Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load

33

YOMNA ABDELRAHMAN, University of Stuttgart, Germany
EDUARDO VELLOSO, The University of Melbourne, Australia
TILMAN DINGLER, Osaka Prefecture University, Japan
ALBRECHT SCHMIDT, University of Stuttgart, Germany
FRANK VETERE, The University of Melbourne, Australia

Current digital systems are largely blind to users' cognitive states. Systems that adapt to users' states show great potential for augmenting cognition and for creating novel user experiences. However, most approaches for sensing cognitive states, and cognitive load specifically, involve obtrusive technologies, such as physiological sensors attached to users' bodies. This paper present an unobtrusive indicator of the users' cognitive load based on thermal imaging that is applicable in real-world. We use a commercial thermal camera to monitor a person's forehead and nose temperature changes to estimate their cognitive load. To assess the effect of different levels of cognitive load on facial temperature we conducted a user study with 12 participants. The study showed that different levels of the Stroop test and the complexity of reading texts affect facial temperature patterns, thereby giving a measure of cognitive load. To validate the feasibility for real-time assessments of cognitive load, we conducted a second study with 24 participants, we analyzed the temporal latency of temperature changes. Our system detected temperature changes with an average latency of 0.7 seconds after users were exposed to a stimulus, outperforming latency in related work that used other thermal imaging techniques. We provide empirical evidence showing how to unobtrusively detect changes in cognitive load in real-time. Our exploration of exposing users to different content types gives rise to thermal-based activity tracking, which facilitates new applications in the field of cognition-aware computing.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Cognitive science**; • **Hardware → Displays and imagers**;

Additional Key Words and Phrases: Thermal Imaging, cognitive load, Thermal latency

ACM Reference Format:
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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 33 (September 2017), 20 pages.
https://doi.org/10.1145/3130898

This work is supported by the Victorian State Government and Microsoft through their contributions to the Microsoft Research Centre for Social Natural User Interfaces (SocialNUI). It was partly conducted within the Amplify project which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 683008) and received funding from the German Research Foundation within the SimTech Cluster of Excellence (EXC 310/2).
Author's addresses: Y. Abdelrahman, A. Schmidt, Institute for Visualization and Interactive Systems, University of Stuttgart, email: {yomna.abdelrahman, albrecht.schmidt}@vis.uni-stuttgart.de. E. Velloso, F. Vetere, School of Computing and Information Systems, The University of Melbourne, email: {eduardo.velloso, F.vetere}@unimelb.edu.au. T. Dingler, Graduate School of Engineering, Osaka Prefecture University, email: tilman@cs.osakafu-u.ac.jp.

2474-9567/2017/9-ART33
https://doi.org/10.1145/3130898

1 INTRODUCTION

Building systems that extend our cognitive abilities, augment our intellect [10], work in symbiosis with humans [35], and provide ubiquitous access to information [66] has been a core theme in human-centered computing since its inception. These aspirations have carried on through multiple research programs, including Affective Computing [47], Physiological Computing [12], and more recently Symbiotic Interaction [25] and Human Amplification [51]. These cognition-aware systems aim to sense users' internal states and to adapt their interface and behaviour accordingly. Such systems offer opportunities to tailor educational activities in online learning environments, to dynamically optimize work-flows for knowledge, to improve performance for assembly line workers [14], and to focus users' attention in critical systems. A crucial step in building cognition-aware systems is capturing different aspects of users' mental states, such as their cognitive load, loci of attention, and affect. Despite over 50 years of work in the area, how to sense cognitive load in a robust, accurate, timely, and unobtrusive way is still an open challenge.

Cognitive load has been measured traditionally in two ways: (1) by subjective self-reporting and (2) by observing user performance in a task or in a set of parallel tasks. The NASA TLX is a common example of the first category, where participants are asked to report their own load with regard to 6 different categories. Another example where study participants are asked to report their own estimates can be found in Sweller et al. [62]. The drawback of these approaches is that the answers are highly subjective. Furthermore, the self-reporting itself adds to the cognitive load. Measuring cognitive load through the performance in the task itself or in a secondary task (e.g. Lane Change Task for Automotive user interface, ISO 26022) only provides a rough estimate and is typically only suitable to laboratory studies and not for creating cognition-aware real-time systems. For interactive systems to be able to adapt their behavior accordingly, cognitive load information must be captured continuously and automatically—introspection is often not sufficient. Physiological sensors, such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and galvanic skin response (GSR) sensors show potential as possible solutions to this problem, but are limited in their application in ubiquitous computing environments since they require users to wear obtrusive additional hardware (e.g. electrodes on their skin).

By adding sensing capabilities to the environment rather than burdening the user, thermal imaging is a strong candidate for the task of measuring cognitive load. Thermal cameras are both unobtrusive and able to capture information from multiple users at a distance and at the same time. Previous research has shown that thermal patterns reveal different aspects of our internal states, including affect [24, 60], stress [48], and deception [49]. Further, advances in miniaturization and mass production have continuously brought down the prices of these devices. With consumer-grade cameras readily available in the market for a few hundred dollars, measuring cognitive load at a larger scale becomes feasible.

In this article we present a novel method for estimating cognitive load based on users' facial temperature patterns using a commercial thermal camera. Our method works with off-the-shelf hardware and is applicable for ubiquitous computing environments. It automatically estimates in near real-time the cognitive load of the user and opens up new opportunities for large scale deployments of cognition-aware technologies.

In this paper, we advance the state-of-the-art of automatic cognitive load estimation through the following contributions:

- We propose a method for estimating four-levels of cognitive load by computing the temperature difference between the user's forehead and nose temperatures using computer vision techniques on the feed of a thermal camera.
- We demonstrate the validity of our metric through a user study, showing that the estimate for cognitive load strongly correlates with the task difficulty both in an artificial controlled task (Stroop tests [61] under different levels of time pressure) and in a naturalistic task (reading text fragments of different complexity).

- We demonstrate (in a second study), that our method can estimate changes in cognitive load in near real-time, with a latency of 0.7s making it suitable for creating cognitive-aware interactive systems.
- We release both our system implementation and our dataset as open source for future researchers to build upon, replicate, and extend [1].

2 RELATED WORK

Awareness of cognitive load and processes is an ongoing challenge for HCI research. Our work builds on two strands of prior work: (1) cognitive load estimation and (2) thermal imaging.

2.1 Cognitive Load Inference

A technology with the capability of sensing and inferring cognitive load has the ability to provide a "window into our mind" that can further be used to adapt system behavior accordingly [63]. However, capturing cognitive state information is a challenging task—cognitive processes are largely invisible from the outside of the users' brain and introspection often fails to reason about them in a unbiased and objective way.

Monitoring users with the help of sensors can give us clues about different cognitive states. Though certain physiological sensors can be highly specialized, expensive, and therefore only applicable under lab conditions, advances in sensor technology have led to inexpensive solutions that can be easily integrated into personal devices.

