively, using an automatic algorithm followed by a visual inspection to get valid data. Subsequently, the HR and HRV measures were computed.

The spectral procedure used is based on the direct Fourier transform of cardiac event sequences, which directly computes spectral values from the time points of R peaks in the ECG, without having to interpolate or resample interbeat interval values or HR values. In this way, spectral values are obtained for HR (instead of interval). This method, known as sparse discrete Fourier transform or Spectra of point events (DeBoer *et al.* 1984, Mulder 1992, van Steenis *et al.* 1994) can compute spectral measures from data series of any length. On the data segments a 5% taper window was applied. The obtained spectral values were re-sampled to a fixed resolution of 0.01 Hz to make spectral resolution independent from the length of the data series, using a linear integration algorithm. Before computing the spectral band values, a moving average with a width of three spectral points was applied in order to achieve smoothed spectra (Bartlett window).

HRV measures were estimated on the basis of HR spectra, using measures relative to mean HR (squared modulation index; Mulder 1992). The spectral profile method used is a kind of time-frequency analysis, in which the power of a specified frequency band is calculated as a function of time. In this moving window technique, the window length determines the time resolution while the amount of overlap between adjacent time segments is specified by the step size. Mean HR beats per min (bpm) was calculated with both a window and step size of 4 s (no overlap). For HRV the spectral power in the mid- and high-frequency bands was calculated with a window of 30 s and a step size of 1 s. This resulted in a 97% overlap with the adjacent windows. Since the power values of HRV follow a chi-square distribution (van Roon *et al.* 2004), a natural log transformation was performed to obtain a normal distribution. The mean HR and HRV values for the flight segments and flight elements were calculated by averaging the number of values that resulted from the selected step size. The resulting overlap with the adjacent segments and/or elements was accepted. Finally, the baseline values, which were obtained by averaging the measurements of a 5-min resting period directly before and after the experimental flight, were subtracted from the mean values.

An important element in the present approach is that weighted averages are computed from the short-segment band values, improving the spectral reliability. Although, at present, several types of time-frequency methods are available and widely accepted, there is not much insight in the comparability of spectral band measures from short segments (for instance 30 s) and time series of longer duration (for instance 300 s). For this reason, the results of a small pre-study are presented, in which such data are compared (see section 3).

2.5.2. Eye activity measures

The Jazz Synchronic system (version RS-232 Ober Consulting, Poland) was used to record eye activity with a sampling rate of 1 kHz. Both eye activity measures were derived from the recorded signal using the JazzManager application (version 2.0). This software package allows the user to manually adjust the detection parameters of a saccade detection algorithm. A saccade can be defined as an eye movement that is executed with the purpose of displacing the central field of view. The function of the detection algorithm is to distinguish genuine saccades from artefacts such as blinks and unintentional sensor movements. A short description of the parameters and the values that were used in this study are included in Table 3 (see the JazzManager user manual for a more detailed description).

Fixation duration was provided by the software directly and is defined as the time period between respective saccades. In this study, dwell time is defined as the time period during which the central field of view remains within the region of a single instrument. The calibrated amplitudes of the saccades in both the x-axis and the y-axis were combined to determine when the central field of view transitioned from one instrument to another. Due to the close proximity of the altimeter and the vertical velocity indicator (VVI), transitions between these two instruments could not be distinguished (see Figure 3).

2.6. *Procedure*

The experiment consisted of a familiarisation flight (no data recorded) and an experimental flight that lasted 24 and 28.5 min respectively. Before each flight the participants received a briefing. During this briefing, written instructions were handed out on how to execute the flight task in the simulator. After the experimental flight, participants estimated the invested effort during the six flight segments with the aid of the RSME (Zijlstra 1993). The entire experiment lasted about 4 h.

2.7. *Analysis*

First, the effects of gradual changes in task demands (and mental effort) were analysed. This was achieved by analysing the effects of the differences in task demands during the six flight segments on the selected physiological measures. Next, the effects of momentary changes in task demands were analysed. This was achieved by analysing the effects of the variations in task demand levels during the 37 flight elements on the selected physiological variables.

Table 3. Description of the parameters for the saccade detection algorithm with the values used in parentheses.

Parameter (value used)	Description
Speed points (3)	Number of eye positions that are used in velocity calculation
Local size (5)	Number of velocity samples that are used to determine local minima and maxima in the velocity signal
Minimum fixation (100 ms)	Minimum length of the fixation duration that must precede a detected saccade
Minimum duration (20 ms)	Minimum duration of a detected saccade
Maximum duration (200 ms)	Maximum duration of a detected saccade
Speed threshold (auto)	Minimum peak velocity of a detected saccade
Slope threshold (auto)	Minimum slope of the velocity profile during the onset of a saccade
Sharpness threshold (auto)	Minimum value of the ratio between the peak velocity and the duration of a detected saccade

