Cognitive Aid: Task Assistance Based On Mental Workload Estimation

Qiushi Zhou
University of Melbourne
Melbourne, VIC, Australia
qiushi.zhou@unimelb.edu.au

Joshua Newn
University of Melbourne
Melbourne, VIC, Australia
joshua.newn@unimelb.edu.au

Namrata Srivastava
University of Melbourne
Melbourne, VIC, Australia
srivastavan@student.unimelb.edu.au

Tilman Dingler
University of Melbourne
Melbourne, VIC, Australia
tilman.dingler@unimelb.edu.au

Jorge Goncalves
University of Melbourne
Melbourne, VIC, Australia
jorge.goncalves@unimelb.edu.au

Eduardo Velloso
University of Melbourne
Melbourne, VIC, Australia
eduardo.velloso@unimelb.edu.au

ABSTRACT

In this work, we evaluate the potential of using wearable non-contact (infrared) thermal sensors through a user study (N=12) to measure mental workload. Our results indicate the possibility of mental workload estimation through the temperature changes detected using the prototype as participants perform two task variants with increasing difficulty levels. While the sensor accuracy and the design of the prototype can be further improved, the prototype showed the potential of building AR-based systems with cognitive aid technology for ubiquitous task assistance from the changes in mental workload

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313010

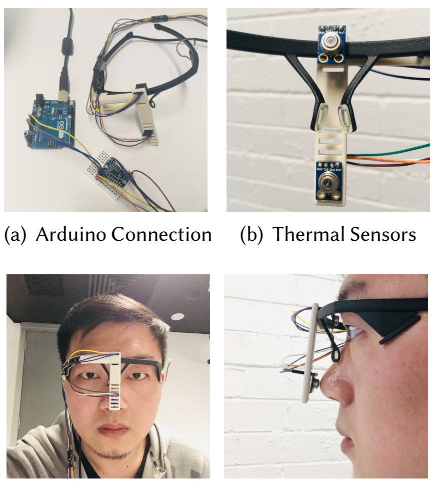
demands. As such, we demonstrate our next steps by integrating our prototype into an existing AR headset (i.e. Microsoft HoloLens).

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Cognitive science**; *Mixed / augmented reality*; • **Hardware → Sensor applications and deployments**;

KEYWORDS

Cognitive load, thermal sensor, affective computing



(a) Arduino Connection (b) Thermal Sensors

(c) Front View (d) Side View

Figure 1: Prototype wearable non-contact thermal sensors with 3d-printed frame.

INTRODUCTION

Advances in psycho-physiological sensing technology have substantially contributed to the development of research revolving around understanding and augmenting human cognition [5, 6, 13]. Researchers in HCI and psychology has used a wide range of technology including electroencephalogram (EEG), eye tracking and thermal imaging as a potential method to measure psychological attributes such as emotion, stress and mental workload [2, 9, 18].

The use of Cognitive Load Theory (CLT) to interpret sensor data has shown to have potential by interpreting estimated overall mental workload as different types of cognitive load [16, 17], few were able to validate the data with direct measuring methods of mental workload. Despite the rapid miniaturisation of sensors in recent years, there is still a lack of ideal implementation due to its obtrusive nature and poor mobility. In this work, we estimated mental workload with small non-contact **infrared thermal sensors** and conducted a user study to **validate the sensor data with CLT**. The application of this technology enables ubiquitous mental workload sensing and provides opportunities for automatic task assistance based on observed mental workload and its reference within the task at hand. The small footprint of the hardware also allows it to be easily embedded in most wearable devices. We envision an Augmented Reality (AR)-based system (e.g. Microsoft HoloLens) of today but with integrated thermal sensors for the purposes of **workload-aware automatic task assistance**.

RELATED WORK

Cognitive Load Theory has been an ongoing topic of investigation for application such as knowledge transfer in learning [4, 12, 16]. For instance, Brunken et al. categorized physiological measurement and dual-task performance measurement of cognitive load as indirect and direct respectively due to their direct causal relationship with cognitive load [4]. The authors, through a user study, have shown that a secondary reaction task can be used for estimating the cognitive load induced by the primary reading comprehension task.