Different approaches have been introduced to infer cognitive states ranging from using facial expressions [65], eye movements [18] and pupil size [6, 46], skin conductance [20, 32, 56], brain signals [8, 57, 58], Electrodermal activity (EDA) [17, 41, 42], heart and respiration rate [32].

EEG: Haak et al. [18] reported that the blinking rate is directly proportional with cognitive load. They extracted the blinking rate from Electroencephalography (EEG) signals by isolating events from the signals. Hosni et al. [21] used EEG signals for task classification. Shirazi et al. [58], classified reading and relaxing tasks based on EEG signals retrieved from a single electrode BCI. Petersen et al. [45] used the EPOC and to distinguish emotional responses when viewing different content.

Galvanic Skin Response (GSR): Analysis of GSR data from user experiments has shown that GSR across users increases along with cognitive load [9]. Yu et al. explored the applicability of using GSR as an indicator of cognitive load [56]. Elise et al. used heart rate, respiration rate and GSR as an indicator [32].

Respiration Rate: Parnandi et al.considered real-time adaptive biofeedback games [41, 42] . They monitored players' EDA to infer their arousal states [41]. Additionally, they used biofeedback sensors (respiration rate sensor and adaptive games) to manipulate their behavior [42].

Combined Sensors: Wang et al. [9, 65] explored how to build an adaptive system that helps workers who use computer heavily on a daily basis by extracting user's features, such as face pose, eye blinking, yawn frequency and eye gaze from a recorded video, in order to monitor the users state. Healey et al. [19] and Schneegass et al. [52] used physiological monitoring for driver stress indication.

A major limitation with both voice and facial-based approaches is that users can be quite skilled at manipulating the parameters being sensed by the system. On the other hand, physiological metrics, such as heart rate, GSR, Blood Volume Pressure (BVP), and Electromyography (EMG) have the advantage that they are primarily under the control of the Autonomic Nervous Systems (ANS) and are therefore less susceptible to conscious manipulation. However, a major limitation of current physiological approaches is the need for sensors to be in direct contact with the user, or to be implanted. As a result, such sensors are impractical for most routine user environments. For instance, the long setup time and contact requirements of BCIs, or the drift over time [34] and fluctuations due to arm movements in GSR.

[1] https://github.com/Yomna-Abdelrahman/Cognitive_Thermal.git

As promising tools within the HCI domain, thermal cameras show high potential for estimating cognitive load. It overcomes the limitation of using contact sensors utilized in previous research and is more robust than other contactless approaches, since the temperature signature is more resistant to conscious manipulation [24]. In this paper, we investigate the use of thermal imaging to estimate cognitive load, exploiting the fact that cognitive load influences the skin temperature (which is directly related to the conduction of heat from the blood to the facial skin [43]) as a reflection of the activation of the ANS [11]. We therefore aim to leverage the correlation between cardiovascular physiology and mental state, where they are capable of reliably differentiating between levels of cognitive load [59].

2.2 Thermal Imaging

Thermal imaging operates in the infrared band in the electromagnetic spectrum, i.e. it senses light waves invisible to the human eye. The infrared spectrum is divided into three sub-bands: (1) Near Infrared (NIR), (2) Mid Infrared (MIR) and (3) Far Infrared (FIR). All three bands are used to passively capture a heat map (i.e. temperature profile) of the camera's field of view. However, the three bands are used for monitoring different temperature ranges, and hence require different operating imaging technology.

NIR imaging operates between the 0.7 to 2.5μm wavelengths range and can monitor temperature ranging from 600°C to 1000°C. NIR is typically used in the industrial space, given its temperature and price range. MIR imaging operates in wavelengths from 1.3μm to 8μm and observes temperature ranges from 5°C to 300°C. Thermal cameras operating in the FIR spectrum with wavelengths between 7.5 and 13μm can capture temperature ranges between -20°C and 900°C. FIR thermal cameras are considered to be the most commercially available thermal cameras in terms of size and cost [2].

Due to the operation nature and price history of thermal imaging, previous applications were limited to specific domains such as medical, firefighting, and industrial settings. However, the evolution of commercial, relatively cheap thermal cameras operating in the FIR band allowed the evolution of a diverse new set of applications. In a recent survey paper, Gade and Moeslund provided an overview of current applications of thermal cameras, highlighting the potential of using thermal cameras in the HCI domain [15] for instance, for physiological motoring [24], vision extension [3, 4], gestural interaction [1, 5, 33, 50] and touch points detection [2].

Thermal cameras can provide information about the observed body's temperature, which can be used to infer the physiological [24] and cognitive state of users in an unobtrusive manner by, for example, evaluating their stress levels [48]. The reason why this is possible is because our skin temperature is modulated by ANS activity. ANS controls the organs of our body, such as the heart, stomach, and intestines. It is responsible for activating glands and organs for defending the body from threats. Its activation might be accompanied by many bodily reactions, such as an increase in heart rate, rapid blood flow to the muscles, activation of sweat glands, and increase in the respiration rate. These physiological changes can be measured objectively by using sensors [31, 53].

Temperature changes on the forehead have been shown to be linked to changes in brain temperature [16, 37]. There is a direct relationship between workload and facial temperature based on the involvement of the autonomic nervous system (ANS) [22]: increased brain activity causes a surge in blood supply. Hence, higher workloads lead to blood flowing from the adjacent facial areas to the brain causing the facial temperature to vary. Zajonc et al. [67] showed different facial areas to be effective temperature indicators, namely the tip of the nose, above the eyes, and at the center of the forehead.

Previous works have explored the usage of thermal imaging to observe users' mental states. We particularly build upon previous work that assessed stress based on the variations in forehead and nose temperature [11, 44, 55]. Emotions like stress [48], fear [38], startling [54], empathy [36], anxiety [68], and guilt [23] could be detected by monitoring facial temperature changes. Ioannou et al. summarized these states and how they correlate to the

[2] http://www.flir.com

facial temperature in terms of region of interest and direction of temperature change (i.e. increase or decrease in the temperature) [23].

Compared to previously established sensors, the great advantage of thermal imaging is its contactless and non-invasive operation. The contact-free recording of facial temperature with an thermal camera allows us to isolate unsystematic data variation (e.g., users' bias due to their awareness of being monitored, the movement of the sensor or the stressful attachment of the sensor on the users' body). Additionally, instrumenting the environment is more user-friendly and allows the tracking of multiple users. Most of the research so far has used MIR thermal cameras. For instance, *StressCam* [48] used the Indigo Phoenix thermal camera costing over 20, 000 USD. Jenkins and Brown [26], utilized the supraorbital region to identify cognitive state, yet they used a non-commercial thermal camera. However, because FIR thermal cameras are commercially available and relatively affordable, they present a compelling opportunity for expanding the reach of these applications. For instance, it is now possible to buy smartphones with built in thermal cameras [3]. These cameras are becoming increasingly smaller, with sizes as small as a 20mm [4], yet maintaining high thermal sensitivity around 0.05° degrees. This enables thermal cameras to be used in a diverse set of applications, by enhancing existing application scenarios and exploring new ones. Previous research has shown that thermal imaging in the NIR and MIR bands can be used to reveal different cognitive states. Stemberger et al. [60] explored the use of FIR to estimate cognitive load levels, but they used a wearable tracking headset to identify the region of interest, and a neural network to build a user dependent classification to three-levels of workload based on six region of interest. Or and Duffy [40], and Kang et al. [27, 28], used the variation in the nose temperature as an indicator of cognitive workload. However, they didn't report on how different levels of workload influence the temperature change. Additionally, their findings are confounded by facial temperature stress indicators [29].

