



Marinescu, Adrian and Sharples, Sarah and Campbell Ritchie, Alastair and Sanchez Lopez, Tomas and McDowell, Michael and Morvan, Herve (2017)
Physiological parameter response to variation of mental workload. Human Factors . ISSN 1547-8181

Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/45218/7/Physiological%20Parameter%20Response%20to%20Variation%20of%20Mental%20Workload_latest_version.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see: http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Physiological Parameter Response to Variation of Mental Workload

A. Marinescu*. S. Sharples**.

A. C. Ritchie**. T. Sánchez López***. M. McDowell***.

H. P. Morvan*.

*Institute of Aerospace Technology, University of Nottingham, Jubilee Campus, Nottingham, UK

**Faculty of Engineering, University of Nottingham, University Park, Nottingham, UK

***Airbus Group Innovations UK, Data Analysis and Interaction, Newport, UK

Structured Abstract

Objective: To examine the relationship between experienced mental workload and physiological response by non-invasive monitoring of physiological parameters including facial skin temperature, breathing rate, pupil diameter and heart rate variability.

Background: Previous studies have examined how individual physiological measures respond to changes in mental demand and subjective reports of mental workload. This study explores the response of multiple physiological parameters, measured simultaneously and quantifies the added value of each of the measures when estimating the level of demand.

Methods: The study presented was conducted in laboratory conditions and required participants to perform a custom-designed visual-motor task that imposed varying levels of demand. The data collected consisted of: physiological measurements (heart inter-beat intervals, breathing rate, pupil diameter, facial thermography); subjective ratings of workload from the participants (ISA and NASA-TLX); and the performance measured within the task.

Results: Facial thermography and pupil diameter were demonstrated to be good candidates for non-invasive mental workload measurements; for 7 out of 10 participants, pupil diameter showed a strong correlation (with R values between 0.61 and 0.79 at a significance value of 0.01) with mean ISA normalized values. Facial thermography measures added on average 47.7% to the amount of variability in task performance explained by a regression model. As with the ISA ratings, the relationship between the physiological measures and performance showed strong inter-participant differences, with some individuals demonstrating a much stronger relationship between workload and performance measures than others.

27 **Conclusion:** The results presented in this paper demonstrate that physiological monitoring can be used for non-invasive real-
28 time measurement of workload, assuming models have been appropriately trained on previously recorded data from the user
29 population. Facial thermography combined with measurement of pupil diameter are strong candidates for real-time monitoring
30 of workload due to the availability and non-intrusive nature of current technology. The study also demonstrates the importance
31 of identifying whether an individual is one who demonstrates a strong relationship between physiological measures and
32 experienced workload measures before physiological measures are applied uniformly. This is a feasible proposition in a setting
33 such as aircraft cockpits, where pilots are drawn from a relatively small, targeted and managed population.

34 **Application:** The methods presented in this article, with current technological capabilities, are better suited for workplaces
35 where the person is seated, offering the possibility of being applied to fixed wing and rotorcraft pilots and air traffic controllers.

36 **Précis:** Objective, real time non-invasive estimation of mental workload level based on physiological measures is now a
37 realistic proposition that will have an impact in the way human-machine systems are designed and evaluated. The results
38 presented in this paper demonstrate that physiological monitoring, especially pupil diameter and facial thermography
39 combined with a facial landmark tracking algorithm, can be used for non-invasive real-time measurement of workload.

40 **Keywords:** mental workload, human performance, facial-thermography, pupil diameter, physiological measures

41

42 1. INTRODUCTION

43 Since the 1980s, passenger air traffic has doubled every 15 years and it is expected to double again by 2034, with 70% of the
44 traffic relying on the extant network (Airbus, 2015). Near future air transport challenges such as increased air traffic, the need
45 for more efficient routes or the introduction of free flight raise new issues of relevance to human factors. The pilot of the future
46 will have to operate in a more congested airspace, aided by more complex technology. One human factors notion that has
47 potential to support the management of increased demand against available cognitive capabilities is workload. This study
48 explores techniques for measuring the mental workload experienced by participants by using non-invasive and minimally
49 intrusive physiological measurements. These measurements have tremendous potential to aid the real time understanding of
50 human workload, but presents a number of challenges in the design and implementation of methodologies (see e.g.
51 (Parasuraman & Mehta, 2015), (Sharples & Megaw, 2015)).

52 Mental workload has been suggested to have a strong relationship with human performance, the current consensus being that
53 both excessively high and excessively low levels of mental workload influence performance negatively (Young, Brookhuis,
54 Wickens, & Hancock, 2014) (Sharples & Megaw, 2015). Traditionally, some methods of workload assessment have been
55 difficult to implement in-situ in a real work environment due to being invasive (e.g. interrupting tasks or requiring

56 uncomfortable equipment to be worn). Advances in physiological sensors and data analytic techniques mean that tools such as
57 facial thermography (Ora & Duffy, 2007) are now realistic candidates for non-invasive capture of workload in real time. Lehrer
58 et al., 2010 concluded in a flight simulator study that the minimum R-R intervals (time interval between heart beats extracted
59 from ECG data) in a task significantly discriminated between high and combined moderate and low-load tasks. Eye movement
60 activity was used by Ahlstrom and Friedman (2006) in an air traffic control study and they concluded that blink duration, blink
61 frequency, mean saccade distance and pupil diameter can provide a sensitive measure of mental workload. They have
62 established that an increase in experienced mental workload level (subjectively measured on a scaled from 1 to 10 using the
63 ATWIT method) is correlated with an increase in pupil diameter (Ahlstrom & Friedman-Berg, 2006). For such candidate
64 measures to be deployed, new knowledge is required to establish the validity, reliability and sensitivity of such tools.

65 Previous studies have explored whether it is possible to infer mental workload by using facial thermography. These studies
66 have shown a high correlation of workload with the decreasing temperature of the skin covering the tip of the nose. Ora and
67 Duffy (2007) first used a simulator driving task together with a mental arithmetic loading task to increase the mental workload
68 while measuring nose and forehead temperature followed by performing a study in a real car driving situation. They
69 demonstrated that there is a strong correlation between the change in nose surface temperature and the subjective ratings for
70 mental workload while the forehead temperature remained relatively constant (Ora & Duffy, 2007). Another study in a ship
71 simulator showed that nasal temperature and heart rate variability are good indices for effective navigation, and also connected
72 the measures to the variation of mental workload (Murai, Hayashi, Okazaki, & Stone, 2008). The reason for the nose
73 temperature drop identified by Ora & Duffy (2007) is the vasoconstriction response of the autonomic nervous system to mental
74 stress or negative emotion, mediated primarily by the sympathetic nervous system. Thermal imaging of the forehead, nose,
75 eyes, cheeks and chin during a cognitive stress test was able to classify mental workload into three levels with 81% accuracy
76 (Stemberger, Allison, & Schnell, 2010). However these studies did not establish the 'added value' of facial thermography as a
77 physiological tool over other techniques such as heart rate or pupil diameter/eye movements (both of which can require the use
78 of more intrusive and personally worn monitoring equipment).

79 The study presented in this paper explores the changes in the physiological parameters that occur as the level of mental
80 workload varies and examines whether a combination of these parameters could be used for estimating the level of mental
81 workload. The study uses a task that has varying level of demand with the aim of eliciting different levels of experienced
82 workload which are then captured by subjective and physiological measures.

83 The hypotheses of this study are that:

- 84 1. There will be a measurable difference in subjective workload between the two levels of task difficulty
- 85 2. The subjective ratings of workload will be associated with changes in physiological measures

86 3. Multiple physiological measures can be used in combination to analyze workload.

87

88 2. METHOD

89 This research complied with the American Psychological Association Code of Ethics and was approved by the Faculty of
90 Engineering Research Ethics Committee at University of Nottingham. Informed consent was obtained from each participant.
91 Participants were presented with an information sheet and consent form, it was verified that they were over 18 years old, and
92 had no pre-existing heart-related condition and had no skin conditions or allergies that could prevent them from wearing the
93 heart rate chest strap.

94

95 2.1. Participants

96 Fourteen students and staff from the University of Nottingham took part in the study (11 men and 3 women; M age = 28.3
97 years; SD = 4.9; range = 21-38). The participants were recruited via e-mail and were compensated with a £20 Amazon voucher
98 for their time. The data from four participants was discarded due to data recording problems and difficulties in tracking the
99 facial features. Data from the remaining ten participants is presented here.

100 2.2. Apparatus

101 The Zephyr BioHarness 3 chest strap was used for measuring posture, heart and breathing activity. The device outputs raw
102 ECG data at a sampling rate of 1000 Hz and also a processed version of the raw signal including the R-R intervals, heart rate
103 and breathing rate. (Medtronic, Annapolis USA).

104 For eye-tracking, the RED 250 eye tracker was used in stand-alone configuration, measuring pupil diameter and gaze data at
105 60 Hz (SensoMotoric Instruments, Teltow-Germany).