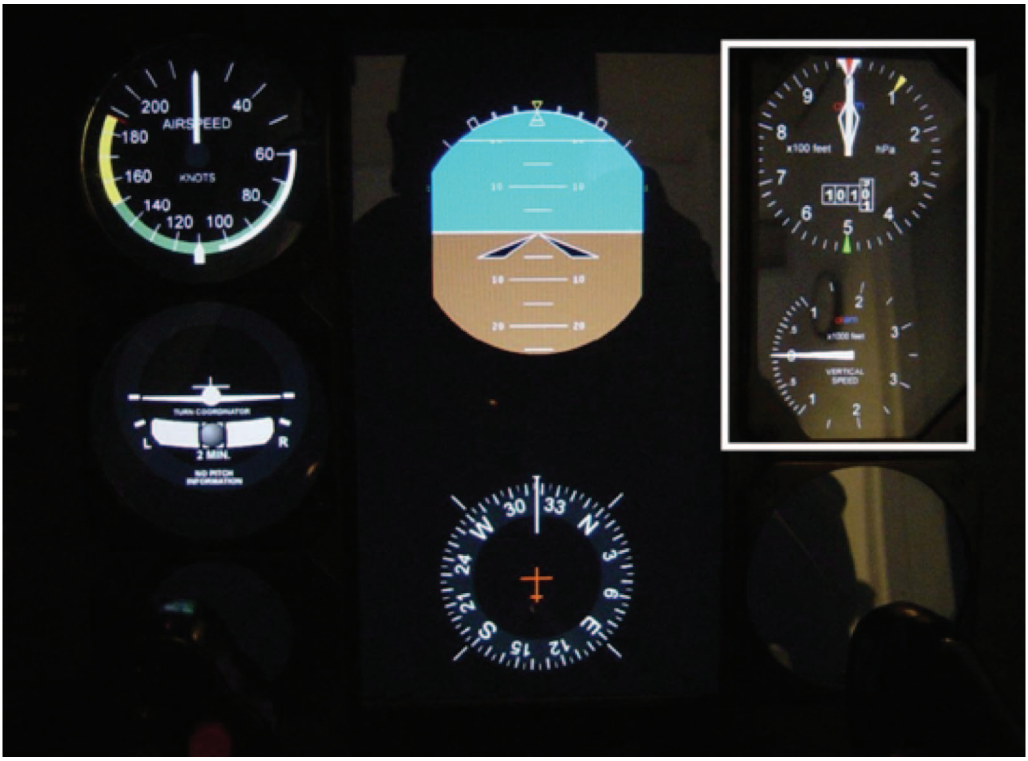


Figure 3. Relative locations of the flight instruments. The altimeter and the vertical velocity indicator are located within the white rectangle.

2.7.1. Statistical analysis

First, the means and standard errors of the RSME rating and the selected physiological variables were calculated. Then, random effect models were used to determine whether the means of these variables were equal in the different segments and task demand levels. This statistical analysis was performed with the software package MlwiN, version 2.00 (for a review, see Wright 2004).

A two level model was specified with individual measurements as level 1 and participants as level 2. Different models were evaluated by comparing the deviance statistic. The model building process was as follows. First, the empty model was built. Then, a saturated model was built with segments or task levels as fixed and random effects. This saturated model was compared to models with more parsimonious covariance structures (e.g. compound symmetry). Finally, the equality of different subsets of fixed effects was tested using models with this parsimonious covariance structure. *p*-values (one-sided) smaller than 5% were considered to be significant.

3. Comparing spectral band values of different segment lengths (pre-study)

Reliability of spectral data is directly dependent on the duration of analysis segments. Unsmoothed spectral power values have a chi-square distribution with two degrees of freedom. Reliability can be improved by taking spectral values together in spectral band values, by taking longer data segments and/or by (weighted) averaging of log transformed band values. Log transformed power values have a normal distribution (Bendat and Piersol 1986, van Roon *et al.* 2004). Several rules of thumb exist about the minimum required data length when spectral components are not averaged, such as 'minimally 100 data points' (Berntson *et al.* 1997).

In the present work, however, the authors did not work with separate spectral power values; spectral values are both averaged with regard to frequency (band values) and with regard to time (repeated segments). Nevertheless, the question remains whether small segment spectral analysis delivers stable spectral outcomes. For this reason a small pre-study was performed in which spectral power values from 240 s segment length are compared with those of averaged values (30, 60 and 120 s).

Data segments of 19 subjects of the original dataset were used. For each subject, nine data segments, consisting of a rest period (240 s) and eight consecutive task segments of 240 s (independent of flight phase), were selected. Each 240-s segment was divided into 8 × 30, 4 × 60 or 2 × 120 s. Spectral power values of each of these sub-segments were averaged (after log transformation) in order to get spectral power estimates of each sub-divided 240-s segment. Spectral band values of these sub-divided segments were compared to the band values of the original 240-s segments, following the spectral approach as described in section 2. Both the mid-frequency band and the high-frequency band will be used.

Results of these data are summarised in Table 4. Pearson correlations between the 240-s segments and the 30-, 60- and 120-s segments, over all 171 (19 × 9) values were between 0.972 and 0.994 (see Table 4, column 3). The differences between the spectral power values per frequency band and per segment length had an almost perfect Gaussian distribution, with mean values around zero (see Table 4, column 5). Largest deviations (–0.29) were for the short (30 s) segments. None of the mean differences deviated significantly from zero. Standard deviations for the distributions of the differences were again largest for the shortest segments (0.23, see Table 4, column 6). In order to get an idea of the magnitude of these differences, the range of the log power-values was between 5 and 10. Also the inter-subject correlations over the rest period and eight consecutive task segments were computed (nine values); mean Pearson correlations were between 0.88 and 0.95. As may be expected, because of the low number of individual data, these values are lower than the overall correlations (see Table 4, column 4). More detailed analysis and visual inspection of individual data showed that both restriction of range and extreme deviations in one or two data points strongly influenced

Table 4. Comparison of spectral band values: 240 s vs. 30, 60 and 120 s respectively.

Segment duration	Spectral band	Overall correlation	Mean correlation	Mean difference	SD of differences
30	mid	0.972	0.88	−0.29	0.23
60	mid	0.985	0.93	−0.06	0.16
120	mid	0.987	0.94	0.03	0.15
30	high	0.980	0.85	−0.13	0.22
60	high	0.983	0.89	0.08	0.21
120	high	0.994	0.95	0.05	0.11

Note: Differences are ln-values.

individual correlations. This detailed analysis also showed that it would be wise not to smooth the spectral data before computing the band values in future work.

Knowing the inherent restrictions in reliability of spectral power data, it can be concluded that the short-segment averaging approach, as used in the present paper, is well suited for obtaining spectral power estimates of short data segments with varying length.

4. Results

The results of increasing task demands in segments will be reported first, followed by the results of increasing task demand levels. The dependent variables are arranged in three groups: the RSME ratings; the cardiovascular measures; and the eye activity measures. Since the eye activity signal of two participants could not be calibrated reliably, the results of the mean dwell time are based on the data of 17 participants.