Figure 2: Tasks. *Top*: Stroop Test, *Bottom*: Reading Comprehension with Secondary Task.

Different approaches are used by researchers to measure mental workload. Over the years, a strong relationship has been demonstrated between facial temperature change and mental workload. For instance, Zajonc et al.'s work on emotion and facial efference revealed the link between facial temperature dynamics and stress [18]. In another study, Kataoka et al. found a high correlation between stressful task performance and facial temperature. Or and Duffy utilized facial thermal imaging in driving simulation and claimed nose temperature as a reliable indicator of mental workload [11]. Stemberger et al. analyzed the effect of cognitive workload on facial temperature with a neural network classifier [14]. Jenkins and Brown conducted a study to indicate that forehead temperature change can reflect variation in cognitive demand [10]. Abdelrahman et al. demonstrated that a significant effect can be found between cognitive load and task difficulty using a system that implemented thermal imaging with face recognition [1].

Estimation of mental workload using sensor technology can help us build Automatic Task Assistance systems. For example, Bonanni et al. designed intelligent kitchen systems to help reduce cognitive load in daily life using attention-tracking [3]. Similarly, Gerry et al. designed a virtual reality system that assists in a search task with EEG mental workload sensing [8].

USER STUDY

We conducted a user study to investigate the feasibility of building a wearable sensor implementation for estimating mental workload. Our study is driven by three hypotheses based on prior work:

(1) With increasing task difficulty, a subsequent increase can be observed in forehead temperature and a substantial decrease in nose temperature.
(2) When participants perform the tasks, there will be a greater change in temperature as compared to the rest periods.
(3) Performance on a secondary task will be affected as the difficulty level of primary tasks increase.

To test the hypotheses, we employed a repeated-measures design where all participants were subjected to four difficulty levels of the Stroop test [15] and a reading comprehension task containing an additional secondary reaction test. The order of the tasks was counterbalanced using a Latin Square.

Tasks: The **Stroop test** required participants to quickly react to a word-color combination to decide if they match (Figure 2-Top). The difficulty levels of this test correlated with the amount of time allowed to react to each stimulus i.e. the more difficult, the faster the participant has to perform the task. On the other hand, the **reading task** required participants to read and understand each material within three minutes followed by a comprehension question (Figure 2-Bottom). The difficulty levels of the materials were based on their topics, sources and Flesch Reading Ease scores [7]. We used four different materials– a travel blog, a financial news article from an entry-level textbook, a GMAT 500 scientific article and a GMAT 700 article about media theory– with a Flesch Reading Ease scores of 68.6, 54.1, 29.1 and 4.0 respectively. We used the performance of a secondary reaction task added to

Figure 3: Experimental Setup.

the reading comprehension task to directly measure mental workload. During reading, participants reacted to the randomly-appearing circle as quickly as possible by pressing a button to make it disappear (Figure 2-Bottom). The performance of the reaction task was recorded as reaction time.

Experimental Setup: Our setup consisted of a desktop computer with two thermal sensors connected by an Arduino board (Figure 3). The sensors are Melexis MLX90614 Single Zone Infra Red Thermometers with 5 Hz frequency, -40 to +85°C Operating Temperature Range, ± 0.5°C accuracy and 90°FOV. The sensors were attached to a wearable frame similar to everyday eyeglasses together with a 3D-printed component designed as adjustable to fit different facial features (Figure 1).

