Occupational Ergonomics 7 (2007) 83–94
IOS Press

Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement

Calvin K.L. Or[a] and Vincent G. Duffy[b,*]
[a]*Department of Industrial & Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*
[b]*School of Industrial Engineering and Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA*

Abstract. The research aimed at developing a non-intrusive physiological measure for mental workload using human facial skin temperature change. It demonstrated initial results in two driving experiments that showed the potential of using this physiological parameter to infer mental workload. Participants completed driving tests in a simulator in the first experiment. Results of simulator and real vehicle testing were used in a second experiment. Forehead and nose temperature were obtained via thermography. Nose temperature dropped significantly after the drives for all conditions in the simulator tests. A secondary task during driving led to higher subjective workload score and a greater nose temperature drop. Simulator drives led to a higher subjective workload score and a greater nose temperature drop than the real driving task. A significant correlation between the nose skin temperature change and the subjective workload score was yielded in both experiments. Potential applications of this research include real-time, non-intrusive, and automated mental workload assessment for advanced human-system interface development and performance prediction.

Keywords: Facial skin temperature, mental workload, non-intrusive physiological measure, thermography

1. Introduction

With the advent of the complex systems that have multiple and competing task demands, operator mental workload has been recognized as an important concern in relation to human errors and human-systems failures. Excessive mental workload has negative impact on performance and can lead to errors with disastrous outcome. In the domain of ground transportation, drivers with high mental workload can have reduced alertness, diverted attention, and inadequate time for information processing [4]. Those are significant contributors to road-related accidents [20].

Operator mental workload can be assessed by subjective measures, such as the Modified Cooper-Harper (MCH) scale [26]. However, subjective measures do not allow one to understand operator's workload level in task performance without either a loss of information during the task, or some intrusion. Physiological measurement techniques, such as heart rate variability, pupil size, blood pressure variability

---

[*] Address for correspondence: Vincent G. Duffy, School of Industrial Engineering and Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA. Tel.: +1 765 496 6658; Fax: +1 765 494 1299; E-mail: duffy@purdue.edu.

and visual demand, have been widely used to infer operator's mental state in a quantified manner [1, 2,4,9,19,22]. However, these assessment tools are sometimes intrusive. A non-intrusive physiological measure, facial temperature change, is considered in this study. The results suggest a multi-modal approach that incorporates measurement of facial temperature could be considered in the future.

Investigation of the change of human facial skin temperature in relation to mental workload is an area in which little study has been done to date. This present study attempted to explore the use of facial skin temperature change as an objective index of mental workload. Few related works could be found in the literature [8]. The previous research focused on the ability to create an individual modulus for establishing a relationship between workload and facial temperature, rather than a general measure applicable across a sample or population [8]. Recently there are some scientists in Europe and Asia who have shown similar interest and perspective in recent conferences [10,23]. They are again investigating possible relationships between the physiological measure of facial temperature and workload [10,23].

Prior to the beginning of the investigation via experiments, a review of the psychological factors and human skin temperature literature provides some insight into how the physiological response is influenced by events in the human brain. A reduction of the skin temperature in the nasal area has been found after the exposure to task load during the monitoring and manual-tracking task with keyboard and trackball [8,14]. To create variability in their task, a bell-ring noise was presented to the subjects and they were required to correctly recall and enter the previously learned 12-character password to stop the noise. The resulting nose temperature drop was possibly due to the activity alteration in the Autonomic Nervous System (ANS) [8]. Naemura and colleagues [14] adopted the technique of loud noise presentation as a task load event presented to the subjects. For those who were presented the 100 dB-noise for two minutes, their nasal temperature began to drop after the onset of the stimulus presentation. The nasal temperature continued to decrease and started to recover about 1 minute after the offset of the stimulus. It was explained that the temperature drop was ascribed to the vasoconstriction response of the ANS. The underlying mechanism is that mental stress or negative emotion can induce peripheral metabolic responses, which are mediated primarily by the sympathetic nervous system. When the sympathetic nervous system is activated, vasoconstriction occurs. Vasoconstriction leads to a reduction of blood flow in the peripheral capillary vessels, and consequently the nose skin temperature decreases [25].

The findings in Genno and colleagues [8] and Naemura and colleagues [14] also revealed the stability of forehead skin temperature. The forehead skin temperature remained unchanged in the experiments. As the human forehead skin temperature is known as the most stable of any of the body surface [21], the temperature of that area is expected to remain fairly stable even though task load was imposed.