For each dependent variable the descriptive statistics will be reported first, followed by the results of the tests for the equality of fixed effects. The main effect of level on mean dwell time and fixation duration was tested using a model that was based on compound symmetry. All other fixed effects were tested with models that were based on a weakened version of compound symmetry.

4.1. Increasing task demands in segments

4.1.1. Rating Scale of Mental Effort

The mean RSME rating increases from 24 points (little effort) in segment 1 to 80 points (severe amount of effort) in segment 6 (see Figure 4).

Multilevel analyses demonstrated a significant main effect of segment on mean RSME rating ($p < 0.001$). The differences between segments 3 and 4 and between segments 5 and 6 were not significant.

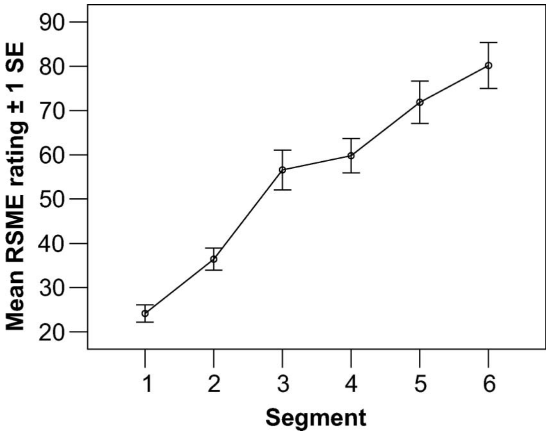


Figure 4. Mean Rating Scale of Mental Effort (RSME) rating (±1 SE) per flight segment.

4.1.2. Cardiovascular measures

In relation to the baseline, the mean increase in HR ranges from 9 bpm in segment 1 to 22 bpm in segment 6 (see Figure 5). Both HRV in the mid- and high-frequency bands decreased in relation to the baseline. The decrease in the mid-frequency band ranges from -0.11 in segment 1 to -0.41 in segment 6 (see Figure 6) and the decrease in the high-frequency band ranges from -0.12 in segment 1 to -0.71 in segment 6 (see Figure 7).

Multilevel analyses demonstrated a significant main effect of segment on the difference in mean HR ($p < 0.001$) and both HRV in the mid- ($p = 0.039$) and high-frequency bands ($p < 0.001$). For HR, the difference between segments 5 and 6 was not significant.

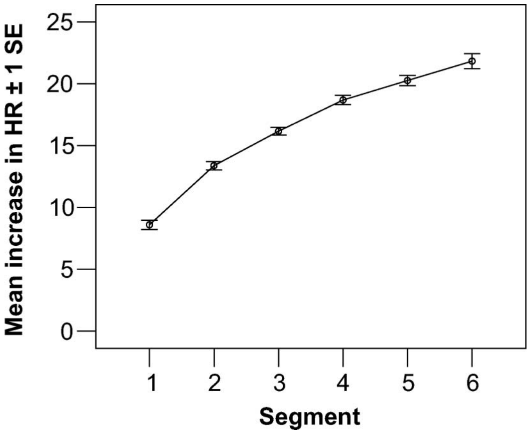


Figure 5. Mean increase in heart rate (HR) (±1 SE) per flight segment in relation to the baseline in beats per min.

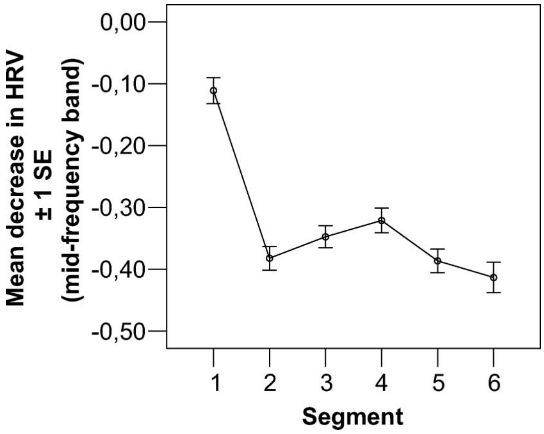


Figure 6. Mean decrease in heart rate variability (HRV) (±1 SE) in the mid-frequency band per flight segment in relation to the baseline (ln-values).

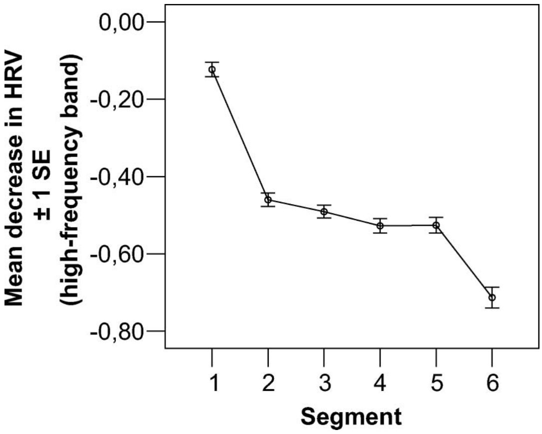


Figure 7. Mean decrease in heart rate variability (HRV) (±1 SE) in the high-frequency band per flight segment in relation to the baseline (ln-values).

For HRV in the mid-frequency band, the differences between levels 2, 3, 4, 5 and 6 were not significant. For HRV in the high-frequency band, the differences between segments 2, 3, 4 and 5 were not significant. In addition, the difference between the rest period and level 1 was not significant in both frequency bands.

4.1.3. Eye-based measures

Mean dwell time decreased from 520 ms in segment 1 to 456 ms in segment 6 (see Figure 8). Mean fixation duration showed a similar decrease from 493 ms in segment 1 to 418 ms in segment 5 (see Figure 9).

Multilevel analyses demonstrated a significant main effect of segment on mean dwell time and fixation duration ($p < 0.001$). For mean dwell time, the differences between segments 2, 4, 5 and 6 were not significant. For fixation duration, the differences between segments 2 and 3 and segments 4, 5 and 6 were not significant.