Participants and Procedure: We recruited 12 participants (7 females) with an average age of 26.3 years through the University's mailing list. All participants were required to have native English reading ability. We informed the participants of the purpose of the study upon their arrival. After a brief training with the Stroop test, we asked the participants to perform each level for 3 minutes, each followed by a rest period of 3 minutes. Then, they were asked to perform each level of the reading task (with the secondary reaction task) in the same fashion (3 minutes task, 3 minutes rest). The study took approximately 60 minutes on average to complete. We recorded the temperature with the sensors pointing at participants' forehead and nose (as shown in Figure 1(c) and 1(d)) for both tasks as participants performed them including the rest periods in between. The experiment was conducted in a maintained room temperature of 24°C. Participants were compensated with coffee vouchers for their time.

Results: We analyzed the effect of task difficulty on forehead and nose temperature as well as the difference between forehead and nose temperature. We examined two aspects of the change in temperature: change-over-time and the velocity-of-change. We define change-over-time as the mean temperature during the last minute of a 3-minute task or a 3-minute break minus the mean temperature during the first five seconds of the same period. We define the velocity-of-change as the velocity of temperature change during any 30-second window (during which the temperature changed the most) in a 3-minute task or a 3-minute break.

We tested the effect of the objective and subjective difficulty levels of the tasks on temperature change with one-way ANOVA tests, but found no significant results. We also tested the effect of the difficulty levels of reading task on the average reaction time of the secondary task with a one-way ANOVA without significant results.

We compared the temperature change during the 3-minute task periods and their preceding 3-minute break periods. We tested the effect of task on/off on temperature change for each difficulty level with one-way ANOVA tests. We found significant effect of between task and rest periods on change-over-time of forehead-nose temperature difference for 3 conditions: level-3 Stroop test ($F(1, 11) = 13.09, p < .01, ges = 0.27$) (Figure 4-Top), level-4 Stroop test ($F(1, 11) = 11.32, p < .01, ges = 0.43$) (Figure 4-Bottom) and level-4 reading task ($F(1, 11) = 4.97, p < .05, ges = 0.29$) (Figure 5).

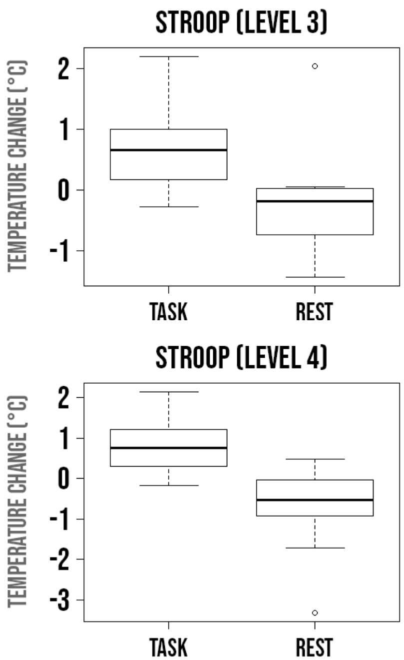


Figure 4: Stroop Test Results.

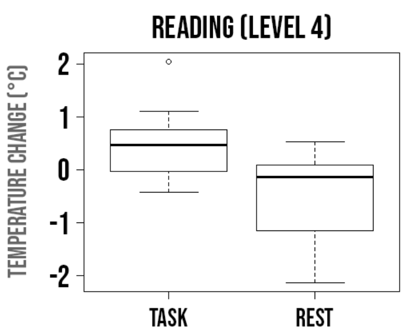


Figure 5: Reading Task Result.

Discussion: Though the results did not show a significant effect of task difficulty on facial temperature change, the effect of task on/off on temperature change did reflect our second hypothesis. In other words, though our sensor did not provide enough granularity to differentiate between levels of difficulty within a task, it could differentiate between doing the task and not doing the task. Among all levels of both tasks, only the two most difficult levels of Stroop tests and the one most difficult level of reading task showed significant increase between the forehead-nose temperatures.

We observed individual differences from the results. While the recorded facial temperature of some participants demonstrated a clear increase-decrease pattern during task-on and task-off periods, the temperature change pattern of other participants was not as clear. We summarize possible reasons for such individual difference as the different facial structure of different participants resulted in different position and distance of the sensor from their skin; different participants may have a different perception of task difficulty; different participants may have a different threshold of facial temperature change. While we attempted to measure facial temperature with small foot-print thermal sensors, the lack of accuracy of the sensors utilised in this study compared to more professional thermal cameras is another possible reason for the lack of significant results.