In this present study, the driving tasks required monitoring (e.g., observing the road and entities on the road) and manual-tracking (e.g., controlling the steering wheel) in engaging the task. The experiments were performed in both the simulated and real car driving environments. In Experiment 1 the driving task took place in a simulated driving environment in which the driving courses were a city-like and highway-like terrains. For the purpose of creating different complexity level of the task, the participants were asked to perform a secondary task while driving. The secondary task used was a mental arithmetic test. This secondary task technique had been widely used in mental workload research [5,17,24]. In experiment 2, data from the simulator and real vehicle driving conditions were compared for the same drivers. The experiment was intended to demonstrate the facial skin temperature measure in a field and a laboratory setting. Also, it attempted to test the effect of different driving environments on facial temperature change and subjective workload.

Changes in facial skin temperature were expressed as differences of the initial and final temperature ([initial: temperature obtained immediately pre-stimulus]-[final: temperature obtained immediately at the end of stimulus]). The MCH questionnaire was administered to the drivers after the driving tasks and results were compared to facial temperature changes. Prior to testing, the methodology was approved by the Institutional Review Boards.

2. Experiment 1

Introducing a mental arithmetic test as a mental loading task (MLT) to the mental workload experiment is a common technique in human factors transportation research. To manipulate the mental workload level, a mental arithmetic test has commonly been added into driving studies [5,17,24]. Subjective mental workload and physiological measures appeared to be sensitive to concurrent tasks in which a mental arithmetic test is used as a secondary task [24].

In this present study, two simulated carriageways (i.e., city-like and highway-like) were created for the test. The four experimental conditions were randomly presented to the participants. The conditions were city-like with and without the MLT and highway-like with and without the MLT. In order to evaluate the possible dual-task effect on the facial temperature response and subjective workload, a comparison was made of the mean values between the levels was performed. It was expected that driving in the MLT conditions would provoke a larger effect on the nose temperature change and the perceived workload level.

2.1. Method

2.1.1. Participants

Thirty-three healthy licensed drivers (16 males, 17 females) participated in the experiment. Of the thirty-three participants, 16 were categorized as younger ranged from 18 to 35 years (mean age = 24.2; $\sigma = 4.9$) and 17 as older ranged from 36 to 64 years (mean age = 48.8; $\sigma = 7.1$). Participants were recruited with newspaper advertisements, handouts, and posters. All participants reported having normal or corrected vision and having no color blindness. They were paid at the rate of $30 per hour for their participation. All participants self reported annual mileages to be at least 2,000–7,999.

2.1.2. Apparatus

The infrared thermal imaging camera MikroScan 7200V (Mikron) was used to measure the participants' facial temperature. The camera was installed on a tripod and was positioned nearby the participants who were sitting in the driving simulator. The sensitivity/NETD (Net Equivalent Temperature Difference) of the thermal imaging camera was 0.08°C.

The simulator tests were carried out with a desktop driving simulator that was purchased from DriveSafety Inc. The simulator was comprised of a portable built up cab with adjustable real vehicle seat and real vehicle steering wheel attached to a platform. Also included were two computer systems that were mounted in a rack. Images of the driving scenarios were projected onto a four foot diagonal screen. Two speakers located on each side of the simulator were used to deliver driving sounds. The city-like scenario included metropolitan scenes of buildings, crowded intersections and surrounding traffic. It was created as a two-lane, two-way circuit formed by four right-turn intersections with a traffic light at each of the turns. The highway-like scenario was a four-lane, two-way road with the classic characteristics of a freeway. For the two MLT conditions, participants were required to perform the mental addition test while driving. The mental test was a 2-digit addition task in which the numbers for the problems were generated using a random number generator. The MCH scale was administered with paper and pen.

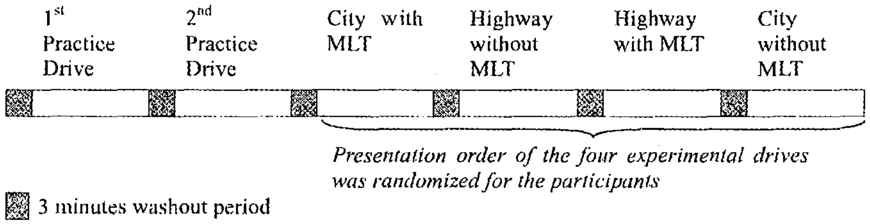


Fig. 1. Sequence of the practice drives, washout period, and the experimental conditions for a participant in Experiment 1.