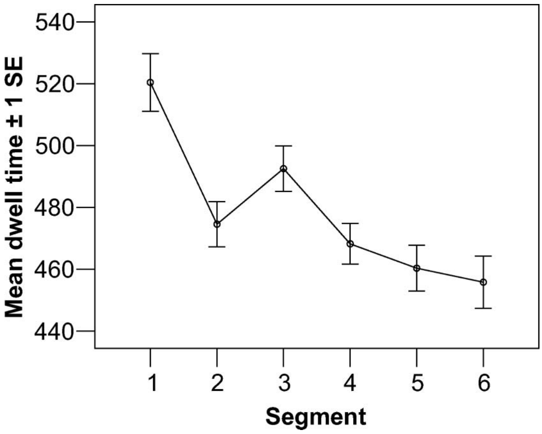


Figure 8. Mean dwell time (ms) per flight segment (±1 SE).

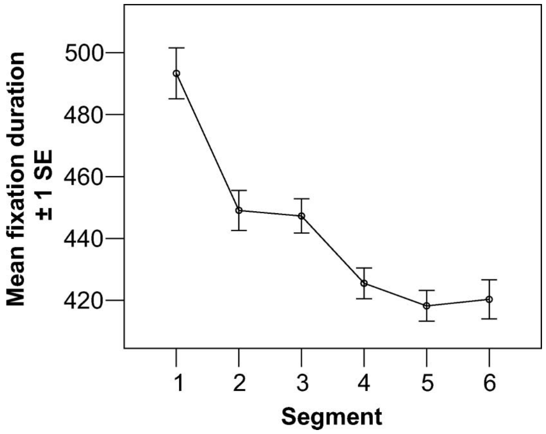


Figure 9. Mean fixation duration (ms) per flight segment (±1 SE).

4.2. Increasing task demand levels

4.2.1. Cardiovascular measures

In relation to the baseline, the mean increase in HR ranges from 12 bpm in level 1 to 18 bpm in level 3 (see Figure 10). Both HRV in the mid- and high-frequency bands decrease in relation to the baseline with increasing task demand levels. The decrease in the mid-frequency band ranges from -0.08 in level 1 to -0.36 in level 3 (see Figure 11) and the decrease in the high-frequency band ranges from -0.18 in level 1 to -0.54 in level 3 (see Figure 12).

Multilevel analyses demonstrated a significant main effect of level on the difference in mean HR and both HRV in the mid- and high-frequency bands ($p < 0.001$). For HR, the differences between levels 1 and 2 and levels 3 and 4 were not significant. For HRV in the

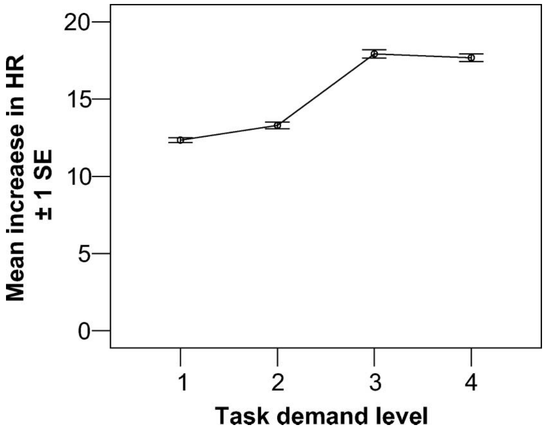


Figure 10. Mean increase in heart rate (HR) (±1 SE) per task demand level in relation to the baseline in beats per min.

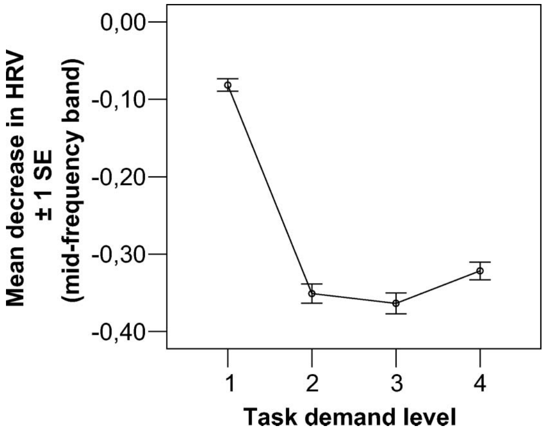


Figure 11. Mean decrease in heart rate variability (HRV) (±1 SE) in the mid-frequency band per task demand level in relation to the baseline (ln-values).

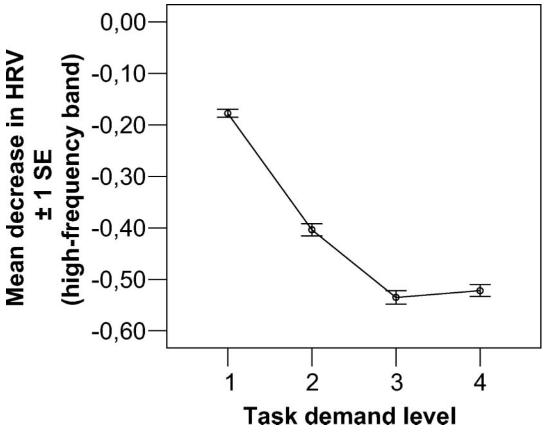


Figure 12. Mean decrease in heart rate variability (HRV) (±1 SE) in the high-frequency band per task demand level in relation to the baseline (ln-values).

mid-frequency band, the differences between levels 2, 3 and 4 were not significant. For HRV in the high-frequency band, the difference between levels 3 and 4 was not significant. In addition, the difference between the rest period and level 1 was not significant in both frequency bands.

4.2.2. *Eye-based measures*

Mean dwell time ranged from 498 ms in level 1 to 465 ms in level 3 (see Figure 13) and mean fixation duration ranged from 469 ms in level 1 to 419 ms in level 3 (see Figure 14).

Multilevel analyses demonstrated a significant main effect of level on mean dwell time and fixation duration ($p < 0.001$). For mean dwell time, the differences between levels 1, 2 and 4 were not significant. For fixation duration, the difference between levels 2 and 4 was not significant.