In this paper, we explore thermal imaging operating in the FIR band. We leverage the advances in miniaturization and reduction in the prices of these devices, to explore the feasibility to not only detect the cognitive load, but also to estimate four-levels of cognitive load, while maintaining the unobtrusive operation manner of thermal imaging. We aim to investigate the possibility of using thermal imaging as a user-independent cognitive state detector. Additionally, we also explore different metrics to avoid any possible overlap between other states e.g. stress.

In summary, the aim is to address two major shortcomings of previous work concerned with estimating cognitive load—the obtrusive and contact nature of traditional physiological sensors and the limitation in detecting different cognitive load levels in a user-independent manner.

3 THERMAL IMAGING FOR COGNITIVE LOAD ESTIMATION

Our cognitive and affective states strongly influence how our blood flows through our bodies. When we are scared, blood flows to our legs in reaction to the *fight or flight* response; when we are embarrassed, blood flows to our face, making us blush. Because blood carries heat, as it flows through our bodies it changes the temperature distribution in our skin [59], underlying tissues, and vessels [54, 59]. Therefore, monitoring changes in this distribution can give us an insight into the changes in cognitive load or arousal that caused them.

Previous works have suggested several points in the body to measure this temperature fluctuation, such as the nose, the cheeks, the areas around the eyes (periorbital and supraorbital), the jaw, the next, the hands (fingers and palms), the lips, and the mouth [24]. The face is particularly promising for this task for several reasons. First, it is often exposed, making it easy to observe with a thermal camera. Second, it features a thin layer of tissue, making temperature changes more pronounced. Therefore, in this work we explore how facial temperature fluctuations can give us an insight into changes in cognitive load. We focus on two of the points of interest suggested in the

[3]http://www.catphones.com/en-gb/phones/s60-smartphone
[4]http://www.flir.com/cores/lepton/

literature—the forehead and the nose—as these can be monitored even if the user is wearing glasses. In summary, in this work we focus on two research questions:

(1) Can we distinguish different cognitive load levels using a relatively low-cost, commercially available thermal camera? More specifically, do the changes in facial temperature correlate with the level of difficulty of the task? (**RQ1**)
(2) If so, how long after the increase in cognitive load can we detect the corresponding temperature fluctuation (**RQ2**)? In other words, what is the latency of our method as a cognitive load sensor?

To answer the first research question, we built a system that monitors users facial temperature (see Section 4.3) and observed how it changed as users performed two tasks with four levels of difficulty each (see Section 4.1). We hypothesize that the higher task difficulty will result in a greater temperature difference. To address the second research question we conducted a second study and measured how long the temperature started to change after the task started or ended, and how long it took for it to reach its maximum level (see Section 5). We present our results regarding the applicability of thermal cameras as a cognitive load sensor and suggest directions for future work (see Section 6).

4 STUDY I: CORRELATING COGNITIVE LOAD WITH FACIAL TEMPERATURE

To answer **RQ1** and to test our hypothesis of the ability of thermal cameras to elicit cognitive states and classify tasks based on face temperature variation, we conducted a user study in which we recorded the temperature of participants' nose and forehead in three activity states:

(1) Relaxing as the baseline.
(2) Reading four different types of text.
(3) The Stroop test [61] with four levels of difficulty.

4.1 Design

We applied a repeated-measures design, where all participants were exposed to all three task conditions. We studied the effect of the tasks on the facial thermal print. For the baseline we asked the participants to relax. For the reading task we provided four types of different content types: 1) a comic, 2) an easy blog article, 3) a scientific article and 4) a literary piece. We chose these content types because of their presumed differences in cognitive demand. Additionally, we computed the readability index [5] for each text, which indicates the text difficulty: the higher the value the more difficult the text is to read. The text found in the comic, easy blog, science article and literary piece reported 26.6, 52.9, 68.2 and 77.9 respectively. All content used during the user study was in German and all participants were native German speakers.

The Stroop test is a classic Psychology task for evaluating executive functions [61]. During the test, users are asked to name the color of the font in which different words are written. The difficulty of the task lies in the fact that the words displayed represent a different color to the one in which it is colored. For example, the word 'red' would appear colored in blue, and the participant had to say 'blue'. In our study we also introduced four levels of difficulty in the task by adding four levels of increasing time pressure: the higher the level, the less time users had to respond. For varying the difficulty of the Stroop test, we considered four levels of difficulty provided by the app *Magic Colors*[6]. To overcome the effect of the repeated-measures experimental design, namely order effect, the order of the tasks was counter-balanced using a Latin Square.

[5]https://www.psychometrica.de/lix.html
[6]https://play.google.com/store/apps/details?id=com.accountmaster.in.MagicColors

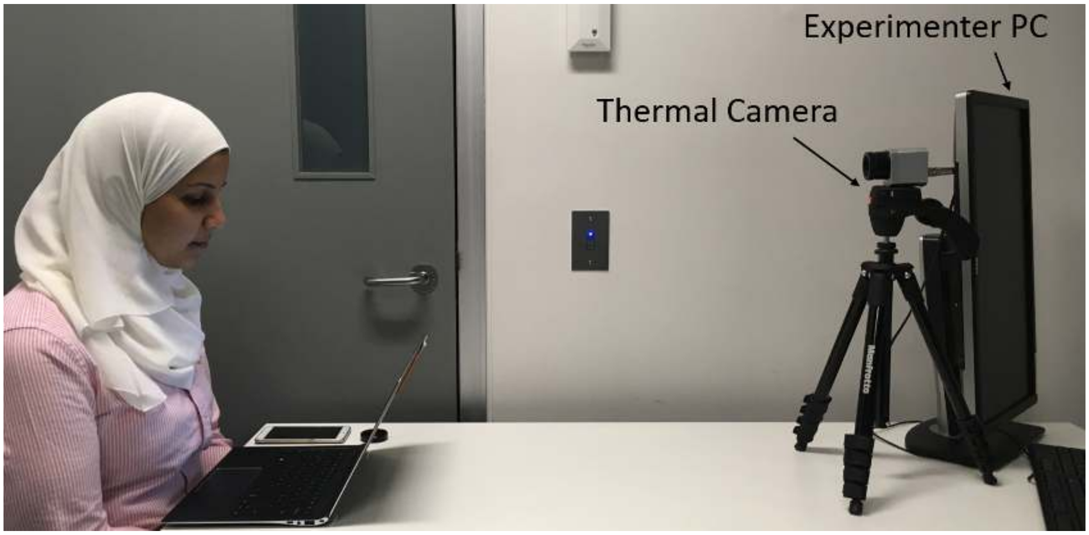

Fig. 1. Study setup consisting of thermal camera facing the participant while performing the reading task.