106 The FLIR SC7000 thermal IR camera with a spectral range of 3-5 μm was employed for monitoring the facial thermal features
107 of the participants. The resolution of the camera is 640x512 and was used at a sampling frequency of 50 Hz. The camera offers
108 a noise equivalent differential temperature of less than 25 mK (FLIR Systems, Wilsonville, Oregon-USA).

109 The near-infrared light used by the eye-tracker for illumination has a wavelength of 870 nm (Sensomotoric Instruments, 2011,
110 pg. 186), which is outside the 3-5 μm (FLIR, 2012, pg.2) spectral range of the thermal camera, therefore it does not influence
111 the measurements of the thermal camera.

112 To perform the task, the participants were seated at about 1.5 m away from a 55 inch (1397 mm) LCD flat screen display; the
113 task covered a rectangular area of 652mm x 718mm (H x W) while the rest of the area was black. Although no data was
114 recorded with regards to the light intensity in the room, this was kept as constant as possible by keeping the light off and having
115 the blinds closed. The background of the screen during the task was black and it is expected that as the number of balls onscreen

116 increased, the light intensity coming from the screen would increase as well, inducing a pupil contraction response (Winn,
 117 Whitaker, Elliott, & Phillips, 1994). In fact, results demonstrated the opposite of that was observed, meaning that the most
 118 likely dominant factor inducing the dilation of the pupil was the task difficulty. Had the light intensity from the screen been
 119 constant, the observed effect may have been even larger.

120 **2.3. Materials**

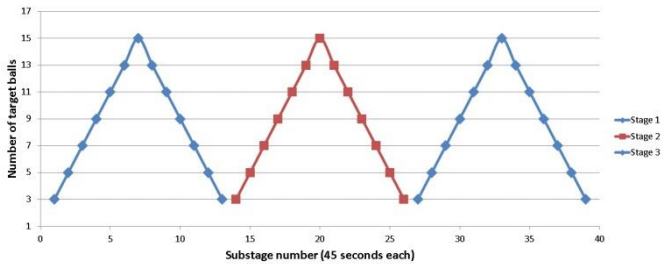
121 In order to explore the relationship between mental workload, variation of performance and objective physiological parameters,
 122 a specific computer-based task was designed to impose different levels of mental demand on the participant.

123 The task consisted of a computer game with 3 stages of two levels of difficulty, in total lasting 29 minutes; each stage consisted
 124 of 13 sub-stages (45s each) of varying difficulty, a task paradigm previously used in our research group (Sharples, Edwards,
 125 & Balfe, 2012). Table 1 describes the task stages in terms of targets, difficulty level and number of sub-stages.

	Stage 1	Stage 2	Stage 3
Targets	Red balls	Odd numbered balls	Red balls
Difficulty	Level 1 – low difficulty	Level 2 – high difficulty	Level 1 – low difficulty
No. of sub-stages (45s each)	13	13	13

126 Table 1 Task stages description

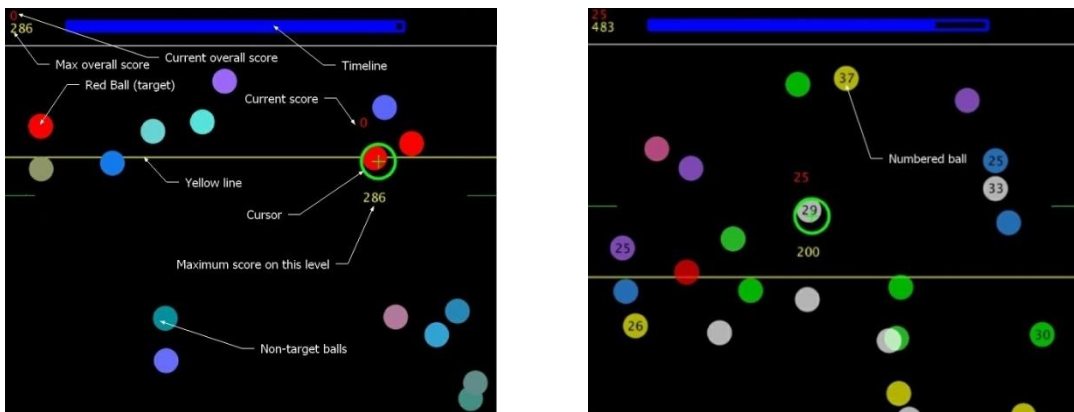
127 During each of the stages, the participant is presented with moving coloured balls on a black background. The movement of
 128 the balls gives the impression that they are falling from the top of the screen. At the beginning of each of the three stages, the
 129 participant is told which the target balls are; the task is to aim at the target balls using a joystick and shoot using a button on
 130 the joystick before the balls reach the yellow line and drag it down. During stages 1 and 3 of level 1 difficulty, the target balls
 131 are red (Fig. 2 Left) while during stage 2 of level 2 difficulty (Fig. 2 Right) the color of the balls no longer represents an
 132 identifier of the balls to be targeted. Instead the ones having odd numbers written on them now represent the target, introducing
 133 an additional cognitive element with the intent of increasing mental demand. Each of the stages is made up of 13 sub-stages,
 134 each presenting the participant with a set number of target balls on the screen at any time; when a target ball is shot, the game
 135 generates another one. The number of balls per sub-stage was varied as presented in (Fig. 1) in order to control the level of
 136 demand.



137
 138 Fig. 1. Description of stages

139 The position of the joystick is indicated by a red circular cursor that turns green once it is within range of the target balls and
140 the participant can make a successful shot (Fig. 2) using the front button on the joystick. At the beginning of the stage, the
141 horizontal yellow line finds itself at the top of the screen; when a target ball reaches the yellow line it will drag it down. The
142 participants are told that they have to fight the balls from dragging the yellow line down by shooting at them. Whenever a
143 target ball has been shot, the yellow line goes up by a small increment and whenever the participant misses a shot, the yellow
144 line goes down by the same increment. The main reasons for using the horizontal yellow line were:

- 145 • To prevent participants from focusing on the balls that are high on the screen and abandoning the ones that are lower and
146 will soon disappear off the screen, in this way subjecting all participants to the same number of targets at one time;
- 147 • To give participants a simple goal to fight towards – keeping the yellow line high up on the screen;
- 148 • To obtain a continuous measure of performance in terms of how high on the screen they were able to maintain the yellow
149 line at any moment.



157 *Fig. 2. Left: Level 1 difficulty stage (Stage 1) - Right: Level 2 difficulty stage (Stage 2)*

158 After each sub-stage, lasting 45s, the participant was prompted by a voice in the task for their subjective assessment of mental
159 workload, saying: ‘Level please’. The task was not frozen while asking for the ISA level, the participant just has to say a
160 number from 1 to 5. At the end of each stage, the task was paused and the participants were shown the task score they achieved
161 in comparison with the other participants as a means of increasing motivation.

162 Sample screen recordings of the task can be found at the following links:

- 163 • Stage 1 sample: <https://www.youtube.com/watch?v=7a4MaTZ5PzE>
- 164 • Stage 2 sample: <https://www.youtube.com/watch?v=FNwAnWgM024>

165 2.4. Design

166 The independent variable that was manipulated during the study was the task difficulty (i.e. imposed demand). The dependent
167 variables were the physiological measures, the subjective assessment of the perceived level of mental workload and the task
168 performance.

169 The Instantaneous self-assessment workload scale (ISA) (Brennen, 1997) was used once every 45s to collect subjective data
170 about the level of perceived mental workload. The ISA scale was developed primarily as a subjective measure of mental
171 workload for air traffic controllers and it involves the participants self-rating their workload on a scale from 1 (low) to 5 (high).
172 The main reason for using the ISA scale throughout the task was the low level of intrusion, as the participant would verbally
173 rate the perceived level of mental workload when prompted by an auditory message ('Level please').
174 At the end of each of the three task stages, the participant filled in a NASA-TLX (Hart, California, & Staveland, 1988)
175 questionnaire for a subjective assessment of workload. The reason for using NASA-TLX was to get a more detailed
176 retrospective multidimensional subjective assessment of each of the three stages to determine whether the manipulation of
177 imposed demand through task difficulty had resulted in a perceived experience of increased workload.

178 **2.5. Procedure**

179 Each participant was invited to read the information sheet, describing the details of the study, and then fill in a consent form.
180 They were then asked to play a training version of the stimulus task until they became familiar with the rules and the controls.
181 After the training was finished, the participants were invited to attach the Zephyr sensor around their chest in a private space;
182 the thermal and visual cameras were then aligned to match the height of the participant. Before starting the actual task, the eye
183 tracker was calibrated. When the participant was ready, they played stage 1 of the stimulus task, which lasted for almost 10
184 minutes, at the end of which the participant's score was shown in comparison to the participants before. During the game-play,
185 the participant rated the level of mental workload on the ISA scale once every 45s. After the first stage was over, they filled in
186 the NASA-TLX questionnaire. Stage 2 (higher demand level 2) and 3 (original demand level 1) of the task then followed.
187 Before starting each of the stages, the eye-tracker was recalibrated and after finishing each of the stages, the participant was
188 shown their score and filled in a NASA-TLX questionnaire. After stage 3 ended and the questionnaire had been completed, the
189 participant was invited to remove the Zephyr sensor in a private space. They were then offered a £20 voucher as a reward for
190 their time.