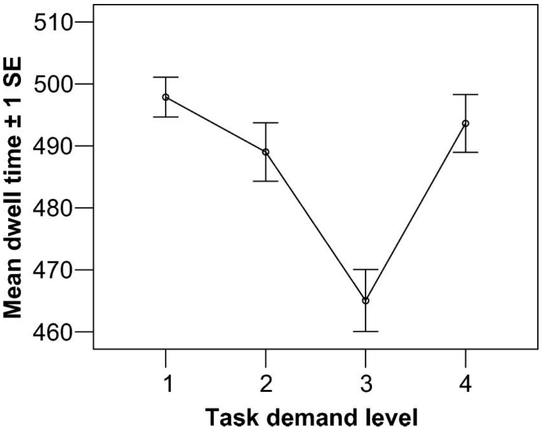


Figure 13. Mean dwell time (ms) per task demand level (±1 SE).

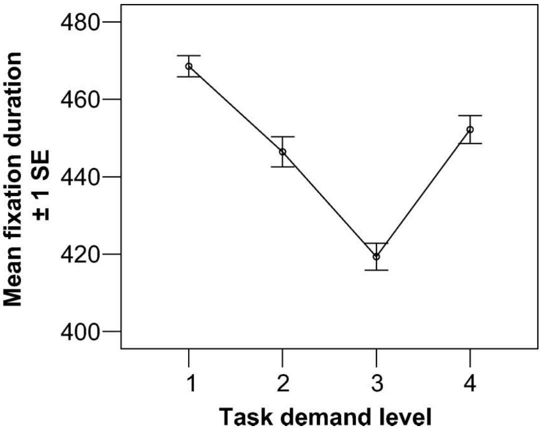


Figure 14. Mean fixation duration (ms) per task demand level (±1 SE).

5. Discussion

5.1. Effect of gradual changes in mental effort

The mean RSME rating increased from 24 points (little effort) in segment 1 to 80 points (severe amount of effort) in segment 6. Therefore, it seems that task demands were successfully increased in each consecutive flight segment by including more difficult flight elements.

In accordance with earlier research results, HR increased with increasing task demands. The maximum increase in relation to the baseline (rest period) was about 22 bpm. This is quite substantial; in most studies the increase in HR during the execution of a mental laboratory task (Turner *et al.* 1988) or an operating task (e.g. car driving; De Waard 1996, 2002, Richter *et al.* 1998) ranges from 5 to 13 bpm. Veltman and Gaillard (1998) also found large effects in HR in a flying task with different levels of task difficulty, but the effects were smaller than in the present study.

In line with other aviation research (Veltman and Gaillard 1996a, 1998, Wilson 2002), HRV was found to be less sensitive to gradual changes in mental effort than HR. The reduced sensitivity of the HRV measures might be explained by the relatively high workload conditions during the execution of the instrument flight task. While HRV might be suitable to distinguish gradual fluctuations in task demands at low to intermediate workload levels, HR might be more suitable to distinguish these fluctuations at high workload levels. Still, HRV in the high-frequency band distinguished three intermediate levels of mental workload: segment 1; segment 2 to segment 5; and segment 6. The maximum decrease in relation to the baseline (–0.71) is relatively large in comparison to the results of other studies, where maximum decreases generally ranged from -0.5 to -0.6 (De Waard 1996, van Roon 1998). The study of Veltman and Gaillard (1998), however, showed an even larger decrease from rest to (flying) task.

HRV in the mid-frequency band distinguished two intermediate levels of mental workload: segment 1; and segment 2 to segment 6. It seems that the effects of gradual changes in mental effort in the mid-frequency band are less stable, resulting in larger standard errors in comparison to the high-frequency band.

In general, dwell time and fixation duration decreased with increasing task demands. These results are consistent with other aviation research, where the experimental task consisted of a flight task without additional mental tasks (e.g. Svensson and Wilson 2002). In addition, these results are in line with the expectations based on the study of van Orden *et al.* (2000), who used a visual compensatory tracking task.

Both dwell time and fixation duration provided similar information. The only exception is the relatively short mean dwell time in segment 2, which is probably related to the method of determining dwell time in this study (see section 5.5). In general, the mean dwell time was approximately 35 ms longer than the mean fixation duration. The relatively small difference between both eye activity measures can be explained by the characteristics of an instrument flight task. During the execution of this task, participants scan their flight instruments to detect deviations from the prescribed parameters. Since the detection of these deviations will normally take only one or two fixations, the small time difference between both eye activity measures could have been anticipated.

The sensitivity to variations in task demands of fixation duration and HRV (high-frequency band) are more or less comparable. Based on the mean fixation duration, three intermediate levels of mental effort could be distinguished: segment 1; segments 2 and 3; and segments 4, 5 and 6.

5.2. *Effect of momentary changes in mental effort*

Since the RSME was only used to estimate invested effort during the six flight segments, this rating scale could not be used to see if subjective ratings also increased with increasing task demand level. Therefore, the discussion will concentrate on the associations between the different physiological measures.

In general, all cardiovascular measures indicated an increase in mental effort with increasing task demand level. In contrast with the results of gradual changes in mental effort during the flight segments, no single measure was found to be most sensitive to momentary variations in task demands. HR was the only cardiovascular measure that could differentiate between the invested effort during the rest period and level 1. HRV (both frequency bands), on the other hand, could differentiate between level 1 and level 2, whereas HR could not. Both HR and HRV in the high-frequency band could distinguish level 2 from level 3 and none of the cardiovascular measures indicated a difference in mental effort between level 3 and level 4.

In summary, all cardiovascular measures provided valuable information about the variations in task demands at the four different levels. Combining the information of both HR and HRV, it seems that task demands increased up to level 3 and then stabilised at level 4.

These results are in line with cardiovascular effects in the following two studies. Richter *et al.* (1998) related HR and HRV to road characteristics in a car-driving task on the basis of task analysis. They found that cardiovascular responses varied in relation to driving difficulty of road segments. Highest HR was found at intermediate curvature change rate, which was related to processes of information intake. HRV effects were strongest in intermediate and high curvature change conditions. Effects in HRV were in the same range as in the present experiment, while HR effects were lower.