4.2 Apparatus

Our experimental setup consisted of a 13.3" laptop in front of a thermal camera (Optris PI160[7]) mounted on a tripod. The optical resolution of our camera was 160×120 pixels and its frame rate was 120 Hz. It is able to measure temperatures between -20°C and 900°C, and operates with a thermal sensitivity of 0.08°C represented by the noise equivalent temperature difference (NETD)[8]. The wavelengths captured by the camera are in the spectral range between 7.5μm and 13μm. The lens we use provides a 23°× 17° field of view. The thermal camera uses USB as power source as well as to transfer data. It provides temperature information in the form of 16-bit color values encoding the temperature information. The participants were asked to look to the front facing the thermal camera placed at 1m from the participants and the screen as shown in figure 1.

4.3 Implementation

To answer the first research question, we built a system consisting of a thermal camera and an image processing software that recognizes and analyzes the user's facial temperature. Our application receives the data from the thermal camera, recognizes the face of the user, and extracts the temperature of the forehead and nose. We used the OpenCV library[9] for image processing and facial points extraction.

To enhance the face recognition, we performed a series of preprocessing steps for each retrieved frame:

(1) **Frame extraction:** We sample each frame of the camera feed at 120fps, based on the cameras frequency.
(2) **Noise filtering:** We apply a 5×5 median filter to smooth the image. We convert the output to gray-scale and apply a 2D Gaussian filter to further remove high frequency noise as performed by Shirazi et al. [50].
(3) **Face Recognition:** We detect faces in the frame using the Viola-Jones classifier [64] built into OpenCV.

[7]http://www.optris.com/thermal-imager-pi160
[8]NETD refers to the electronic noise that is interpreted as a temperature difference of an object
[9]OpenCV: http://opencv.org/ (last access: July 9, 2019)

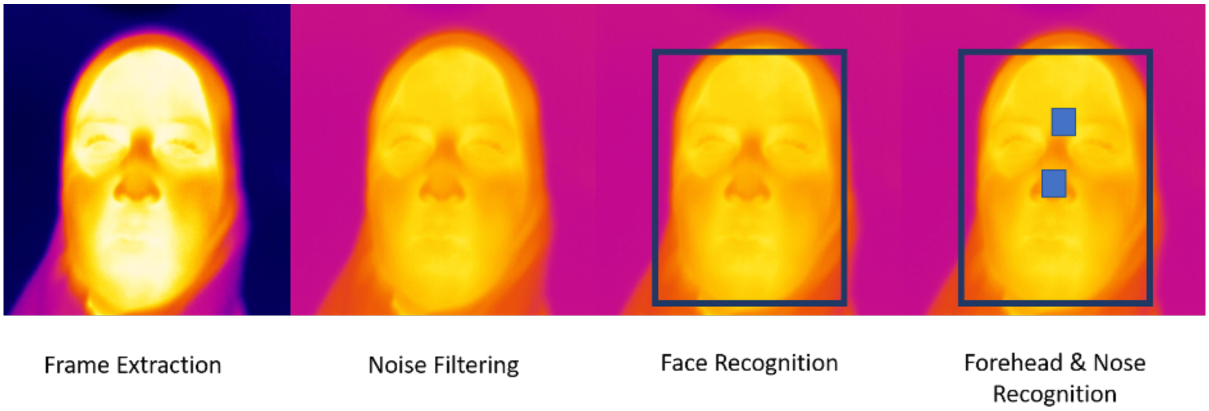

Fig. 2. Nose and Forehead ROI extraction.

(4) **ROI Identification:** We identify the nose tip and forehead as the ROI (Regions of Interest). These ROI are computed relative to the face coordinates extracted as shown below. We identified a 5x5 pixels window to represent the ROI. We used a simple ROI identification approach to maintain a fast frame rate for the algorithm. Unlike Stemberger et al. [60], we aimed to rely on the thermal camera solely without any wearable tracking headsets.

```
xForehead = xFace + (4 * face.Width / 7);
yForehead = yFace + (face.Height / 6);

xNasal = xFace + (4 * face.Width / 9);
yNasal = yFace + (face.Height / 2);
```

(5) **Temperature Recording:** We record the average temperature of the 5x5 window CSV file to represent the temperature of the nasal tip and forehead, as well as the difference in temperature between the two.

4.4 Participants and Procedure

We recruited 12 participants (7 females) with an average age of 28.3 years ($SD = 4.6$) using university mailing lists. None of the participants had any previous experience with thermal cameras. After arriving in the lab, participants signed a consent form and received an explanation of the purpose of the study. Next, we asked participants to perform the set of reading tasks, video watching and Stroop tests, each for 12 minutes (3 mins × 4 levels). The order of the tasks was counter-balanced using Latin-square.

The study took approximately 60 minutes. During the entire experiment, we recorded the temperature of the participant's face, extracting the forehead and nasal temperatures. Additionally, we recorded the whole study using an RGB video camera. The experiment was conducted in a maintained room temperature of 24°C. Participants were rewarded with 10 EUR.

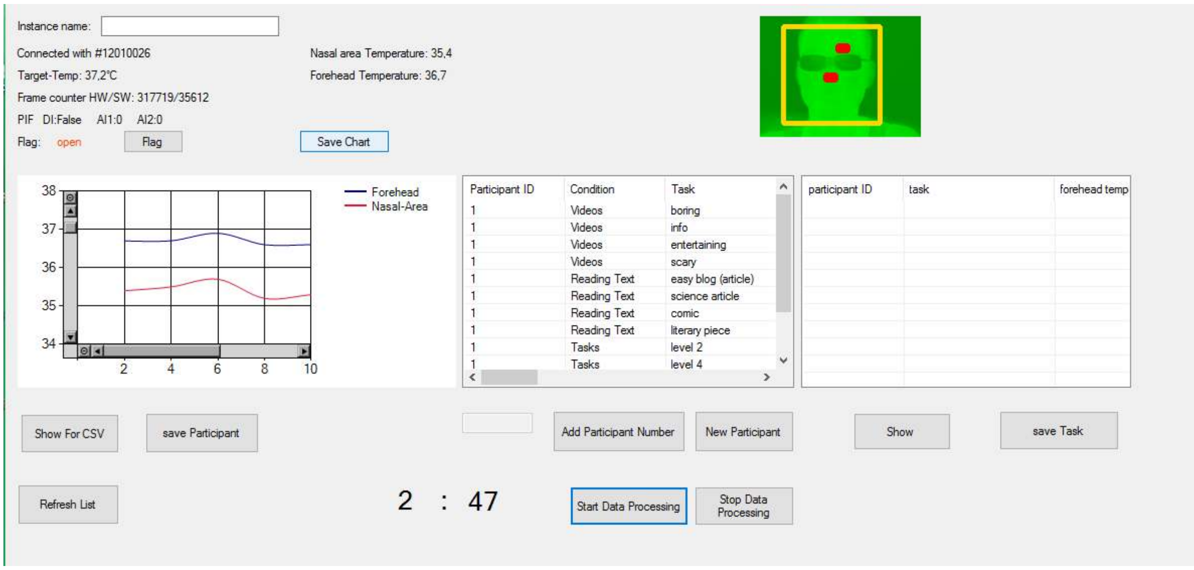

Fig. 3. The application interface used during the study, showing the detected face and ROI as well as visualizing their temperature and listing the tasks order.