2.1.3. Procedure

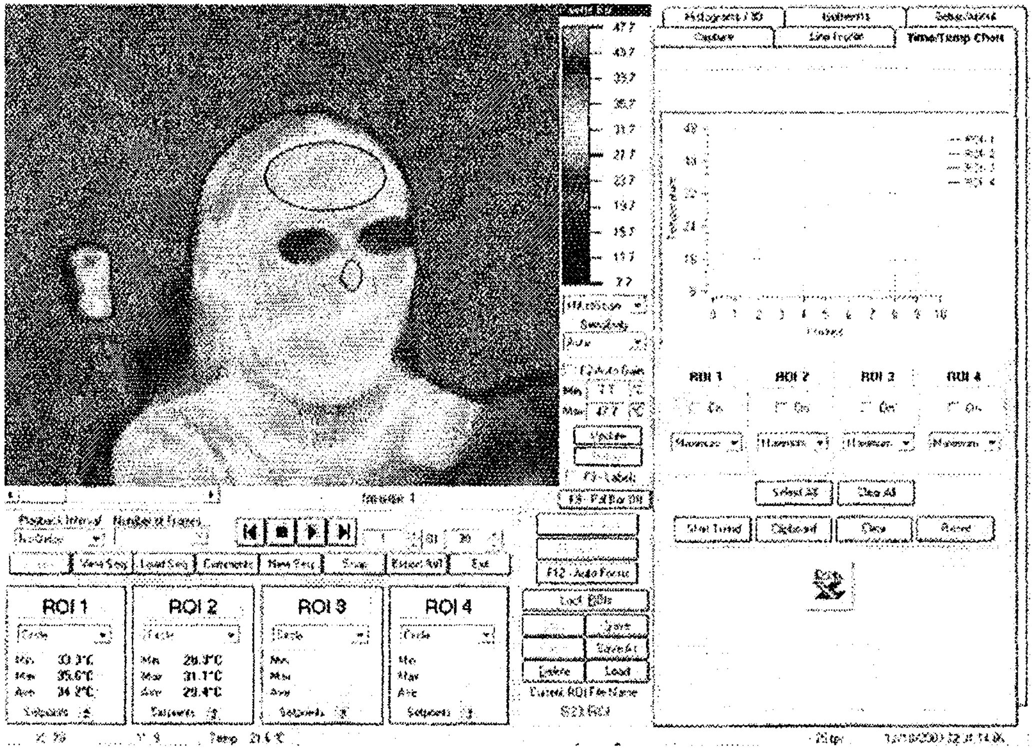
The participants drove in each of the four experimental conditions after the completion of the two practice drives in a temperature-controlled laboratory with thermostat set at 22°C. The presentation order of the experimental conditions was randomized. The driving test for each condition lasted approximately 3 minutes, with a rest period of about 3 minutes allocated before each next trial. The rest period was assigned between two consecutive trials to wash out the residual effect of any facial temperature from the previous trial. Participants were asked to perform the practice drive twice prior to the actual test. The sequence of the experimental condition presentation for a participant in Experiment 1 is illustrated in Fig. 1.

In the experimental conditions with the MLT, the participants were asked to accomplish the mental addition test as quickly as possible as long as they still complete the primary driving task successfully. They were asked to mentally solve and report their answer out loud. A tape recorder placed in the experimental laboratory was used to deliver a female voice recording of the 2-digit numbers at intervals of 10 seconds. The facial areas of interest were nose and forehead. Temperature values were obtained for the analyses by placing the Regions of Interest (ROI) on the thermal images (see Fig. 2) in the computer software. The software computed the temperature in the ROI for data analysis. The temperatures of the facial areas were measured immediately pre-stimulus and right after the completion of a drive. The MCH questionnaire was administered after each drive (test condition) was completed.

2.2. Data analysis

Three primary steps were adopted to examine the facial skin temperature responses and mental workload throughout the two experiments of this study. First, initial temperatures were compared between different conditions to confirm that the washout period was sufficient to eliminate any residual effect of the previous test condition. Then, a comparative analysis of the initial and final temperatures was performed in order to determine whether the facial areas would exhibit any noticeable changes during the tasks. The third part of the analysis was to examine whether the facial temperature change was different between different experimental conditions. The temperature change was obtained by subtracting the temperature measured immediately at the end of stimulus from the temperature measured immediately pre-stimulus. The associated subjective workload score (MCH) was also compared between conditions.

A distribution free test was performed to determine the correlation between the objective measure and the subjective measure afterward. The data from the subjective workload test did not follow a normal distribution. Therefore, some additional post processing was done considering methods used to treat data that is psychophysical (psychological response to a physical input). Such data is typically log normal, skewed and has increasing error. Individuals tend to have scales that are different from one another.



ROI 1: Min 33.3°C, Max 35.6°C, Ave 34.2°C

ROI 2: Min 28.3°C, Max 31.1°C, Ave 29.4°C

Fig. 2. Regions of Interest (ROI) on the thermal image.

In order to correct the data, one would need to consider a subset of the six steps approach used by Snow and Williges [18]. In this case, the data was not lognormal, so we did not take the log transfer on both axes, nor the antilog. However, it was necessary to create an individual modulus from individual mean, create a common modulus from mean of all responses, subtract the common modulus from individual modulus, and then subtract that individual modulus from all individual scores to create a more normally distributed data set which was at least adjusted for a individual differences in scaling.