Veltman and Gaillard (1998) studied HR and HRV reactions in a flight simulator; in their study, pilots had to fly through a tunnel with varying levels of difficulty. A large decrease of HRV was found when comparing task with rest conditions, while task conditions did not differentiate from each other. Part of the changes in HRV could be attributed to changes in respiratory pattern. HR was considerably increased during the high task load conditions compared to lower task load, while the magnitude of the effects was lower than in the present experiment. To be more precise, highest effects on HR were in the range of the present level 1 and 2 conditions.

In accordance with the results of the cardiovascular measures, the decrease of mean dwell time and fixation duration also indicates an increase in task demands from level 1 to level 3. Mean fixation duration distinguishes all three levels and seems to be the most sensitive measure to momentary changes in mental effort. At level 4, however, the results of the cardiovascular and eye activity measures diverge. Although the cardiovascular measures do not differentiate between level 3 and level 4, the increase of mean dwell time and fixation time between these two levels suggests a sharp decrease in task demands. These results might be explained by the sensitivity of eye activity measures to task characteristics.

5.3. Sensitivity to task characteristics

Physiological measures are not only affected by mental effort. Physical exercise, for example, has a similar effect on cardiovascular measures as mental effort; HR increases and HRV decreases (Mulder 1992). Eye activity measures are also influenced by other factors. An increase in pupil diameter, for instance, could be the result of an increase in mental effort or a decrease in illumination level. The eye activity measures, especially, seem to be highly task dependent (van Orden *et al.* 2000). As described in section 1, an increase in dwell time and fixation duration could either indicate an increase or a decrease in mental effort. The direction of the effect seems to be dependent on specific characteristics of the experimental task. Therefore, the diversion of results between the cardiovascular and eye measures might be explained by a closer examination of the specific task characteristics at each task demand level.

During the execution of the flight elements with task demand levels 1, 2 and 3, the aircraft is in a state of equilibrium. The participants essentially perform a tracking task; they continually scan their flight instruments for deviations from the prescribed parameters and make adjustments accordingly. This 'scanning behaviour' results in relatively short dwell times and fixation durations. During the flight elements in task demand level 4, however, the participants change the attitude of the airplane with the aid of the artificial horizon. This attitude change can take up to several seconds, which will result in relatively long dwell times and fixation durations. Therefore, the increase in dwell time and fixation duration at task demand level 4 might not be caused by an increase in mental effort but by a change in task characteristics.

5.4. Heart rate variability spectral profiles

When using the spectral profile method to detect momentary changes in mental effort, the choice of the window length is of crucial importance. The use of a wide window will smooth the HRV-pattern, which will complicate the detection of momentary changes. On the other hand, a narrow window will lack the statistical power to achieve significant results. To be able to distinguish changes in mental effort in the mid-frequency band, Mulder (1992) stated that segments of at least 30 to 40 s are required. The pre-study in this paper shows, indeed, that HRV measures with lengths of at least 30 s are stable enough to use in applications such as these. As a matter of fact, the crucial criterion is whether the total amount of data (total time included in the analysis of the averaged segments) is long enough to get reliable data. If this were a problem, the sensitivity of the applied HRV method would have been lower than usual for spectral HRV analysis of longer periods. This is certainly not the case in the present study (Richter *et al.* 1998, Veltman and Gaillard 1998, Althaus *et al.* 1999). Therefore, the spectral power in the mid- and high-frequency bands was calculated with a window length of 30 s. This window length was expected to provide an optimal trade-off between time resolution and statistical power.

Spectral analysis of the HRV-pattern appears to be a powerful method for detecting momentary changes in mental effort. The high-frequency band, especially, turned out to be sensitive to momentary changes in task demands. The sensitivity of this band might even be increased by narrowing the length of the window in future studies. According to the literature (Berntson *et al.* 1997, van Steenis 2002), the mid-frequency band requires longer time segments than the high-frequency band to achieve enough power for statistical significance. Therefore, a window length of 15 s for the high-frequency band might increase time resolution while maintaining enough statistical power.

The results in the present study indicate once again that the cardiovascular system can react very rapidly on changes in task load and related mental effort. This is well known for HR (Richter *et al.* 1998). The present approach shows, however, that HRV measures can also be used for this purpose if short segments can be applied. It may be expected that, in particular, the parasympathetic part of the autonomous nervous system is involved in these short-term changes because of its small time constants, compared to the sympathetic part (van Roon *et al.* 2004). Also, for this reason, it may be expected that the high-frequency band of HRV is more sensitive to short-term changes in mental effort than the mid-frequency band.

5.5. Limitations of the study

5.5.1. Level of experience

The participants in this study all had less than 100 flight hours and were relatively inexperienced. As a result, successful completion of the instrument flight task probably required them to exert a relatively high level of mental effort. For experienced pilots, however, instrument flying is a basic flight task that is highly automated. Therefore, it is likely that less variation in mental effort will be found during the performance of an instrument flight task by an experienced pilot than found in the present study with less experienced pilots. It should be realised, however, that it was not the aim here to estimate absolute workload levels of pilots during instrument flying. In contrast, the aim was more to study the effects of different work load levels on HRV and eye-derived measures, which makes the present group of relative inexperienced pilots even more suitable.

5.5.2. *Time-on-task effects*

The instrument flight task was designed to produce increasing task demands during flight. This approach should prevent an early mental overload. An early mental overload could lead to the incorrect execution of the prescribed flight elements, making it impossible to accurately assess the effect of varying task demands on the physiological measures. In addition, it was expected that participants would need a considerable amount of time to recover from a mental overload condition. As a result, a large amount of the experimental data might be lost. A disadvantage of the present approach is that it is hard to discriminate task load effects from time-on-task effects. In principle, mental effort in each flight segment might have increased as a result of fatigue instead of increased task demands.