192 **3. RESULTS**

193 The results are presented in several stages. Firstly the results of the inferential tests to examine the impact of the manipulation
194 of the task demand on the measures of workload and performance are presented. The aim of these tests is to confirm that the
195 demand manipulation affected workload and performance in the manner anticipated. The second analysis examines the
196 relationship between the different measures of workload, using bivariate correlations and reporting both correlation
197 significance and the coefficient of determination to indicate effect size. The final analysis uses multiple linear regression to

198 determine the percentage of variability in task performance explained by the physiological measures and the relative
199 contribution of each of the measures.

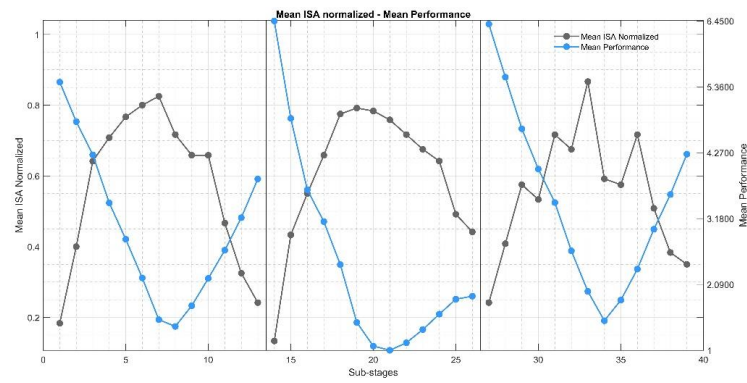
200 3.1. Subjective and Performance Data

201 A one way ANOVA ($F(1,28) = 4.56, p = 0.041, \eta^2 = 0.14$) confirmed that there was a difference between the two levels of
202 difficulty in terms of the NASA-TLX mental demand scale, confirming that stage 2 (odd numbered balls as targets) was
203 perceived to be more mentally demanding than stages 1 and 3 (red balls as targets), however the effect size is small, group
204 differences explaining about 14% of the variance.

205 One of the disadvantages of using the ISA technique is subjectivity in interpretation of the absolute meaning of numbers on
206 the rating scale, and thus the limited absolute validity that can be inferred from the ratings. However, it can be assumed that
207 the relative validity of the ratings is robust, and therefore in order to compare the results across the participants, the data were
208 normalized to a common scale ranging between 0 and 1.

209 Fig. 3 shows the mean performance score for all participants (better performance in the task results in a higher score) at sub-
210 stage scale, plotted against the mean normalized ISA rating for all participants. There is a negative correlation between the two
211 mean scores, the Pearson correlation coefficient is $R(37) = -0.74$ with $p < 0.01$, showing that as the mean subjective level of
212 mental workload increased, the mean task performance decreased.

213 While Fig. 3 looks at the mean performance and level of mental workload, Table 2 shows the individual correlations with
214 performance of both the mean ISA normalized and to each participant's rating. It can be observed that for the individual (non-
215 normalized) ISA ratings, three of the participants did not have significant correlations to the 0.05^1 level and that the R^2 value
216 is smaller in general compared with the mean ISA normalized correlation. Overall this data demonstrates a clear association
217 between performance and subjective workload.



219 Fig. 3. Mean ISA ratings – mean score

¹ Note that no familywise corrections such as Bonferonni were applied, as tests were conducted on independent (participant-based) data sets, but it should be acknowledged that as normal when multiple tests are conducted one in twenty will be significant by chance if a $p < 0.05$ level of significance is adopted

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.632	0.401	<0.01	-0.300	0.090	0.0632
2	-0.652	0.426	<0.01	-0.574	0.330	<0.01
3	-0.648	0.420	<0.01	-0.620	0.385	<0.01
4	-0.706	0.499	<0.01	-0.421	0.178	<0.01
5	-0.729	0.532	<0.01	-0.551	0.304	<0.01
6	-0.659	0.434	<0.01	-0.434	0.188	<0.01
7	-0.759	0.576	<0.01	-0.229	0.053	0.15
8	-0.783	0.613	<0.01	-0.754	0.569	<0.01
9	-0.681	0.465	<0.01	-0.038	0.001	0.81
10	-0.742	0.551	<0.01	-0.769	0.592	<0.01

Table 2 – Mean and normalized ISA ratings correlated with Performance

3.2. Physiological Data

The physiological data collected consisted of heart R-R inter-beat intervals, breathing rate, pupil diameter and facial skin temperatures measured by thermography. All physiological data reported is the mean of the readings taken during the 45s duration of each of the sub-stages.

Due to the fact that physiological data depends so much on the physiology of each of the participants and also on the reaction each of them has to the stimulus task, the correlations of each of the physiological signals with the ISA subjective ratings (both mean normalized and individual values) will be presented in

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.696	0.484	<0.01	-0.535	0.286	<0.01
2	-0.197	0.039	0.22	0.016	0	0.92
3	-0.173	0.030	0.29	-0.183	0.033	0.26
4	0.47	0.221	<0.01	0.079	0.006	0.62
5	-0.276	0.076	0.08	-0.323	0.104	0.04
6	-0.573	0.328	<0.01	-0.454	0.206	<0.01
7	0.185	0.034	0.25	-0.202	0.041	0.21
8	-0.222	0.049	0.17	-0.198	0.039	0.22
9	-0.349	0.122	0.02	-0.327	0.107	0.04
10	-0.05	0.003	0.75	-0.112	0.013	0.49

Table 3 – R-R intervals correlated with subjective ISA reports

tabular form for each of the participants individually, together with strong and weak correlation example plots. This helps us understand whether any association between physiology and subjective ratings applies across a population or whether there are different levels of strength of relationships between different predictive variables in different populations.

Table 3, shows the correlation of the R-R inter beat intervals with both the mean normalized ISA values and the individual ISA ratings; correlations with p values smaller than 0.05 are bolded. For three of the participants (1, 6 and 9), the R-R values were significantly correlated to both the mean normalized ISA and to their individual ISA ratings. A negative moderate correlation was found for participants 1 and 6 while participant 9 showed a weak correlation with the subjective ISA ratings. The R-R values for participants 1 and 6 showed a moderate negative correlation with their individual ISA ratings but not a significant correlation with the mean normalized values. Participant 4 was the only participant to show a positive significant correlation with the mean normalized values. Participant 4 was the only participant to show a positive significant correlation between R-R and mean ISA normalized. Figure 4 shows the R-R measure for participant 1 plotted against mean ISA normalized and individual ISA, representing an example of strong correlation while Figure 5 shows the same measures for participant 10, representing the weakest correlation.

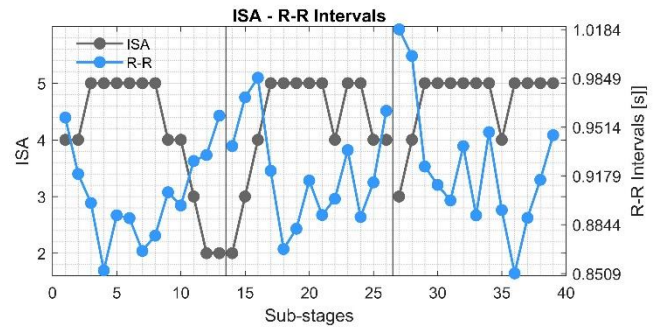
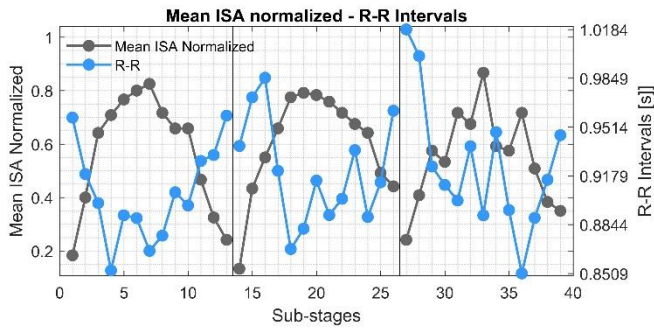


Figure 4 R-R – mean ISA normalized and individual ISA for participant 1 (strongest correlation)

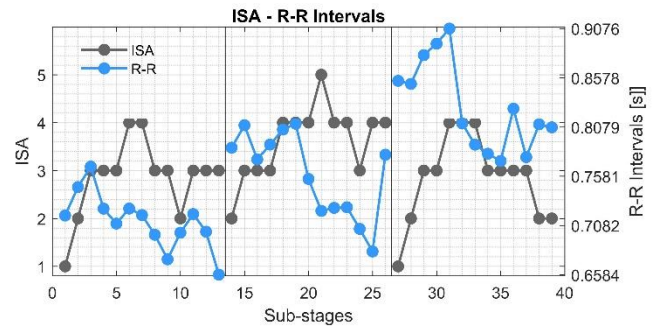
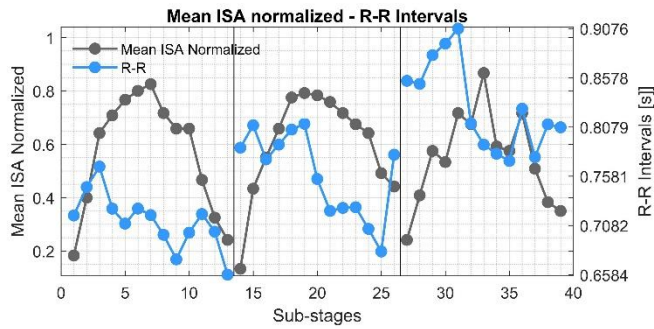


Figure 5 R-R – mean ISA normalized and individual ISA for participant 10 (weakest correlation)

Table 4 shows the correlations of pupil diameter with both the mean normalized ISA values and the individual ISA ratings; pupil diameter data from all participants except for 7 and 10 have moderate to strong positive correlations with the mean ISA normalized. Participants 7 and 10 show a weak positive correlation with the individual ISA ratings. Only participants 1 and 9 do not show a significant correlation to the individual ISA ratings. For most participants, a clear increase in pupil diameter was observed with the increase of workload.