2.3. Results

2.3.1. Facial temperature response

No significant difference was shown in the mean initial forehead ($\chi^2 = 5.55$, $p = 0.14$) and nose ($\chi^2 = 1.17$, $p = 0.76$) temperature across the four trials. This confirms that the rest period between the successive trials was long enough to washout the residual on the facial temperatures from a previous trial. Figure 3 showed the initial forehead and nose temperature of the four experimental trials.

To understand the effect of the driving task on forehead temperature in each condition, the initial and final temperature was compared. The workload tasks had no significant effect on forehead temperature (highway without MLT: $\chi^2 = 0.0012$, $p = 0.71$; highway with MLT: $\chi^2 = 1.09$, $p = 0.30$; city without MLT: $\chi^2 = 1.19$, $p = 0.28$; city with MLT: $\chi^2 = 3.56$, $p = 0.06$). The temperature remained stable during the 3-minute driving task in the simulator test.

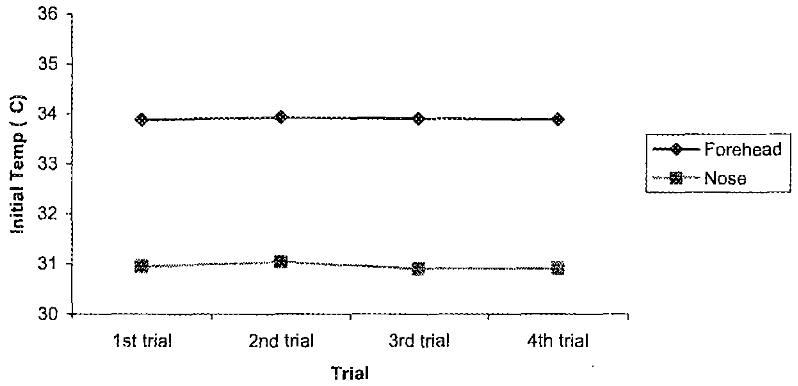


Fig. 3. Initial ($t = 0$) forehead and nose temperature of the four trials of Experiment 1.

Comparative analysis of nose temperature before and after driving in the simulator showed consistency in all experimental conditions. The distribution-free Friedman test for the nose temperature showed a significant change after completing tasks for all conditions (highway without MLT: $\chi^2 = 13.13$, $p < 0.0001$); highway with MLT: $\chi^2 = 28.13$, $p < 0.0001$; city without MLT: $\chi^2 = 16.33$, $p < 0.0001$; city with MLT: $\chi^2 = 28.13$, $p < 0.0001$). After the experimental tasks, the nose temperature dropped by 0.32°C ($\sigma = 0.36$°C) for *highway without MLT*, 0.52°C ($\sigma = 0.51$°C) for *highway with MLT*, 0.34°C ($\sigma = 0.37$°C) for *city without MLT*, and 0.56°C ($\sigma = 0.55$°C) for *city with MLT*. A significant difference was found for the mean nose temperature drop across the four conditions ($\chi^2 = 9.20$, $p = 0.03$). The two dual-task conditions (i.e., city with and high with MLT) provoked a significantly greater drop of nose temperature than did those either in city or highway drive without the secondary task. The preceding two groups were not significantly different, and neither were the latter two groups. Age had no significant effect on nose temperature drop. No significant interaction was found between condition and age.

2.3.2. Subjective workload score

The mean perceived mental workload for the *city with MLT condition* was 6.21 ($\sigma = 2.51$); *highway with MLT* was 5.88 ($\sigma = 2.34$); *city without MLT* was 3.39 ($\sigma = 1.37$); *highway without MLT* was 3.18 ($\sigma = 1.33$). The ANOVA was performed on the workload scores. It yielded a significant effect for conditions ($\chi^2 = 64.24$, $p < 0.0001$). The post hoc test showed that the subjective workload score for *city with MLT* condition and *highway with MLT* condition was significantly higher than in the *city without MLT* and *highway without MLT* conditions. There were no significant differences in the workload scores among the two dual-task conditions nor the two single-task conditions. Age had no significant effect on the perceived workload. Again, there was no significant interaction found between condition and age.

2.3.3. Correlation

The Spearman correlation analysis is used to compare the nose temperature drop against the MCH score. The test result showed a significant correlation between the nose temperature drop and the subjective workload score ($r = 0.32$, $p = 0.009$).

2.4. Discussion for experiment 1

The findings revealed that an increase in the MCH ratings correlated with the nose skin temperature drop while the participants' forehead temperature remained stable throughout the tasks.

The introduction of the mental loading task significantly increased the perceived workload and engendered a larger drop of the nose skin temperature. Since other influential factors that could lead to facial temperature change had been controlled for all trials, the significant larger temperature drop could be ascribed to the larger mental demand of the tasks. It appears that the mental load associated with the experimental tasks led the nose temperature to drop and is possibly due to the vasoconstriction response mediated by the ANS [8,14].