During the experiment, mental overload turned out not to be a significant factor. Although small mistakes were made during the execution of the flight task, all participants were able to continue the profile as instructed. This might be the result of their successful selection for the pilot training programme. Future pilots are selected on their capability to set priorities in a highly demanding environment, which prevents them from being overloaded. In retrospect, it would have been preferable to vary task demands more randomly to be able to distinguish task load effects from time-on-task effects. On the other hand, the flight task lasted less than 0.5 h and was performed by highly motivated participants. In addition, HR generally decreases with time-on-task (Mascord and Heath 1992, Brookhuis and De Waard 1993) and the opposite effect was found here. Therefore, no substantial effect of fatigue is expected.

5.5.3. *Performance measures*

In addition to subjective ratings and physiological measures, performance measures can also provide a valuable insight in invested effort. In this study, no performance measures were used as a result of an inherent limitation of the ALSIM AL 100 Flight Trainer. This flight trainer registers performance as the deviations from the prescribed flight parameters (e.g. airspeed, altitude, turn rate). These deviations, however, cannot be registered continuously as the following example demonstrates. During straight and level flight, the deviations from the prescribed altitude, airspeed and heading are registered. During a climbing turn, however, the prescribed values for these parameters are undefined because altitude, airspeed and heading are continuously changing. As a result, it is not possible to compare the level of performance between these two flight elements. Still, performance measures are an important aspect of measuring mental effort and should be incorporated in future studies.

5.5.4. *Method for measuring dwell time*

In this study, the calibrated amplitudes of the saccades in both the x-axis and the y-axis were combined to determine whether the central field of view transitioned from one instrument to another. Due to the close proximity of the altimeter and the VVI, transitions between these two instruments could not be distinguished. As a result, when the central field of view remained within the boundaries of these two instruments, it was counted as a single dwell. This should have had no effect on the mean dwell time during the execution of the flight elements that only required one of these instruments to be monitored. However, during the execution of flight elements that required both instruments to be monitored, the mean dwell time was probably artificially raised.

The relatively short mean dwell time in segment 2 might be explained by this limitation in determining dwell time. Segments 3, 4, 5 and 6 all contain altitude changes. This manoeuvre requires the participants to monitor both the altimeter and the VVI to determine the correct rate of climb, which will result in an artificial lengthening of the mean dwell time. In contrast, segment 2 consists of a horizontal turn, which only requires the monitoring of the altimeter. Therefore, the mean dwell time is not artificially lengthened during this manoeuvre, which provides an explanation for the relatively short mean dwell time in comparison to the other segments.

5.6. Conclusion

The main aim of the present study was to determine to which degree HR and eye activity measures reflect momentary changes in mental effort during the execution of an instrument flight task. Although HR was found to be the most sensitive measure for detecting gradual changes in mental effort, the selected measures were complementary in the detection of momentary changes. HR was the only cardiovascular measure that could differentiate between the rest period and task execution. The combined information of the HRV in both frequency bands, however, provided more insight in the intermediate levels of mental effort.

The eye activity measures were sensitive to momentary changes in mental effort as well, but the results were also influenced by specific task characteristics. Therefore, a task analysis is required before these measures can be used as valid indices of mental effort. In addition, dwell time and fixation duration are unsuitable to distinguish the rest period from task execution since a baseline for eye movements is hard to establish.

Acknowledgements

The authors wish to thank the staff members of the Center for Man and Aviation (Soesterberg, The Netherlands) for providing them with the logistical support for the completion of this study. They also wish to thank Evelien Kums, Michiel de Leeuw and Siemen Henk Sikkema for their fruitful cooperation in the study design and data collection and Hedwig Hofstetter for the pre-study on short-segment HRV analysis. Finally, they wish to thank the selection candidates and Air Force cadets who participated in this study.

References

Aasman, J., Mulder, G., and Mulder, L.J.M., 1987. Operator effort and the measurement of heart-rate variability. *Human Factors*, 29, 161–170.

Althaus, M., *et al*., 1999. Cardiac activity to attention-demanding tasks in children with a pervasive development disorder not otherwise specified (PDDNOS). *Biological Psychiatry*, 46, 799–809.

Bendat, J.S. and Piersol, A.G., 1986. *Random data: Analysis and measurement procedures*. New York: John Wiley & Sons, Inc.

Berntson, G.G., *et al*., 1997. Committee Report – Heart rate variability: Origins, methods and interpretive caveats. *Psychophysiology*, 34, 623–648.

Boucsein, W. and Backs, R.W., 2000. Engineering psychophysiology as a discipline: Historical and theoretical aspects. *In*: R.W. Backs and W. Boucsein, eds. *Engineering psychophysiology: Issues and applications*. London: Lawrence Erlbaum Associates, 3–30.

Brookhuis, K.A. and De Waard, D., 1993. The use of psychophysiology to assess driver status. *Ergonomics*, 36, 1099–1110.

Callan, D.J., 1998. Eye movement relationships to excessive performance error in aviation. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 42, 1132–1136.

Cnossen, F., Meijman, T., and Rothengatter, T., 2004. Adaptive strategy changes as a function of task demands: a study of car drivers. *Ergonomics*, 47, 218–236.

DeBoer, R.W., Karemaker, J.M., and Strackee, J., 1984. Comparing spectra of a series of point events particularly for heart rate variability data. *IEEE Transactions on Biomedical Engineering, BME-31*, 4, 384–387.

De Waard, D., 1996. *The measurement of drivers' mental workload*. Thesis (PhD). University of Groningen.

De Waard, D., 2002. Mental workload. *In*: R. Fuller and J.A. Santos, eds. *Human factors for highway engineers*. Oxford: Pergamon, 161–175.

Duchowski, A.T., 2002. A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34, 455–471.

Endsley, M.R. and Kaber, D.B., 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42 (3), 462–492.