All the data was visualized by the experimenter in real-time in an accompanying application developed in C# (see Figure 3).

4.5 Results

We analyzed the effect of the task difficulty on the recorded facial temperature. We used three metrics as our dependent variables:

(1) Decrease in nose temperature.
(2) Increase in forehead temperature.
(3) Difference between nose and forehead temperature.

We defined the temperature change as the difference between the mean temperature during the baseline recording and the mean temperature in the final minute of the task.

4.5.1 *Effect of Reading task on ROI Temperature.*

Nose Temperature. We tested the effect of the Content Difficulty on the Nose Temperature with a one-way ANOVA. Mauchly's test showed a violation of sphericity against Content Difficulty (0.07, p<.05), so we report Greenhouse-Geisser-corrected ($GGe = 0.45$) values. We found a large significant effect of Content Difficulty on the Nose Temperature ($F_{1.35,14.9} = 14.0, p < .0001, ges = 0.29$). Bonferroni-corrected pos-hoc tests found a statistically significant difference between all content types ($p < .05$), except between the blog and the science article, and between the science article and the literary piece at $p < .05$. The mean decrease in temperature between levels was of .27 degrees Celsius.

Forehead Temperature. We tested the effect of the reading Content Difficulty (4 levels) on the Forehead temperature (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against difficulty ($0.01, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.37$) values.

We found a significant large effect of Content Difficulty on the Forehead Temperature ($F_{1.12, 12.33} = 19.78, p < .001, ges = 0.16$). Bonferroni-corrected pos-hoc tests found a statistically significant difference between all content types ($p < .05$), except between the science article and the literary piece. The mean increase in temperature between levels of difficulty was .07 degrees Celsius.

Forehead-Nose Temperature Difference. We tested the effect on the difference between forehead and nose temperature (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against Content Difficulty ($0.09, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.44$) values. We found a large significant effect of Content Difficulty on the Forehead-Nose Difference ($F_{1.32, 14.54} = 23.26, p < .0001, ges = 0.38$). Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. The mean increase in temperature difference between the forehead and the nose between levels was of .34(±0.12) degrees Celsius.

Reading Task	Nose Temperature	Forehead Temperature	Forehead-Nose Temperature Difference
Comic	-0.69	0.23	0.92
Blog	-0.92	0.33	1.25
Article	-1.11	0.37	1.48
Old Lit.	-1.49	0.44	1.93

Table 1. Mean temperature change between the baseline and the Reading tasks at different readability index.

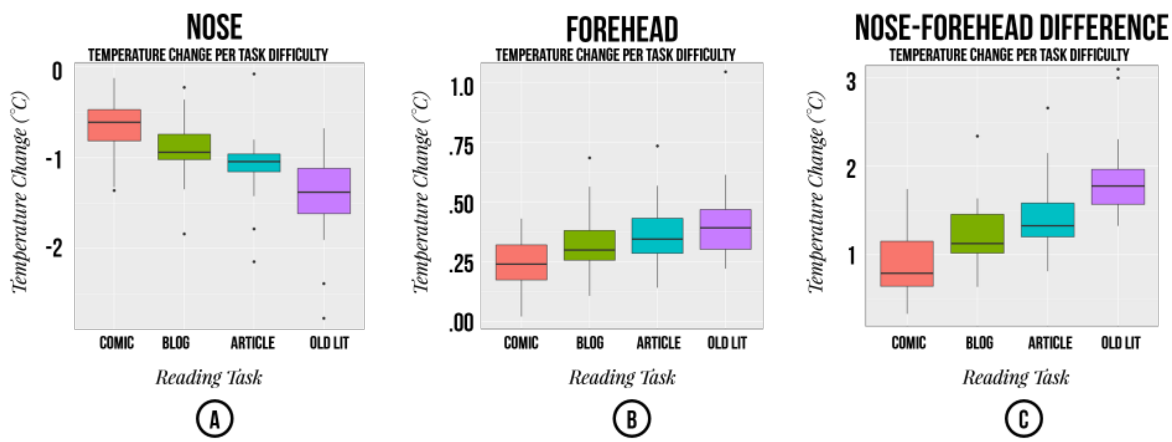

Fig. 4. Temperature change between the baseline and the Reading tasks at different levels of difficulty for (A) the nose, (B) the forehead, and (C) the difference between the forehead and the nose.

In summary, our reading tasks exhibited a significant increase in the forehead temperature and decrease in the nasal temperature. We found a significant difference between all contents for the increase in the forehead-nose temperature difference, and a larger effect size of the task difficulty on this metric. The difference between levels of difficulty in the order of .34 degrees Celsius.

4.5.2 *Effect of Stroop Task Levels on ROI Temperature.*

Nose Temperature. We then tested the effect of the TASK DIFFICULTY (4 levels) on NOSE TEMPERATURE (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against TASK DIFFICULTY ($0.08, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.68$) values. We found a large significant effect of task difficulty on NOSE TEMPERATURE ($F_{2.05,22.57} = 29.1, p < .0001, ges = 0.14$). Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. For each increase in the level of difficulty we found an decrease of 0.33 (±0.12) degrees celsius in nose temperature as estimated by a linear regression model.

Forehead Temperature. We tested the effect of the TASK DIFFICULTY (4 levels) on the FOREHEAD TEMPERATURE (difference to the baseline) with a one-way ANOVA. A Mauchly's test showed a violation of sphericity against Difficulty ($0.06, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.58$) values. We found a large significant effect of TASK DIFFICULTY on FOREHEAD TEMPERATURE ($F_{1.73,19.01} = 14.99, p < .001, ges = 0.31$). However, Bonferroni-corrected pos-hoc tests did not find a significant difference between levels 1 and 2, and between levels 3 and 4. For each increase in the level of difficulty we found an increase of 0.09 (±0.02) degrees Celsius in the forehead temperature as estimated by a linear regression model.

Stroop Level	Nose Temperature	Forehead Temperature	Forehead-Nose Temperature Difference
Level 1	-1.11	0.29	1.40
Level 2	-1.45	0.42	1.87
Level 3	-1.68	0.48	2.16
Level 4	-2.12	0.57	2.69

Table 2. Mean temperature change between the baseline and the Stroop tasks at different levels of difficulty.