It should be noted that though the correlation is significant when including a whole sample of subjects, but it is not large for this experiment. It is also important to note that the range of mental workload tested was also not large relative to the whole scale of the MCH available in the test. It is possible that the range of temperature change tested for the short driving test is reaching the threshold of the sensitivity range for the differences that can be detected reliably with the current instrumentation.

3. Experiment 2

The second experiment was designed to analyze the effect of driving in various driving environments on the facial temperature change and the perceived mental workload. *Driving environment* was a within-subject variable with two conditions, simulator and real vehicle. This experiment was to examine whether the effect of driving in a simulator would be different from that of driving in a real vehicle, and whether the difference may be correlated with facial temperature change.

Driving activities in virtual reality were expected to be different from those were carried out in the real vehicle. The virtual environments provide different auditory cues and haptic and kinesthetic feedback to drivers that could increase task difficulty. Drivers can also experience difficulty with navigation and spatial orientation in virtual world.

3.1. Method

3.1.1. Participants

Fourteen healthy licensed drivers participated in this experiment. However, one missing data set for a participant occurred in the car testing. Data from the simulator for that participant was excluded from further analysis. Thus, only the data sets from the 13 participants (9 males, 4 females) were used. Of the thirteen participants, 6 younger participants ranged from 18 to 35 years (mean age $= 25.7$; $\sigma = 5.2$) and 7 older participants ranged from 36 to 64 years (mean age $= 46.7$; $\sigma = 6.4$). All participants reported having normal or corrected vision and having no color blindness. All participants self-reported annual mileages to be at least between 2,000–7,999.

3.1.2. Apparatus

City-like, highway-like, and rural-like carriageways were created for the simulated driving test of this experiment. The rural was a one-lane, two-way road surrounded by classic rural-area scenarios including open, sometimes winding roads, farmhouses and trees. The other two settings and other apparatus in the simulator testing were identical to those of Experiment 1.

For the real car testing, the vehicle used was an automatic transmission Buick Rendezvous CX 2002. It was outfitted with an additional thermal imaging camera of the same model as that used in the simulator. An additional brake was installed on the passenger side of the vehicle to be used by the experimenter to override the driver controls if needed. Bright orange magnetic signs were located on the sides and rear of the vehicle that read "Caution: Test Vehicle". The participants were asked to drive from the university research park to US Highway 82 West, then merge onto US Highway 12 West in the city downtown, and then drive back to the research park. US Highway 82 is a four-lane, two-way highway with a 55 mph speed limit. Route 12 in the city downtown is the main artery of Starkville, Mississippi. It has some two-lane, but is primarily a four-lane, two-way downtown carriageway with speed limits ranging from 35 to 45 miles per hour (mph).

3.1.3. Procedure

In the simulator driving, the participants performed the experimental test after the completion of two 5-minute practice drives. The simulator driving test lasted about 25 minutes. Each carriageway took 5 minutes to complete. The MCH evaluation was administered at the end of each simulator drive. The participants performed the on-road driving test afterward. Prior to the on-road test, a break was allowed in the laboratory for the participants to recover their facial skin temperature to "baseline". After the break, they were seated and waited in the compartment of the car for approximately 5–10 min prior to the drive began. It allowed participants' bodies to acclimate to the temperature of the driving compartment. The experimenter sat in the front passenger seat of the vehicle and would tell the participants the directions for the drive. The drive lasted approximately 45 min for a participant from the time he/she left the research park until his/her return. The experimenter delivered the MCH questionnaires to the participants during the break which took place after about 20 min of driving, and again at the end of the drive. The on-road experiment was scheduled between the hours of 9:00 a.m. to 4:30 p.m. during daylight hours when it was not raining.

3.2. Results

3.2.1. Facial temperature response

The initial forehead temperature did not significantly differ from its final temperature after the real vehicle driving task in the Wilcoxon Matched Pairs Signed Rank tests ($S = 13$, $p = 0.14$). The forehead temperature demonstrated its stability in the real car driving environment as well as in the simulator. No significant difference was detected in the response of the nose temperature after the driving in the car environment. The simulator drive induced an average decrease in nose temperature of 0.47 °C that was significantly different ($S = 43.5$, $p < 0.001$) from the real car drive that induced a nose temperature increase of 0.12°C (not significantly different from the starting temperature).

3.2.2. Subjective workload score

Subjective MCH workload scores indicated a significant difference between simulator test and real-vehicle drive ($S = 33.5, p < 0.01$). The mean subjective workload scores were: simulator MCH = 4.05 ($\sigma = 1.42$) versus real vehicle MCH = 2.35 ($\sigma = 0.45$). From the definitions of the MCH questionnaire, a workload score of 4 represents "moderately high operator mental effort is required to attain adequate system performance" and a workload score of 2 represents "operator mental effort is low and desired performance is attainable".