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	0.617	0.381	<0.01	0.309	0.095	0.05
2	0.675	0.456	<0.01	0.544	0.296	<0.01
3	0.635	0.403	<0.01	0.497	0.247	<0.01
4	0.611	0.373	<0.01	0.677	0.458	<0.01
5	0.449	0.202	<0.01	0.35	0.123	0.02
6	0.705	0.497	<0.01	0.668	0.446	<0.01
7	0.268	0.072	0.09	0.435	0.189	<0.01
8	0.658	0.433	<0.01	0.601	0.361	<0.01
9	0.79	0.624	<0.01	0.073	0.005	0.65
10	0.308	0.095	0.05	0.489	0.239	<0.01

Table 4 – Pupil Diameter correlated with subjective ISA reports

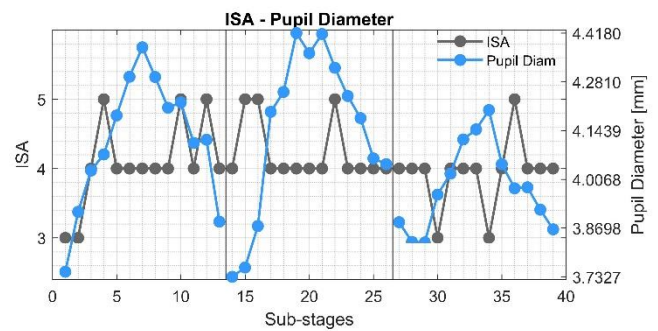
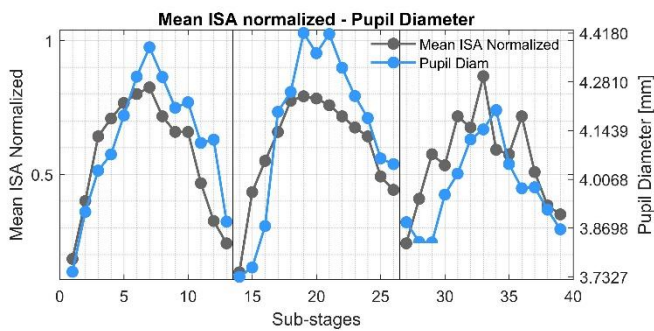


Figure 6 Pupil diameter – mean ISA normalized and individual ISA for participant 9 (strong correlation for mean ISA normalized but weak correlation for individual ISA)

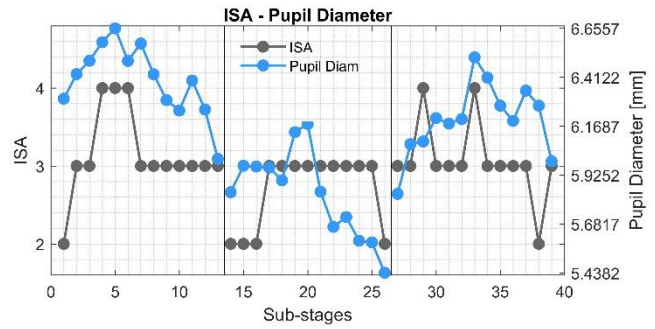
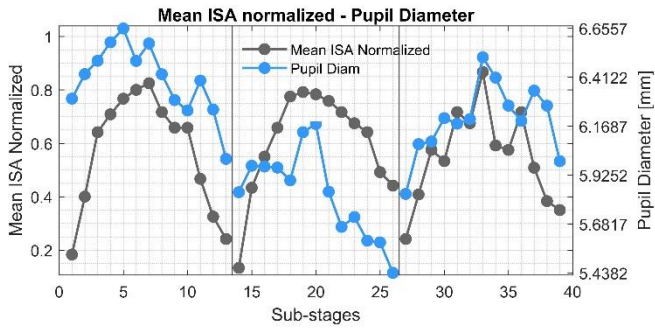


Figure 7 Pupil diameter – mean ISA normalized and individual ISA for participant 7 (non-significant correlation for mean ISA normalized but weak positive correlation with individual ISA ratings)

Table 5 shows the correlations of breathing rate with both the mean normalized ISA values and the individual ISA ratings; only the breathing rate data for participant 7 showed a moderate positive correlation with the mean normalized ISA values and a weak correlation with the individual ISA values. Participant 1 showed a moderate positive correlation between breathing rate and individual ISA ratings.

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	0.115	0.013	0.48	0.419	0.176	<0.01
2	0.0005	0	0.99	-0.144	0.021	0.38
3	-0.095	0.009	0.56	-0.169	0.029	0.30
4	0.061	0.004	0.71	0.242	0.059	0.13
5	0.097	0.009	0.55	0.038	0.001	0.81
6	-0.258	0.067	0.11	-0.122	0.015	0.45
7	0.661	0.437	<0.01	0.393	0.154	0.01
8	0.14	0.02	0.39	0.137	0.019	0.40
9	0.304	0.092	0.05	0.088	0.008	0.59
10	0.297	0.088	0.08	0.232	0.054	0.15

Table 5 – Breathing Rate correlated with subjective ISA reports

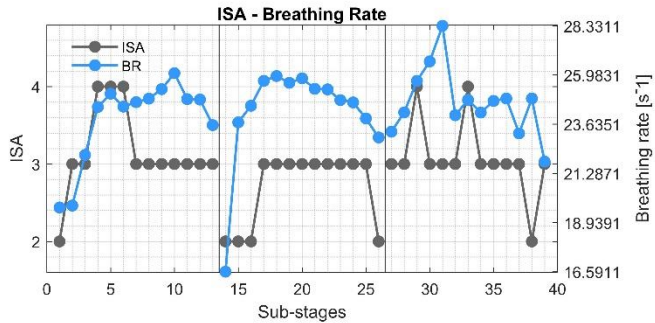
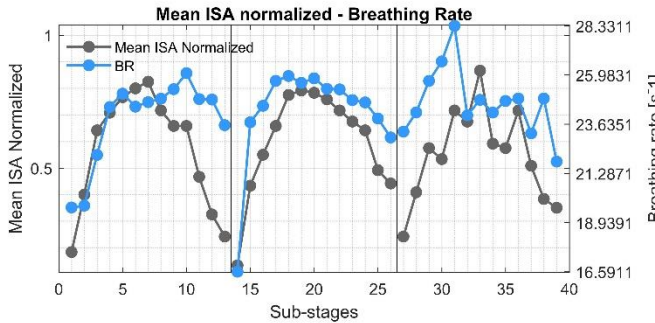


Figure 8 Breathing rate – mean ISA normalized and individual ISA for participant 7 (strongest correlation with mean ISA normalized)