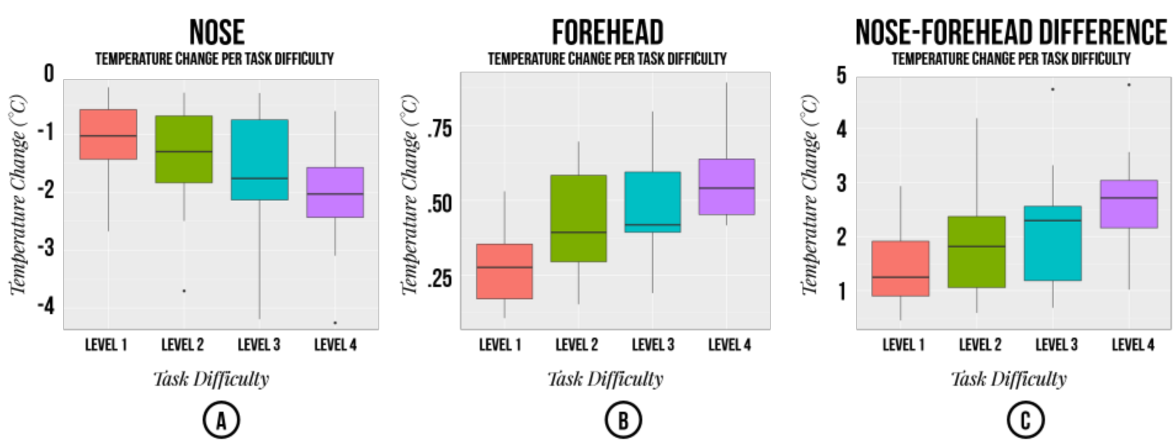

Fig. 5. Temperature change between the baseline and the Stroop tasks at different levels of difficulty for (A) the nose, (B) the forehead, and (C) the difference between the forehead and the nose.

Forehead-Nose Temperature Difference. We tested the effect of the Task Difficulty (4 levels) on the difference between forehead and the nose temperature (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against Task Difficulty ($0.18, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.68$) values. We found a large significant effect of Task Difficulty on Forehead-Nose Difference ($F_{2.03,22.36} = 37.97, p < .0001, ges = 0.20$).Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. For each increase in the level of difficulty we found a difference of 0.42 (±0.13) degrees Celsius in the difference between forehead and nose temperatures as estimated by a linear regression model.

In summary, we found statistically significant effects of task difficulty on temperature measures on (1) forehead, (2) nose, and (3) the difference between the forehead and the nose. The largest effect was found in the difference between the forehead and the nose.

4.6 Discussion

Informed by previous work, we hypothesized that an increase in the task difficulty would lead to a change in the participants' facial temperature patterns. Because forehead and nose are two of the most visible points on users' faces and are two points recommended deemed feasible for temperature measurement by previous work [24], we tested the effects of different tasks and their difficulties on temperature changes in these points.

We elicited increases in cognitive load both through an abstract task and through a naturalistic task. Our abstract task consisted of a variant of the classic Stroop test, in which we increased the task difficulty by introducing a time pressure. In this task, we found that an increase in the task's difficulty lead to a change both in forehead and nose temperature. The corresponding changes were related—an increase in task difficulty lead to an increase in forehead temperature and a decrease in nose temperature. We therefore combined both metrics by calculating the difference in temperature changes between the two. This proved to be the most robust metric, with a statistically significant average increase of 0.42 (±.13) degrees Celsius between each difficulty level.

We confirmed the validity of this finding in a naturalistic scenario, consisting of reading four pieces of text with varying levels of difficulty as measured by a readability scale—a comic book, a blog post, a scientific article, and a snippet of Old German literature. Again, we found a significant effect of the task difficulty for all metrics, in the same directions as in the Stroop task. Though this difference was not significant for all pairs of tasks in the forehead and nose temperatures alone, they were significant for all pairs when combining the two by subtracting the latter from the former. We found an average increase in the temperature difference between the forehead and the nose of .34 degrees Celsius for between each level of difficulty.

The forehead temperature increases are correlated with metabolic increases in this ROI. This is presumed to be due to the influence of muscle activation of the forehead muscle group [39, 59]. In parallel, the vessels in the nose region experience vasoconstriction (tightening in the blood vessels) as response to increased cognitive load [26, 59], reflecting a decrease in nose temperature.

Our findings validates the correlation of cognitive load and the selected region of interest and the measuring metrics we selected. Fernández-Cuevas et al. and Ioannou et al. [13, 24] summarized and presented how facial temperature and region of interest vary with different mental states. However, there were no states (e.g. stress, guilt, joy...etc) that correlated with an decrease in the nose and an increase in the forehead temperature. For instance, fear was correlated with decrease in both nose and forehead temperatures [24] and stress was correlated with variation in the nose temperature [24, 28]. Other works identified stress as an equation of the difference of the temperature of the nose and forehead [29] with a specific reading values between 34 and 36 degrees, rather than the total temperature change.

Our results answer **RQ1** by showing that the order of magnitude of the temperature changes are large enough to be detected by commercial sensors. For example, the FLIR One, a smartphone-compatible thermal camera and one of the most affordable devices currently in the market, is capable of detecting temperature changes of 0.1 degrees Celsius and can hence be used to detect cognitive load.

5 STUDY II: EVALUATION OF TEMPORAL LATENCY OF FACIAL TEMPERATURE CHANGE

Our first study validated the suitability of using the temperature differences between forehead and nose as a metric for cognitive load sensing. We found that the temperature changes are large enough for some of the cheapest thermal cameras in the market to capture. In the first study, we were interested in the *magnitude* of the temperature changes and therefore, we were only concerned with the average temperature at the end of the tasks. However, in a realistic scenario, we would be interested in pinpointing specific times in which changes in the facial thermal pattern could be detected. This would allow us to build cognition-aware systems that detect user state changes in real-time. For this purpose, it is crucial to understand the temporal response of these changes, which was the focus of our second study.

Other physiological sensors like GSR exhibit response times around three seconds [30]. The response latency achieved in previous works with thermal imaging include 10secs [30] on monkeys subjects and 3.8secs using functional thermal imaging [38]. The high latency found in previous works are not ideal for real-time applications. In our work, we wanted to investigate the latency, thereby investigating whether the current state of commercially available thermal imaging is appropriate for measuring cognitive load levels in real-time and in real world cognition-aware applications. To the best of our knowledge, no work has been done in evaluating the temporal latency of temperature changes using commercial thermal cameras which operate in the far infrared spectrum, particularly considering different stimuli durations. To answer **RQ2**, we evaluated the response time of the temperature change while considering different task durations ranging from 5 to 60 seconds.

5.1 Design

For this, study we applied a repeated-measures design, where all participants were exposed to all conditions. We studied the effect of the duration of the task on the latency of the temperature change. We chose the Stroop test as the task/stimulus, with task duration of 5, 15, 30, 45 and 60 seconds. Each duration value was repeated three times.

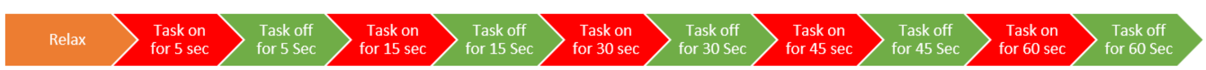

Fig. 6. Example for one study sequence; each task on/off cycle was repeated three times.