3.2.3. Ambient temperature

The ambient temperature inside the vehicle was set at 23°C during the data collection period, and showed no significant difference between the initial and final ambient temperature during the data-collection period.

3.2.4. Correlation

The Spearman correlation test revealed a significant correlation ($r = 0.64$, $p < 0.001$) between the subjective Modified Cooper-Harper (MCH) score and the nose temperature change.

3.3. Discussion for experiment 2

Drivers perceived a lower level of mental workload when driving in a real car environment. The resulting low mental effort in the actual reality may be due to the following reasons: 1) the driving task carried out in the *usual environment* (i.e., a typical car) was not effortful nor difficult for the experienced drivers in this study, and 2) they had acquaintance with the driving vicinity and experienced relatively few unexpected episodes (e.g., startle events) during the drive.

The simulated drive, which led to higher mental effort, may be ascribed to the attributes of the Virtual Environment (VE). A study of VE usability testing revealed that users can experience difficulty with navigation and spatial orientation in virtual world compared with the interaction in real-world context [12]. In the context of automotive driving, the perception of objects in a virtual driving environment is a function of the quantity and quality of visual and auditory cues, haptic and kinesthetic feedbacks [15], and other factors. Drivers can experience difficulty in carrying out the task without the effective feedback and cues in the simulator that they expect based upon their past experiences in the real world driving. This could be a factor contributing to the higher mental workload experienced in the simulator used for this study.

The results revealed that a higher workload condition (i.e., driving in the simulated environment) provoked a greater amount of nose temperature drop, whereas, the low workload condition (i.e., driving in the real car environment) caused only a small change of nose temperature. Human facial skin temperature variations with the change of mental workload are possibly due to the change of activities of the ANS. The experimental results showed a strong trend between the mean facial temperature change and the mean perceived workload scores for both the simulator and real car tests across this sample of participants. A significant correlation of the nose temperature change with the subjective workload score was also found.

4. Discussion

This present study demonstrated the use of a non-intrusive measure of facial skin temperature change to indicate a change in mental workload as measured through the Modified Cooper-Harper test. This study also showed the same facial skin temperature measuring method can be used in a field and in a laboratory setting. The method appears to be robust in that it can indicate change across a sample of subjects after some post processing of the data. Whereas past experiments relied on the individual modulus to explain changes in workload across tasks rather than across a set of individuals tested in different conditions of the same tasks.

The participants' nose temperature dropped significantly after the driving tasks. The nose temperature drop was shown to be related to an increase of mental workload in both experiments. However, only a narrow range of mental workload was tested in these two experiments. In this range, a significant correlation was shown between nose temperature drop and the subjective workload in both experiments. The larger nose temperature drop was associated with a higher mental workload in both experiments. The participants perceived higher subjective workload for the dual-task conditions and concordantly the physiological measure showed a larger nose temperature drop for those conditions. The findings related to the forehead temperature remaining stable and constant are consistent with previous literature.

Infrared thermography was successfully employed to measure participants' facial temperature without physical contact between the instrument and human body for the assessment of mental load. The facial temperature drop due to the psychological event is believed to be due to the constriction or dilation of blood vessels which interferes with blood flow in the areas in which the activities of blood vessels are mediated by the ANS [6,8,13,14].

The infrared camera must be able to see the Region of Interest during the measurement period of interest. For the nose measurements, it appears that hair, eye glasses, and virtual reality goggles, will not interfere. A more detailed analysis is left for future work. On the other hand, hair, eye glasses and virtual reality goggles can interfere with measurements of the forehead. It is fortunate that the present findings demonstrated the stability of forehead skin temperature under different task loading conditions. This methodology may enable an evaluation of emotional changes of state through a physiological measure.

A major disadvantage of most of the physiological measures is that physical contacts of the instruments (i.e., electrodes, sensors, etc.) with body parts are often unavoidable. This is problematic since the contact between the instrument and skin surface could affect test subject and produce noise in the data. As the skin temperature can be measured by means of an infrared camera at a distance, effects of the examination itself can be minimized. Another important consideration when using the thermal imaging camera in data collection is to reduce environmental noise. The rule of thumb of the reduction of environmental noise is to 1) shorten the distance between the thermal imaging camera and the measured object, 2) eliminate high temperature objects behind the measured object, such as sun shining on the back of the measured object, 3) eliminate direct sunlight or any light source with high illuminance from striking the thermal imaging camera, and 4) remove obstacles, such as dust, between the measured object and the camera. All of these had been taken into the account during the measurement process for this study.