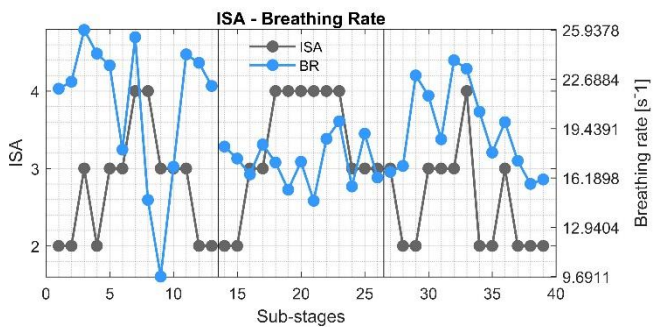
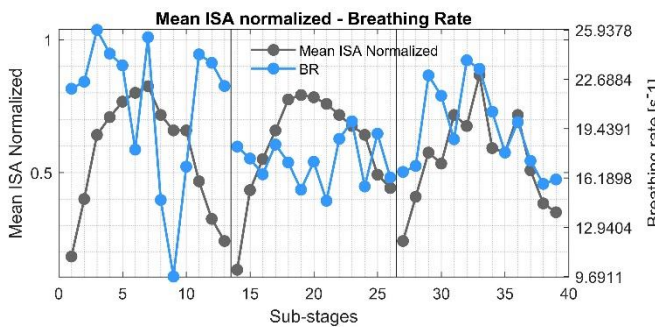
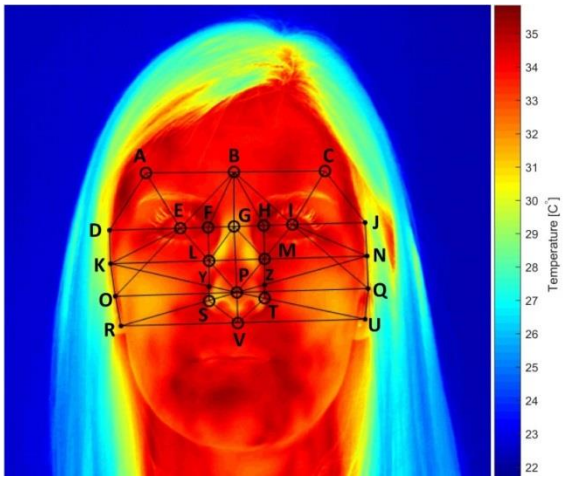


Figure 9 Breathing rate – mean ISA normalized and individual ISA for participant 2 (non-significant correlation example)

278 In order to extract the thermal data from the images, a feature
 279 tracking algorithm was deployed, splitting the face into regions
 280 of interest. For each frame, the temperature was extracted from
 281 inside the circular points, from along the lines and from inside
 282 some of the triangular areas without using markers, making the
 283 technique less intrusive (Fig. 6). Features from below the nose
 284 were not tracked due to the difficulty imposed by facial hair in
 285 some of the participants. The nose and forehead can therefore be considered as ideal sites for skin temperature measurement,
 286 as they would normally be un-occluded, which might present a challenge in a real life application as well.

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.746	0.557	<0.01	-0.507	0.257	<0.01
2	-0.373	0.139	0.01	-0.07	0.005	0.66
3	-0.075	0.006	0.64	-0.137	0.019	0.40
4	-0.152	0.023	0.35	-0.429	0.184	<0.01
5	-0.167	0.028	0.30	-0.008	0.000	0.95
6	0.345	0.119	0.03	0.188	0.035	0.25
7	-0.401	0.161	0.01	-0.459	0.211	<0.01
8	-0.086	0.007	0.60	-0.042	0.002	0.79
9	-0.514	0.264	<0.01	-0.208	0.043	0.20
10	-0.329	0.108	0.04	-0.028	0.001	0.86

Table 6 – Point P temperature correlated with subjective ISA reports



287
 288 Fig. 6. Feature tracking example

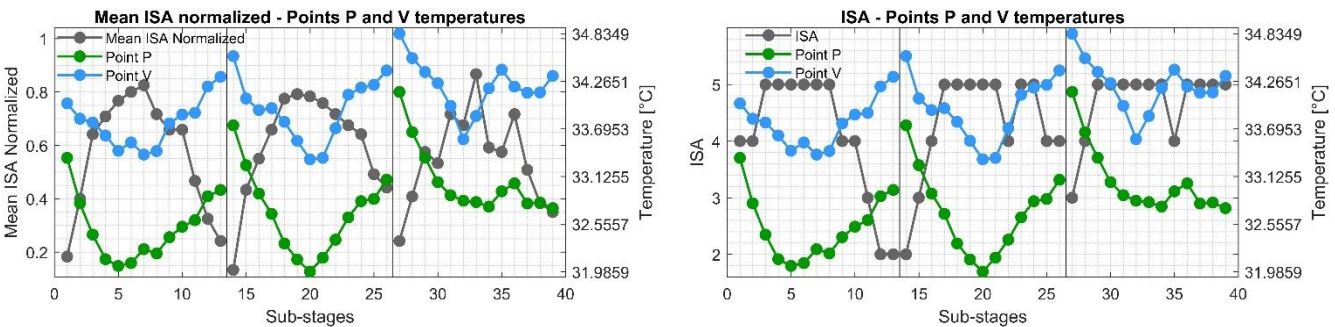
289 Table 6 shows the correlations of the average temperature inside
 290 point P (nose tip) with both the mean normalized ISA values and
 291 the individual ISA ratings; only participants 1 and 9 showed
 292 strong and moderate negative correlations to the 0.01 level for
 293 the mean ISA normalized. Participants 2, 7 and 10 showed weak
 294 negative correlations, significant to the 0.05 level with the mean
 295 ISA normalized values. Participant 6 was the only one to show
 296 a weak positive correlation with the mean ISA normalized values. Participants 4 and 7 showed stronger correlations with the
 297 individual (non-normalized) ISA ratings.

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.724	0.524	<0.01	-0.468	0.219	<0.01
2	-0.354	0.125	0.02	-0.05	0.003	0.76
3	-0.267	0.071	0.09	-0.1	0.010	0.54
4	-0.382	0.146	0.01	-0.381	0.145	0.01
5	-0.284	0.081	0.07	-0.132	0.017	0.41
6	0.146	0.021	0.37	-0.118	0.014	0.47
7	-0.296	0.088	0.06	-0.382	0.146	0.01
8	-0.107	0.011	0.51	-0.075	0.006	0.64
9	-0.035	0.001	0.83	-0.118	0.014	0.47
10	-0.307	0.094	0.05	-0.156	0.024	0.34

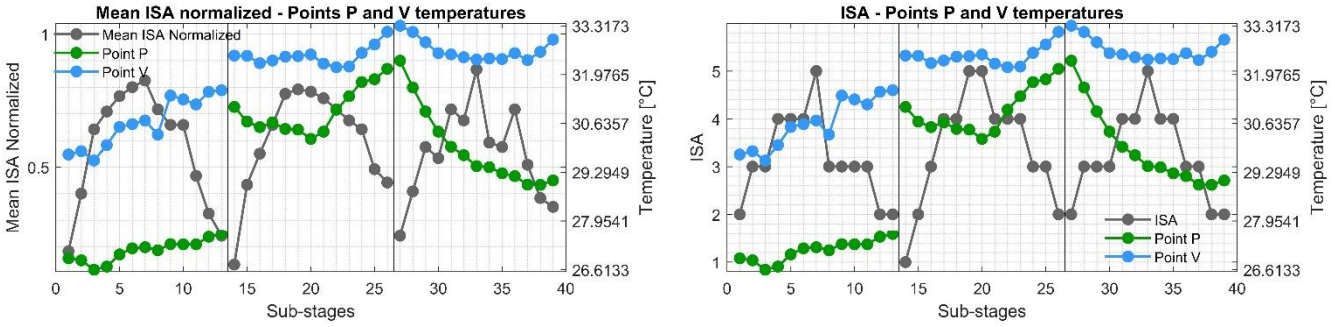
Table 7 – Point V temperature correlated with subjective ISA reports

298 Table 7 shows the correlations of the average temperature inside point V with both the mean normalized ISA values and the
 299 individual ISA ratings; only participant 1 showed a strong negative correlation to 0.01 level for the mean ISA normalized while

300 participants 2 and 4 showed a weak negative correlation, significant to the 0.05 level with the mean ISA normalized.
 301 Participants 1, 4 and 7 showed moderate to weak negative correlations with the individual ISA values.



302
 303 Figure 10 P, V points temperature – mean ISA normalized and individual ISA for participant 1 (strong correlation example)



304
 305 Figure 11 P, V points temperature – mean ISA normalized and individual ISA for participant 8 (non significant correlation
 306 example)

307 Table 8 shows the correlations of the average temperature inside
 308 point L with both the mean normalized ISA values and the
 309 individual ISA ratings; participants 1, 2, 9, 10 showed moderate
 310 negative correlations while participants 7 showed weak negative
 311 correlations with the mean ISA normalized levels. Participants
 312 4 and 7 showed a moderate to weak negative correlation with
 313 the individual ISA ratings.

Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.533	0.284	p<0.01	-0.249	0.062	p=0.12
2	-0.501	0.251	p<0.01	-0.292	0.085	p=0.07
3	-0.196	0.038	p=0.23	-0.006	0	p=0.96
4	-0.08	0.006	p=0.59	-0.471	0.222	p<0.01
5	-0.289	0.084	p=0.07	-0.155	0.024	p=0.34
6	-0.025	0.001	p=0.87	-0.023	0.001	p=0.88
7	-0.373	0.139	p=0.01	-0.377	0.142	p=0.01
8	-0.16	0.026	p=0.33	-0.082	0.007	p=0.61
9	-0.594	0.353	p<0.01	-0.157	0.025	p=0.33
10	-0.457	0.209	p<0.01	-0.169	0.029	p=0.30

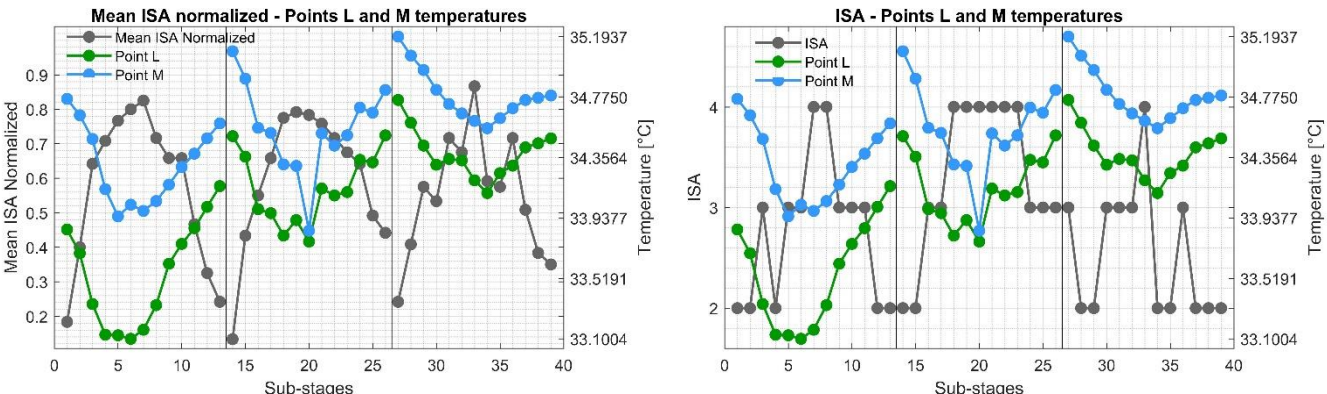
Table 8 – Point L temperature correlated with subjective ISA reports

314 Table 9 shows the correlations of the average temperature
 315 inside point M with both the mean normalized ISA values and
 316 the individual ISA ratings; participants 1, 2, 4, 7, 9 and 10
 317 showed a moderate negative correlation with the mean ISA
 318 normalized while participant 5 showed a weak negative
 319 correlation with the mean ISA normalized. Participants 2, 4 and

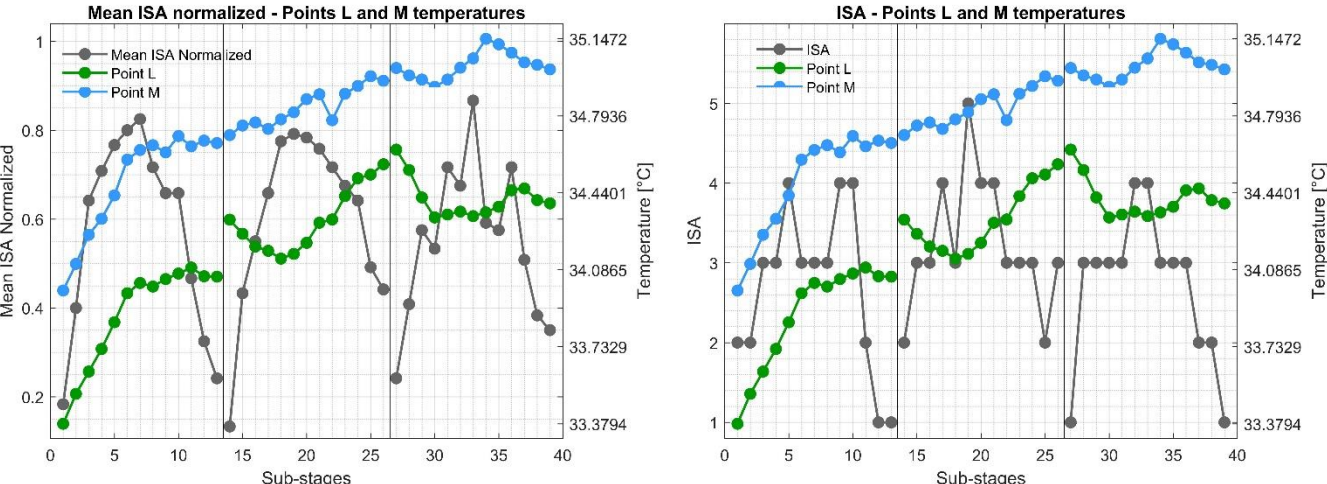
Participant	Mean ISA Normalized			Individual ISA		
	R(37)	R ²	p	R(37)	R ²	p
1	-0.472	0.223	p<0.01	-0.251	0.063	p=0.12
2	-0.674	0.454	p<0.01	-0.511	0.261	p<0.01
3	-0.081	0.007	p=0.62	0.076	0.006	p=0.64
4	-0.419	0.176	p<0.01	-0.509	0.259	p<0.01
5	-0.386	0.149	p=0.01	-0.248	0.062	p=0.12
6	0.116	0.013	p=0.48	0.069	0.005	p=0.67
7	-0.542	0.294	p<0.01	-0.358	0.128	p=0.02
8	-0.16	0.026	p=0.32	-0.071	0.005	p=0.66
9	-0.643	0.413	p<0.01	-0.186	0.035	p=0.25
10	-0.543	0.295	p<0.01	-0.584	0.341	p<0.01

Table 9 – Point M temperature correlated with subjective ISA reports

320 10 showed a moderate negative correlation with the individual ISA ratings while participant 7 showed a weak negative
 321 correlation to the individual ISA ratings.



322
 323 Figure 12 L, M points temperature – mean ISA normalized and individual ISA for participant 2 (strongest correlations)



324
 325 Figure 13 L, M points temperature – mean ISA normalized and individual ISA for participant 6 (non-significant correlations)

326 3.3. Predictive power of combined physiological measures

327 This section explores how the different physiological measures can be combined to produce the most accurate prediction of
 328 performance. Some of the measures presented above show promising correlations to the subjective ISA measure of mental
 329 workload.

330 A multiple linear regression was performed for each participant individually on more combinations of the predictor variables
 331 to test which one explains more of the variability in the response variable and how different physiological parameters can be
 332 combined for more reliable and valid capture of workload. Four combinations of the predictor variables were chosen:

- 333 1. Heart (R-R interval) and Breathing Rate data (Mean RR, Mean BR)
- 334 2. The Heart and Breathing Rate data and pupil diameter
- 335 3. The heart and breathing rate data, pupil diameter and the facial temperatures inside points : 'B', 'F', 'G', 'H', 'L', 'M',
 336 'P', 'V'

337 4. Facial temperatures inside points : 'B', 'F', 'G', 'H', 'L', 'M', 'P', 'V'

338 The reason behind the choice of the predictor variables combinations was to start with features from only one of the sensor and
 339 gradually add the others; category 1 contains just the features produced by the Zephyr sensor, category 2 adds pupil diameter
 340 to category 1, category 3 contains the combined features from the first two categories in addition to the facial thermography
 341 measures and category 4 contains just the facial thermography features.

342 Game performance, rather than ISA ratings, was selected as the response variable for this analysis. Game performance was
 343 used as it strongly correlates with the subjective ISA ratings and it is also a continuous variable; game performance is
 344 represented by the height the participants managed to maintain the yellow line on the screen.

345 Some of the predictor variables for some of the participants were highly correlated to each other. Inter-variable correlation
 346 influences the ability of multiple linear regression to distinguish between the predictive ability of each individual variable. Our
 347 approach to this limitation was to systematically add and remove predictors based on the F-statistic; the tool used for this was
 348 stepwise regression in Matlab. The algorithm starts with a constant model and iteratively adds and removes predictors until the
 349 model can no longer be improved substantially.

350 Each of the sections in Table 10 shows the multiple linear regression results for each of the described groups of predictors for
 351 each participant. The adjusted R^2 column, contains the proportion of variability of the dependent variable accounted for by the
 352 regression model. Because the R^2 value increases by adding more predictor variables in the model, the adjusted R^2 value was
 353 reported in order to make the comparison between models more meaningful. The table also displays the F statistic of the linear
 354 fit versus the constant model, testing the statistical significance of the model; the predictors column contains the names of the
 355 predictors selected by the algorithm for each of the regressions. The Beta column contains the estimate standardized coefficients
 356 of the terms in the regression, indicating how many standard deviations the dependent variable will change with the change of
 357 one standard deviation in the predictor variable, allowing for a comparison of the relative contribution of each of the predictors.
 358 The t-statistic test for the significance of each term given the other terms in the model is used to test the null hypothesis that
 359 the term is equal to zero (versus the alternate hypothesis that the coefficient is different from zero). The associated p values are
 360 also reported in the table.