5.2 Apparatus

The general setup was the same as in the first study, except that we used a more precise thermal camera with higher thermal sensitivity: the Optris PI450[10] with an optical resolution of 382×288 pixels and a frame rate of 80 Hz. It measures temperatures between -20°C and 900°C and operates with a thermal sensitivity of 0.04°C represented by the noise equivalent temperature difference (NETD). The lens we use provides a 38°× 29° field of view.

[10]http://www.optris.com/thermal-imager-pi400

5.3 Participants and Procedure

We recruited 24 participants (13 females) with an average age of 30.8 years ($SD = 9.6$) using university mailing lists. Participants were different from those in our first study and mostly students in different majors and university staff. The participants were two groups: native Egyptians and Canadians. None of the participants had any previous experience with thermal cameras. Participants first signed a consent form and the purpose of the study was explained to them. Next, we asked them to relax for 10 minutes to ensure no other factors influencing the facial temperature for instance rushing into the study room. We recorded baseline temperature measures while participants were relaxing. They were then introduced to the Stroop task with different exposure duration, each with three minutes break between them. The order of the duration was counter-balanced using a Latin square. The study took approximately 60 minutes. During the entire experiment we recorded the temperature of the participant's forehead and nasal area. The experiment was conducted in a maintained room temperature of 26°C.

5.4 Results

We analyzed the effect of the different durations of the stimuli/tasks on the latency of facial temperature variations. As in study I, We used the same three metrics as our dependent variables: nose temperature, forehead temperature and differential temperature. Additionally, we investigated the following:

(1) Temperature change ***onset***, refers to the time taken to first observe a change in temperature after the commencement of the task. It is the time between the start of the task and the temperature reaching 3×standard deviation above the forehead baseline temperature or below nose baseline temperature. We picked this method, as 99.73% of the data should be within ± 3 times the standard deviation, hence values outside this range reflects temperature increase/decrease in the forehead and nose respectively.

(2) Temperature change ***saturation*** is the time taken to reach saturation in temperature change. This measure describes the time between the onset to the time the temperature lies between ± 3×standard deviation.

(3) Temperature change ***offset***, is the time taken after the task is stopped to observe temperature change. We computed based on the time it took between the endof the task and the temperature reaching 3× standard deviation below the forehead saturation temperature or above nose saturation temperature.

The baseline temperature for each participant was determined from the relaxing phase. This temperature was compared to the facial temperature during and after the task. We tested the effect of the task duration on the onset, saturation and offset times of the temperature change both in the nose and forehead area. There was no significant difference observed between the task duration and the three metrics.

Stroop Level	Nose Temperature	Forehead Temperature	Forehead-Nose Temperature Difference
Onset	0.7(±0.2)	1.2(±0.3)	0.7(±0.2)
Saturation	3.1(±1.2)	2.3(±0.9)	2.3(±1.2)
Offset	1.1(±0.5)	1.6(±0.9)	1.1(±0.5)

Table 3. Summary for the Onset, saturation and offset of the temperature changes in seconds.

5.4.1 *Latency in Nose Temperature Change.*

The onset for the nose temperature decrease was observed after 0.7s (± 0.2s) after the start of the task. It took 3.1s (± 1.2s) to reach saturation temperature . The offset for the nose temperature was observed after 1.1s (±0.5s) the end of the task.

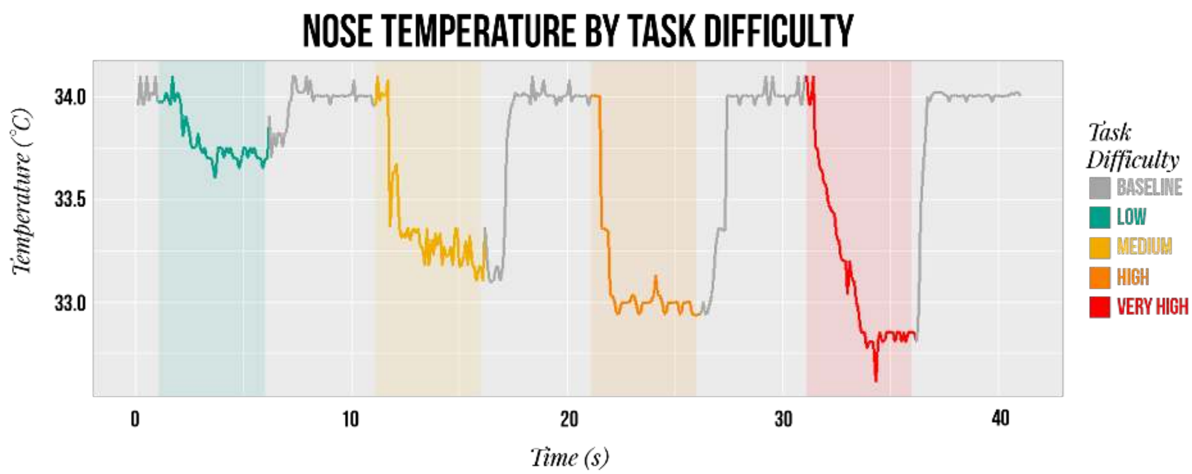

Fig. 7. Temperature decrease on the nose during the 4 levels of difficulty with task duration of 5 seconds.

5.4.2 Latency in Forehead Temperature Change. The onset for the forehead temperature increase was observed after an average of 1.2s (± 0.3s) after the start of the task. It took 2.3s (± 0.9s) to reach saturation of temperature increase. The offset for the forehead temperature was observed after 1.6s (±0.9s) after the task was finished.

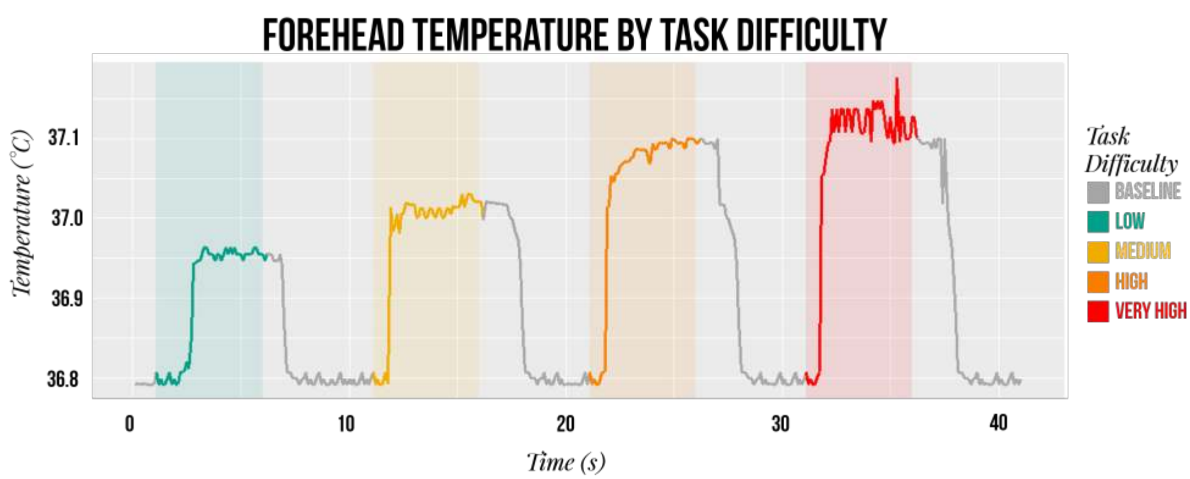

Fig. 8. Temperature increase on the forehead during the 4 levels of difficulty with 5 seconds task duration. The highlighted areas indicates the duration of the task.