The limitations of this study were addressed as follows. For some scientists, mental workload measurement with multidimensional scales may have been preferred. Some scientists consider multidimensional scales to be superior to one-dimensional scales. However, it is not conclusive that a single dimensional rating approach is less useful. As reported in Verwey and Veltman [24], the usefulness of subjective scales with more than one dimension has previously been debated. Comparative evaluations of various workload assessment techniques conducted in a previous driving experiment revealed that a simple one-dimensional rating scale appeared sufficient for assessing workload in driving [24]. Furthermore, Wierwille and Eggemeier [27] compared various subjective scales and found little advantage for the multidimensional scales over the one-dimensional scales in their experiment. In this case, the single dimensional scale could be administered more efficiently.

Future studies should examine whether multidimensional scales such as National Aeronautics and Space Administration Task-Load Index (NASA-TLX) [11] and the Subjective Workload Assessment Technique (SWAT) [16] are superior for understanding the relationships between perceived mental workload with the tasks and facial temperature change. Further examination is recommended under more controlled test conditions for a "full-validation" of the psychophysiological measure used in this study. It may be useful to include additional psychophysiological measures, such as pupil dilation [22], during development and validation of multi-modal measures in future work. The hope would be that the multi-modal approach will provide a more robust, yet still non-intrusive, measure of mental workload with a greater explanation of variance than one made of one measure alone.

The other issue of the present study is that the workload levels of the tasks were in the range of approximately 3 to 5, 3 to 6, and 2 to 4, as shown in the results of the two experiments. It is recommended that future investigation for the proposed physiological measure requires the use of a workload task with multiple levels in which a wider range is provided between two levels.

In terms of implementation, some additional effort is needed to allow for the data collected within the Region of Interest to be utilized in real time. Some better tracking and monitoring of the patterns of data are needed initially. Actual or potential applications of this research include the real-time measurement of mental workload non-intrusively in a variety of military, manufacturing, transportation and healthcare environments.

5. Conclusions

Thermography, when combined with other modes of measurement such as pupil dilation, provides a highly automated and flexible means to objectively evaluate workload when comparing to other physiological measures. This can contribute to design for adaptive automation or to the further development of human performance improvements and human-system interactions. Operator's mental workload could be reduced by means of an adaptive human-system interface. By knowing when to provide assistance with an adaptive interface or automation, the excessive workload imposed on the operator would be relieved with a possible consequence of reducing likelihood of human error.

A multi-modal approach to the measure may help to clarify some apparent inconsistency in the literature with regard to expectations for psychophysiological measures as indicators of performance as well as mental and emotional strain. This can be an important contribution as Boucsein and Backs [3] suggest the greatest gap has closed with regard to empirical data [availability] but not with regard to psychophysiologists contribution to a theoretical approach backing up the measures. It is hoped that this work can contribute some additional understanding in that regard.

We typically use behavioral (video) to explain behavior and performance measures (reaction times) to explain complex behavior (drive a car). We traditionally use physiological variables to explain physical phenonmena (e.g. electromyography to explain levels of muscle tension). Gaillard and Kramer [7] suggest that though physiological measures can possibly be used for behavioral, mental or emotional (daily hassles, job insecurity), traditionally they are not. Continued development in this area is critical as the nature of work changes from an emphasis on physical to cognitive.

Acknowledgements

The authors would like to thank Christy Jinyan Du, at Caterpillar, for her assistance in part of the data analysis of this study, Prof. Srinivasan Chandrasekar for supporting the use of thermography at Purdue, and Prof. Jason McCarley at the University of Illinois Urbana-Champaign for efforts in the seminal stages of the project setup. Finally, thanks to the support staff at Drive Safety and Clayne Woodbury for their support in efforts on the initial simulator start-up.

References

[1] R.W. Backs, J.K. Lenneman and J.L. Sicard, The use of autonomic components to improve cardiovascular assessment of mental workload in flight simulation, *The International Journal of Aviation Psychology* 9(1) (1999), 33–47.

[2] R.W. Backs, J.K. Lenneman, J.M. Wetzel and P. Green, Cardiac measures of driver workload during simulated driving with and without visual occlusion, *Human Factors* 45(4) (2003), 525–538.

[3] W. Boucsein and R.W. Backs, Engineering psychophysiology as a discipline: historical and theoretical aspects, in: *Engineering Psychophysiology: Issues and Applications*, R.W. Backs and W. Boucsein, eds, Lawrence Erlbaum Associates, Mahwah, NJ, 2000, pp. 3–30.

[4] K.A. Brookhuis and D. De Waard, Assessment of drivers' workload: Performance, subjective and physiological indexes, in: *Stress, Workload and Fatigue: Theory, Research and Practice*, P.A. Hancock and P.A. Desmond, eds, Erlbaum, Mahwah, NJ, 2001, pp. 321–333.