Predictors	Participant	Adjusted R^2	RMSE	F statistic	p-value	Predictors	Beta	t-statistic	p-value
Combination 1:	1	0.404	0.77	13.91	<0.01	Mean RR	0.56	4.34	<0.01
						Mean BR	0.53	4.07	<0.01
Mean RR	2	0	-	-	-	-	-	-	-
Mean BR	3	0.187	0.9	9.75	<0.01	Mean BR	0.45	3.12	<0.01
	4	0	-	-	-	-	-	-	-

	5	0	-	-	-	-	-	-	-
	6	0.434	0.75	30.16	<0.01	Mean RR	0.67	5.49	<0.01
	7	0.265	0.85	14.76	<0.01	Mean BR	-0.53	-3.84	<0.01
	8	0	-	-	-	-	-	-	-
	9	0.224	0.88	11.97	<0.01	Mean RR	0.49	3.46	<0.01
	10	0.116	0.93	6.01	0.019	Mean BR	-0.37	-2.45	0.019
Combination 2: Mean RR Mean BR Pupil diameter	1	0.404	0.77	13.91	<0.01	Mean RR	0.56	4.34	<0.01
						Mean BR	0.53	4.07	<0.01
	2	0	-	-	-	-	-	-	-
	3	0.187	0.9	9.75	<0.01	Mean BR	0.45	3.12	<0.01
	4	0.338	0.81	20.43	<0.01	Pupil diameter	-0.59	-4.52	<0.01
	5	0.092	0.95	4.85	<0.05	Pupil diameter	-0.34	-2.2	<0.05
	6	0.434	0.75	30.16	<0.01	Mean RR	0.67	5.49	<0.01
	7	0.265	0.85	14.76	<0.01	Mean BR	-0.53	-3.84	<0.01
	8	0.280	0.84	15.84	<0.01	Pupil diameter	-0.54	-3.98	<0.01
	9	0.696	0.55	88.39	<0.01	Pupil diameter	-0.83	-9.4	<0.01
10	0.474	0.72	18.15	<0.01	Mean RR	0.45	3.82	<0.01	
					Pupil diameter	-0.62	-5.24	<0.01	
Combination 3: Mean RR Mean BR Pupil diameter Temperatures inside points: 'B' 'F' 'G' 'H' 'L' 'M' 'P' 'V'	1	0.786	0.46	36.03	<0.01	Pupil diameter	-0.33	-3.72	<0.01
						B	-1.19	-8.62	<0.01
						F	0.76	5.59	<0.01
						P	0.58	6.57	<0.01
	2	0.888	0.33	51.42	<0.01	Mean RR	-0.24	-2.14	<0.05
						Mean BR	0.32	4.4	<0.01
						G	-0.33	-2.83	<0.01
						M	0.82	9.17	<0.01
						P	-0.49	-3.84	<0.01
						V	0.81	7.58	<0.01
	3	0.915	0.29	103.78	<0.01	Pupil diameter	-0.13	-2.39	<0.05
						B	-1.59	-16.33	<0.01
						G	-0.66	-8.84	<0.01
H						1.34	18.46	<0.01	
B						-1.31	-12.4	<0.01	

	4	0.856	0.37	57.57	<0.01	F	0.26	2.61	<0.05	
						M	0.58	7.03	<0.01	
						V	0.21	2.87	<0.01	
	5	0.692	0.55	29.51	<0.01	Mean BR	0.28	2.54	<0.05	
						F	-0.26	-2.82	<0.01	
						M	0.85	7.51	<0.01	
	6	0.763	0.48	25.6	<0.01	Mean RR	0.62	5.47	<0.01	
						Pupil diameter	-0.27	-2.7	<0.05	
						B	1.47	5.89	<0.01	
						F	-1.38	-7.35	<0.01	
	7	0.787	0.46	24.49	<0.01	P	-0.32	-2.35	<0.05	
						Mean RR	-0.26	-2.46	<0.05	
						Mean BR	-0.3	-3.03	<0.01	
						Pupil diameter	-0.24	-2.34	<0.05	
						L	0.37	2.43	<0.05	
	8	0.712	0.53	24.5	<0.01	M	1.51	6.85	<0.01	
						P	-1.46	-5.01	<0.01	
						Pupil diameter	-0.67	-6.56	<0.01	
						G	-0.63	-2.8	<0.01	
	9	0.841	0.39	51.43	<0.01	M	1.65	4.83	<0.01	
						P	-1.47	-7.38	<0.01	
						Mean BR	-0.2	-3.01	<0.01	
						Pupil diameter	-0.78	-7.49	<0.01	
	10	0.7	0.54	30.55	<0.01	G	0.34	3.17	<0.01	
						P	-0.45	-5.58	<0.01	
						Pupil diameter	-0.62	-6.54	<0.01	
	Combination 4: Temperatures inside points: 'B' 'F' 'G' 'H' 'L' 'M' 'P' 'V'	1	0.708	0.54	31.71	<0.01	V	0.26	2.53	<0.05
							G	0.49	5.01	<0.01
							B	-0.94	-6.65	<0.01
		2	0.818	0.42	43.77	<0.01	F	0.58	3.92	<0.01
P							0.68	6.94	<0.01	
G							-0.33	-2.63	<0.05	
3		0.903	0.3	120.15	<0.01	M	1.01	10.1	<0.01	
						P	-0.76	-5.22	<0.01	
						V	0.55	4.8	<0.01	
					B	-1.55	-15.18	<0.01		
					G	-0.64	-8.09	<0.01		
					H	1.38	18.33	<0.01		
					B	-1.31	-12.4	<0.01		

4	0.856	0.37	57.57	<0.01	F	0.26	2.61	<0.05
					M	0.58	7.03	<0.01
					V	0.21	2.87	<0.01
5	0.679	0.56	27.81	<0.01	F	-0.55	-3.68	<0.01
					H	0.33	2.17	<0.05
					M	0.52	4.42	<0.01
6	0.641	0.59	18.02	<0.01	B	0.79	2.61	<0.05
					F	-1.5	-6.66	<0.01
					G	-0.93	-4.25	<0.01
7	0.724	0.52	20.97	<0.01	L	1.14	5.73	<0.01
					B	-0.49	-3.27	<0.01
					G	-0.97	-3.88	<0.01
					M	1.27	5.04	<0.01
8	0	-	-	-	-		-	-
					P	-0.83	-3.07	<0.01
9	0.564	0.65	25.63	<0.01	G	0.89	6.91	<0.01
					P	-0.3	-2.33	<0.05
10	0.574	0.65	18.1	<0.01	G	1.29	6.71	<0.01
					L	-1.21	-3.9	<0.01
					P	0.58	2.46	<0.05

Table 10 Proportion of the variability accounted for by the regression model in the response variable

The results presented in Table 10 show that for combination 3, when using all the predictor variables, for 7 out of 10 participants, the pupil diameter measure was demonstrated to be a good predictor of performance, followed by temperature in point P for 6 out of 10 participants. On average, facial thermography measures added 47.7% to the amount of variability explained by the regression model.

Figure 14 below shows a boxplot summary of table 10 in terms of adjusted R^2 and RMSE. It can be seen that for the predictors in combination 3 the amount of variability explained is higher than all other combinations but close to combination 4. At the same time, the RMSE is smallest for combination 3, indicating a better fit compared to the other models. Based on the data collected in this study, for most of the participants, pupil diameter together with thermal data measured around the nose area provided the best combination of predictors for inferring the level of performance.

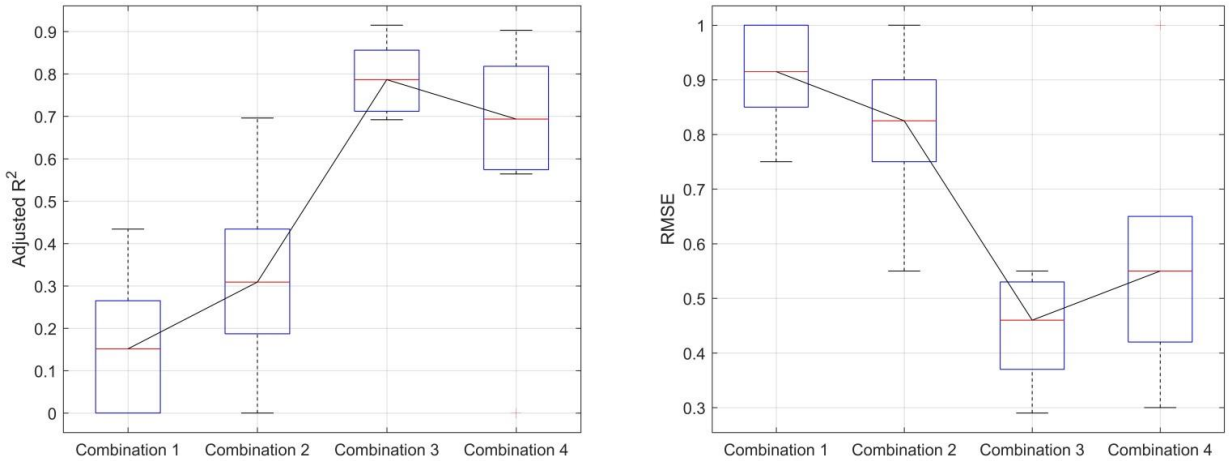


Figure 14 Adjusted R^2 and RMSE for each of the four combinations of predictors

7. DISCUSSION

This research presents novel insights into the relative value of physiological and subjective techniques for assessment of workload and human performance. The main novelty lies in the fact that multiple continuous physiological measures were recorded and synchronized with task performance and subjective ratings. The hypotheses explored in this study were:

1. There will be a measurable difference in subjective workload between the two levels of task difficulty

This hypothesis was found to be true: The mental demand measured using NASA-TLX confirmed that there was a measurable difference between the two levels of difficulty and stage 2 was perceived to be more mentally demanding than stages 1 and 3.