Gopher, D. and Donchin, E., 1986. Workload: An examination of the concept. *In*: K.R. Boff, L. Kaufman, and J.P. Thomas, eds. *Handbook of perception and human performance*, volume II. New York: John Wiley and Sons, 1–49.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive energetical framework. *Biological Psychology*, 45 (1–3), 73–93.

Hoogeboom, P. and Hilburn, B., 2001. *Eye-derived indices of operator state*. NLR/011001–5. Amsterdam: National Aerospace Laboratory NLR.

Johnson, A. and Proctor, R.W., 2004. *Attention: Theory and practice*. Thousand Oaks: Sage Publications, Inc.

Jorna, P.G.A.M., 1992. Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. *Biological Psychology*, 34, 237–258.

Jorna, P.G.A.M., 1993. Heart rate and workload variations in actual and simulated flight. *Ergonomics*, 36, 1043–1054.

Kitney, R.I., 1980. An analysis of the thermoregulatory influences on heart rate variability. *In*: R.I. Kitney and O. Rompelman, eds. *The beat-by-beat investigation of cardiovascular function*. Oxford: Clarendon Press, 81–106.

Kramer, A.F., 1991. Physiological metrics of mental workload: a review of recent progress. *In*: D.L. Damos, ed. *Multiple-task performance*. London: Taylor & Francis, 279–328.

Mascord, D.J. and Heath, R.A., 1992. Behavioral and physiological indices of fatigue in a visual tracking task. *Journal of Safety Research*, 23, 19–25.

May, J.G., *et al.*, 1990. Eye movement indices of mental workload. *Acta Psychologica*, 75, 75–89.

Mulder, G., 1985. Attention, effort and sinus arrhythmia: how far are we? *In*: J.F. Orlebeke, G. Mulder, and L.P.J. van Doornen, eds. *The psychophysiology of cardiovascular control*. New York: Plenum Press, 407–424.

Mulder, G., 1986. The concept and measurement of mental effort. *In*: G.R.J. Hockey, A.W.K. Gaillard, and M.G.H. Coles, eds. *Energetics and human information processing*. Dordrecht: Kluwer, 175–198.

Mulder, L.J.M., 1992. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34, 205–236.

Mulder, L.J.M. and Mulder, G., 1987. Cardiovascular reactivity and mental workload. *In*: R.I. Kitney and O. Rompelman, eds. *The beat-by-beat investigation of cardiovascular function*. Oxford: Clarendon Press, 216–253.

Richter, P., *et al.*, 1998. Psychophysiological analysis of mental load during driving on rural roads – a quasi-experimental field study. *Ergonomics*, 41, 593–609.

Roscoe, A.H., 1992. Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, 34, 259–287.

Svensson, E.A.I. and Wilson, G.F., 2002. Psychological and psycho physiological models of pilot performance for systems development and mission evaluation. *International Journal of Aviation Psychology*, 12, 95–110.

Tole, J.R., *et al.*, 1982. Visual scanning behavior and mental workload in aircraft pilots. *Aviation, Space, and Environmental Medicine*, 53, 54–61.

Turner, J.R., *et al.*, 1988. A comparison of additional heart rates during active psychological challenge calculated from upper body and lower body dynamic exercise. *Psychophysiology*, 25, 209–216.

van Orden, K.F., Jung, T., and Makeig, S., 2000. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, 52, 221–240.

van Orden, K.F., Limbert, W., and Makeig, S., 2001. Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, 43, 111–121.

van Roon, A.M., 1998. *Short-term cardiovascular effects of mental tasks, physiology, experiments and computer simulations*. Thesis (PhD). University of Groningen.

van Roon, A.M., *et al.*, 2004. Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology*, 41, 961–981.

van Steenis, H.G., 2002. *On time frequency analysis of heart rate variability*. Thesis (PhD). Erasmus University.

van Steenis, H.G., Tulen, J.F.M., and Mulder, L.J.M., 1994. Heart rate variability spectra based on non-equidistant sampling: the spectrum of counts and the instantaneous heart rate spectrum. *Medical Engineering and Physics*, 16, 35–36.

Veltman, J.A., 2002. A comparative study of psycho physiological reactions during simulator and real flight. *The International Journal of Aviation Psychology*, 12, 33–48.

Veltman, J.A. and Gaillard, A.W.K., 1993. Measurement of pilot workload with subjective and physiological techniques. *Workload assessment and aviation safety; Proceedings of the conference*, 27–28 April 1993. London, UK. 3.1–3.13.

Veltman, J.A. and Gaillard, A.W.K., 1996a. Physiological indices of workload in a simulated flight task. *Biological Psychology*, 42, 323–342.

Veltman, J.A. and Gaillard, A.W.K., 1996b. Pilot workload evaluated with subjective and physiological measures. *In*: K. Brookhuis, C. Weikert, J. Moraal, and D. De Waard, eds. *Aging and human factors*. Groningen: University of Groningen, 107–128.

Veltman, J.A. and Gaillard, A.W.K., 1998. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41, 656–669.

Vicente, K.J., Thornton, D.C., and Moray, N., 1987. Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors*, 29, 171–182.

Wilson, G.F., 1992. Applied use of cardiac and respiration measures: practical considerations and precautions. *Biological Psychology*, 34, 163–178.

Wilson, G.F., 2001. In-flight psycho physiological monitoring. *In*: F. Fahrenberg and M. Myrtek, eds. *Progress in ambulatory monitoring*. Seattle, WA: Hogrefe & Huber, 435–454.

Wilson, G.F., 2002. An analysis of mental workload in pilots during flight using multiple psycho physiological measures. *The International Journal of Aviation Psychology*, 12, 3–18.

Wilson, G.F. and Eggemeier, F.T., 1991. Psycho physiological assessment of workload in multi task environments. *In*: D.L. Damons, ed. *Multi-task performance*. London: Taylor & Francis, 329–360.

Wright, D.B., 2004. Statistical software review. *British Journal of Mathematical & Statistical Psychology*, 57, 383–384.

Zijlstra, F.R.H., 1993. *Efficiency in work behaviour: a design approach for modern tools*. Thesis (PhD). University of Delft.