[5] I.D. Brown and E.C. Poulton, Measuring the spare "mental capacity" of car drivers by a subsidiary task, *Ergonomics* 4(1) (1961), 35–40.

[6] P.D. Drummond, L. Camacho, N. Formentin, T.D. Heffernan, F. Williams and T.E. Zekas, The impact of verbal feedback about blushing on social discomfort and facial blood flow during embarrassing tasks, *Behaviour Research and Therapy* **41**(4) (2003), 413–425.

[7] A.W.K. Gaillard and A.F. Kramer, Theoretical and methodological issues in psychophysiological research, in: *Engineering Psychophysiology: Issues and Applications*, R.W. Backs and W. Boucsein, eds, Lawrence Erlbaum Associates, Mahwah, NJ, 2000, pp. 31–58.

[8] H. Genno, K. Ishikawa, O. Kanbara and M. Kikumoto, Using facial skin temperature to objectively evaluate sensations, *International Journal of Industrial Ergonomics* **19**(2) (1997), 161–171.

[9] T.C. Hankins and G.F. Wilson, A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight, *Aviation, Space, and Environmental Medicine* **69**(4) (1998), 360–367.

[10] N. Hayashi and H. Murase, Measurement of cognitive and emotional responses to gardening of the elderly using thermography and psychological test, Proceedings of the Automation Technology for Off-Road Equipment, 2004, pp. 201–208.

[11] S.G. Hart and L.E. Staveland, Development of NASA task load index (TLX): Results of experimental and theoretical research, in: *Human Mental Workload*, P.A. Hancock and N. Meshkati, eds, North-Holland, Amsterdam, 1988, pp. 139–183.

[12] K. Kaur, A.G. Sutcliffe and N.A.M. Maiden, Towards a better understanding of usability problems with virtual environments, Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction, INTERACT '99, 1999, pp. 527–535.

[13] K. Mizukami, N. Kobayashi, H. Iwata and T. Ishii, Telethermography in infant's emotional behavioural research, *Lancet* **11** (1987), 38–39.

[14] A. Naemura, K. Tsuda and N. Suzuki, Effects of loud noise on nasal skin temperature, *The Japanese Journal of Psychology* **64**(1) (1993), 51–54.

[15] C.K.L. Or and V.G. Duffy, A paradigm for the display- and control-intensive vehicle, in: *Contemporary Ergonomics*, P.T. McCabe, ed., CRC Press, 2004, pp. 258–262.

[16] G.B. Reid and T.E. Nygren, The subjective workload assessment technique: A scaling procedure for measuring mental workload, in: *Human Mental Workload*, P.A. Hancock and N. Meshkati, eds, North-Holland, Amsterdam, 1988, pp. 185–218.

[17] M.A. Recarte and L.M. Nunes, Mental workload while driving: Effects on visual search, discrimination, and decision making, *Journal of Experimental Psychology: Applied* **9**(2) (2003), 119–137.

[18] M.P. Snow and R.C. Williges, Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments, *Human Factors* **40**(3) (1998), 386–402.

[19] E.J. Sirevaag, A.F. Kramer, C.D. Wickens, M. Reisweber, D.L. Strayer and J.F. Grenell, Assessment of pilot performance and mental workload in rotary wing aircraft, *Ergonomics* **36**(9) (1993), 1121–1140.

[20] A. Smiley and K.A. Brookhuis, Alcohol, drugs and traffic safety, in: *Road Users and Traffic Safety*, J.A. Rothengatter and R.A. de Bruin, eds, Van Gorcum, Assen, 1987, pp. 83–105.

[21] A.M. Stoll, Techniques and uses of skin temperature measurements, Annals of The New York Academy of Sciences, *Thermography and Its Clinical Applications* **121**(1) (1964), 49–56.

[22] J.A. Veltman and A.W.K. Gaillard, Physiological workload reactions to increasing levels of task difficulty, *Ergonomics* **41**(5) (1998), 656–669.

[23] J.A. Veltman and W.K. Vos, Facial temperature as a measure of operator state, Proceedings of the HCI International 11th International Conference on Human-Computer Interaction, 2005, pp. 293–301.

[24] W.B. Verwey and H.A. Veltman, Detecting short periods of elevated workload: A comparison of nine workload assessment techniques, *Journal of Experimental Psychology* **Applied 2**(3) (1996), 270–285.

[25] B.G. Wallin, Sympathetic nerve activity, electrodermal and cardiovascular reactions in man, *Psychophysiology* **18** (1981), 470–476.

[26] W.W. Wierwille and J.G. Casali, A validated rating scale for global mental workload measurement application, Proceedings of the Human Factors Society 27th Annual Meeting, 1983, pp. 129–133

[27] W.W. Wierwille and F.T. Eggemeier, Recommendations for mental workload measurement in a test and evaluation environment, *Human Factors* **35**(2) (1993), 263–281.