2. The subjective ratings of workload will be associated with changes in physiological measures

Hypothesis 2 was partially proved: The study explored which physiological measures showed a change in accordance to the change in mental workload as measured subjectively on the ISA scale. It was found that for some of the participants, the mean normalized ISA ratings showed a stronger correlation with some of their physiological measures than it did with the individual ISA rating.

Table 11 summarizes the results by displaying the number of participants that showed moderate to strong correlations with mean ISA normalized or individual ISA ratings for each of the physiological measures presented above. Overall, the correlations of the thermal data with the individual ISA ratings were weaker for all participants.

Measure	No. of participants showing moderate to strong correlations	
	Mean ISA Normalized	Individual ISA
R-R Intervals	3/10	2/10
Breathing Rate	1/10	1/10
Pupil Diameter	8/10	7/10
Point P Temperature	3/10	3/10
Point V Temperature	1/10	1/10
Point L Temperature	4/10	1/10
Point M Temperature	6/10	3/10

Table 11 – No. of participants showing moderate to strong correlations with the ISA rating

3. Multiple physiological measures can be used in combination to analyze workload

Hypothesis 3 was tested by using a multiple linear regression on the data from each of the participants, showing that when using facial thermography data is combined with other physiological data, the predictive model explains on average 47.7% more of the variability in performance compared to solely using a combination of R-R inter-beat intervals, breathing rate and pupil diameter. As mean performance across the participants was strongly correlated with the mean ISA normalized, it is an indication that these physiological measures could also provide good prediction results for the level of subjectively experienced mental workload.

402 In their discussion section, Ora & Duffy (2007) recommended that further examination under more controlled conditions and
403 the test of additional psychophysiological measures such as pupil dilation should be performed in the hope of developing a
404 more robust approach to the estimation of mental workload in a non-invasive way. In this study, the variation of demand was
405 done in more controlled conditions and additional physiological measures (such as heart rate, breathing rate and pupil diameter)
406 were collected and their relative group contribution was tested. In terms of facial thermography, the landmark tracking was
407 done automatically and included more areas of the face. One of the limitations of the study was the small number of participants;
408 for the limited number of participants (10), there was no physiological measure that proved to work best at predicting mental
409 workload or performance levels across all participants. Although from a physiological point of view people responded
410 differently when being subjected to the type of demand induced by the task, some of the physiological measures, especially
411 pupil diameter and temperatures in points G, M and P, proved to be good and consistent indicators of the level of performance
412 (and implicitly the level of demand) for more than half of the participants.

413 Further studies will concentrate on the collection of more data in environments closer to the real workplace setting and the use
414 of machine learning algorithms to improve prediction accuracy, confirm feasibility of applying the physiological and analytical
415 methods in situ, and ensure generalisability of results. Future work should also consider how facial thermography
416 measurements would vary over longer time periods than have been examined in this study.

417 The results presented in this paper demonstrate that physiological measures, especially face temperature and pupil diameter,
418 can be used for non-invasive real-time measurement of workload when combined with a facial landmark tracking algorithm,
419 assuming models have been appropriately trained on previously recorded data from the user population. This is a feasible
420 proposition in a setting such as cockpits.

421 The demonstration of feasibility of physiological measures as a method as presented within this paper allows the identification
422 of guidance for how this approach can be used in the future, and requirements for further research. The methods presented in
423 this article, with current technological capabilities, are better suited for workplaces where the subject is seated, but the methods
424 can cope with a limited amount of head movement. Continuous real time non-invasive workload measurement techniques is
425 now a realistic proposition that will allow for improved design of human-machine systems, operating procedures and operations
426 scheduling in ways that will bring us closer to the goal of optimizing human well-being and overall system performance.

427 428 Acknowledgments

429 The authors would like to thank the European Union for founding this research from the People Programme (Marie Curie
430 Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no 608322
431 and the Data Analysis and Interaction research team from Airbus Group Innovations UK for their support. We would also like

432 to thank Dr Robert Houghton for offering insightful comments that greatly improved the quality of the manuscript.

433 **Key Points**

- 434 • One of the challenges posed by the future of air transportation, from a human factors perspective, is evaluating the
435 level of mental workload to which the operators are subjected
- 436 • Some methods of workload assessment have been difficult to implement in-situ in a real work environment due to
437 being intrusive (e.g. interrupting task or requiring uncomfortable equipment to be worn).
- 438 • This paper explores multiple physiological measures and their relative significance as indicators of performance and
439 mental workload, demonstrating the feasibility of physiological measures as a method of evaluating the level of
440 mental workload in real time in a non-invasive manner.

442 **References**

- 443 Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload.
444 *International Journal of Industrial Ergonomics*, 36(7), 623–636. <https://doi.org/10.1016/j.ergon.2006.04.002>
- 445 Airbus. (2015). Flying By Numbers, 1–27.
- 446 Brennen, J. (1997). Instantaneous self-assessment of workload technique (ISA), (1992).
- 447 FLIR. (2012). The Ultimate Infrared Handbook for R & D Professionals The Ultimate Infrared Handbook for R & D
448 Professionals, 4–31. Retrieved from <http://www1.flir.com/e/5392/research-development-guidebook/x7hz7/869543448>
- 449 Hart, S. G., California, M. F., & Staveland, L. E. (1988). Development of NASA TLX.
- 450 Lehrer, P., Karavidas, M., Lu, S., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association
451 between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory
452 study. *International Journal of Psychophysiology*, 76(2), 80–87. <https://doi.org/10.1016/j.ijpsycho.2010.02.006>
- 453 Murai, K., Hayashi, Y., Okazaki, T., & Stone, L. C. (2008). Evaluation of ship navigator's mental workload using nasal
454 temperature and heart rate variability. *2008 IEEE International Conference on Systems, Man and Cybernetics*, 1528–
455 1533. <https://doi.org/10.1109/ICSMC.2008.4811503>
- 456 Ora, C. K. L., & Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental
457 workload measurement, 7, 83–94.
- 458 Parasuraman, R., & Mehta, R. (2015). Neuroergonomic Methods for the Evaluation of Physical and Cognitive Work. In J. R.
459 Wilson & S. Sharples (Eds.), *Evaluation of Human Work* (pp. 609–638).
- 460 Sensomotoric Instruments. (2011). iView X System Manual. *Update*, (January).
- 461 Sharples, S., Edwards, T., & Balfe, N. (2012). Inferring cognitive state from observed interaction. In *Proceedings of the 4th*

462 *AHFE International Conference.*

463 Sharples, S., & Megaw, T. (2015). Definition and measurement of human workload. In J. R. Wilson & S. Sharples (Eds.),
 464 *Evaluation of Human Work.*

465 Stemberger, J., Allison, R. S., & Schnell, T. (2010). Thermal Imaging as a Way to Classify Cognitive Workload. *2010*
 466 *Canadian Conference on Computer and Robot Vision*, 231–238. <https://doi.org/10.1109/CRV.2010.37>

467 Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human
 468 subjects. *Investigative Ophthalmology & Visual Science*, *35*(3), 1132–7. Retrieved from
 469 <http://www.ncbi.nlm.nih.gov/pubmed/8125724>

470 Young, M. S., Brookhuis, K. a, Wickens, C. D., & Hancock, P. a. (2014). State of science: mental workload in ergonomics.
 471 *Ergonomics*, *139*(January 2015), 1–17. <https://doi.org/10.1080/00140139.2014.956151>

472

473 **Biographies of co-authors:**

474 Adrian Cornelius Marinescu:

- 475 • Current affiliation: University of Nottingham
- 476 • Highest Degree Obtained: MSc – Alternative and Renewable Energy Sources – University of Bucharest, Faculty of
 477 Physics, 2012

478 Prof. Sarah Sharples:

- 479 • Current affiliation: University of Nottingham
- 480 • Highest Degree Obtained: PhD – Human Factors – University of Nottingham , 1999

481 Dr. Alastair Campbell Ritchie:

- 482 • Current affiliation: University of Nottingham
- 483 • Highest Degree Obtained: PhD – Bioengineering – University of Strathclyde, 1997

484 Dr. Tomas Sanchez Lopez:

- 485 • Current affiliation: Airbus Group Innovations
- 486 • Highest Degree Obtained: PhD – Computer Science – Korean Advanced Institute of Science and Technology (South
 487 Korea), 2008

488 Michael McDowell:

- 489 • Current affiliation: Airbus Group Innovations
- 490 • Highest Degree Obtained: MEng – Engineering Mathematics – University of Bristol, 2013

491 Prof. Herve P. Morvan:

- 492 • Current affiliation: University of Nottingham
- 493 • Highest Degree Obtained: PhD – Computational Fluid Dynamics (CFD) – Glasgow University, 2001