CONCLUSION

While our preliminary results did not show a significant effect of task difficulty on facial temperature change, we did find a significant effect of task performance on facial temperature change in the most difficult tasks. These results indicate the possibility of estimating mental workload with non-contact infrared thermal sensors embedded on a wearable device while requiring improvement in sensor accuracy and individual fitting. We will work on improving the design of the wearable supporting frame for the sensors to accommodate different facial features better and test the effect of task difficulty on larger sample size. We plan to explore the possibility of utilizing more advanced thermal sensing technology while restricting the form-factor of the device to enable more accurate and task-difficulty-sensitive mental workload estimation.

FUTURE WORK

Our work paves the way for a potential AR-based automatic task assistance system with Cognitive Aid technology. We have begun to design an attachment for the Microsoft HoloLens to embed the thermal sensors for mental workload sensing (see Figure 6). With the ability to render virtual objects on surrounding surfaces using wearable AR technology, we will explore the use of mental-workload-aware automatic task assistance system. To do this, we will incorporate the environment of a user, whether working or learning with mental workload sensing to provide real-time feedback to reduce extraneous cognitive load for effective learning and to aid users in everyday tasks.

Figure 6: Thermal sensors embedded on HoloLens with 3d-printed adjustable components for task assistance.

ACKNOWLEDGMENTS

Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Government.

REFERENCES

[1] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 33.

[2] Jackson Beatty and Brennis Lucero-Wagoner. 2000. The pupillary system. *Handbook of psychophysiology* 2, 142-162 (2000).

[3] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. Attention-based Design of Augmented Reality Interfaces. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1228–1231. https://doi.org/10.1145/1056808.1056883

[4] Roland Brunken, Jan L Plass, and Detlev Leutner. 2003. Direct measurement of cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 53–61.

[5] Douglas C Engelbart. 2001. Augmenting human intellect: a conceptual framework (1962). *PACKER, Randall and JORDAN, Ken. Multimedia. From Wagner to Virtual Reality. New York: WW Norton & Company* (2001), 64–90.

[6] Stephen H Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers* 21, 1-2 (2009), 133–145.

[7] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[8] Lynda Gerry, Barrett Ens, Adam Drogemuller, Bruce Thomas, and Mark Billinghurst. [n. d.]. Levity: A Virtual Reality System that Responds to Cognitive Load. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW610.

[9] Martijn Haak, Steven Bos, Sacha Panic, and LJM Rothkrantz. 2009. Detecting stress using eye blinks and brain activity from EEG signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)* (2009), 35–60.

[10] SD Jenkins and RDH Brown. 2014. A correlational analysis of human cognitive activity using Infrared Thermography of the supraorbital region, frontal EEG and self-report of core affective state. QIRT.

[11] Calvin KL Or and Vincent G Duffy. 2007. Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics* 7, 2 (2007), 83–94.

[12] Fred G Paas. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology* 84, 4 (1992), 429.

[13] Rosalind Wright Picard et al. 1995. Affective computing. (1995).

[14] John Stemberger, Robert S Allison, and Thomas Schnell. 2010. Thermal imaging as a way to classify cognitive workload. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*. IEEE, 231–238.

[15] J Ridley Stroop. 1935. Studies of interference in serial verbal reactions. *Journal of experimental psychology* 18, 6 (1935), 643.

[16] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.

[17] John Sweller, Jeroen JG Van Merrienboer, and Fred GWC Paas. 1998. Cognitive architecture and instructional design. *Educational psychology review* 10, 3 (1998), 251–296.

[18] Robert B Zajonc, Sheila T Murphy, and Marita Inglehart. 1989. Feeling and facial efference: implications of the vascular theory of emotion. *Psychological review* 96, 3 (1989), 395.