5.4.3 Latency in Total Difference Temperature Change. The onset for the differential temperature change was observed after an average of 0.7 (± 0.2s) after the start of the task. It took 2.3s (± 1.2s) to reach the saturation temperature difference The offset for forehead-nose temperature difference was observed 1.1s (±0.5s) after the end of the task. As reported in study I, the Stroop test showed a statistically significant difference between the levels and baseline in the three metrics. This was confirmed in the second study.

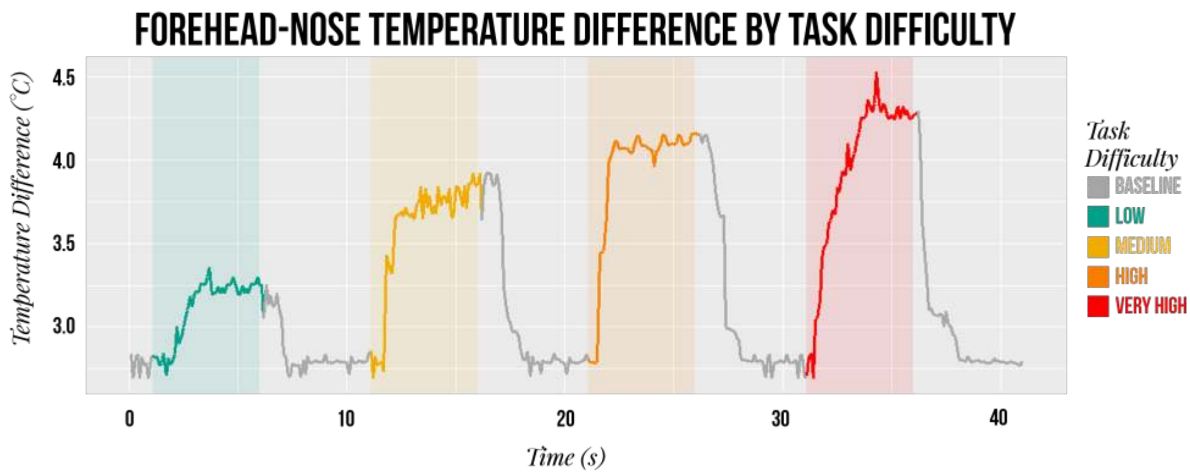

Fig. 9. Increase in the temperature difference between the forehead and the nose during the 4 levels of difficulty with task duration of 5 seconds.

5.5 Discussion

Thermal variations due to vascular changes were consider as slow compared to other physiological monitoring sensors [28]. However, recent research reported a latency of thermal response of 3.8 seconds after the stimuli onset using functional infrared imaging, compared to 3 seconds of GSR [30, 38]. Our findings indicate a response latency of 0.7sec using commercially available, far-infrared thermal imaging. One explanation for the faster response is the camera sensitivity as well as the frame rate, where the camera we used had 0.04K thermal sensitivity and 80 fps, as opposed to the camera used by Kang et al. [28], which had 50 fps and operated in different spectrum. Additionally, we relied on the temperature information of two regions of interest with a simple and real-time ROI extraction approach, which might have influenced the latency of the extracted and observed temperature changes.

As presented above, the onset in the forehead was longer than that of the nose 1.2±0.3 and 0.7±0.2 seconds respectively. This reflects the fact that there are more vessels affecting the subcutaneous temperature in the nose area than the forehead as reported by Berkovitz et al. [7]. This is also confirmed by the temperature variations, where a temperature change of 0.09° was observed in the forehead as opposed to 0.33° in the nose area, for each level.

As mentioned earlier, the participants were two groups native Egyptians and Canadians. However, the observed temperature change is consistent to that reported in the first study. This finding validates and confirms the results from the first study, in which we observed a consistent change in forehead, nose and differential temperatures for the different Stroop levels, although we had different groups of participants from that in the first study where we had only Germans.

Our findings from the second study answers **RQ2** and demonstrate the responsiveness of commercial thermal cameras in estimating cognitive load.

6 LIMITATIONS AND FUTURE WORK

Advances in thermal imaging have recently sparked a renewed interest in their applications within HCI. In this work we demonstrated how to use it to assess cognitive load levels. Our findings suggest that commercial thermal cameras are a suitable, touch-less, and unobtrusive method for assessing individual's cognitive load with minimum latency.

However, our approach has its limitations. We acknowledge that we considered a controlled setup, where the temperature of the surroundings were kept constant. Also participants are not allowed to touch their face (particularly at the forehead and nose), thus changing their temperature patterns. More sophisticated approaches (e.g., using machine learning) in the future could be used to consider the influence of the surrounding temperature and make our approach more robust. Additionally, we used a stationary indoors setup. However, it would be interesting to explore a setup for outdoors. Previous work reported no difference in the operation of thermal cameras in- and outdoors in interaction scenarios [5]. It would be interesting to explore if this holds true to the facial temperature as well.

In future work, our results can be used to generalize thermal monitoring for multiple users. We could consider wider scenarios including classroom and lifelogging scenarios (e.g., teacher monitoring students in a classroom to identify their cognitive load or track content being read). Furthermore, it would be interesting to extend our work to include neural networks and learning mechanisms for a better detection of state changes. Though we considered temperature changes in two ROIs the forehead and nose, it is also worth considering other points in the face or even on other parts of the body. Our findings strongly rely on the sensitivity of the thermal camera used. Higher thermal sensitivity would allow for detecting even smallest temperature variations, which could bring down latency even more to create real-time systems.

In this work, we focused on estimating cognitive load, but our method could also be applied to other scenarios, such as stress or affect recognition. In particular, we are interested in developing algorithms that are optimized to exploit thermal features to build a complete picture of users' affective states.

7 CONCLUSION

In this work we described our approach to unobtrusively derive users' cognitive load based on thermal imaging. Therefore, we investigated the effects of four different task intensity levels on facial temperature changes. We implemented a system capable of monitoring forehead and nose temperature to estimate current cognitive load levels through a novel metric based on the difference between forehead and nose temperature.

Thermal imaging operating in FIR provides novel avenues for studying users' cognitive states. We observed substantial changes in facial temperatures upon the activation of the ANS due to a stimulus. While the nose temperature—reflecting the vasoconstriction limiting the blood flow to the surface i.e., skin—decreases with rising workloads, in parallel, the temperature on the forehead increases as muscular activity leads to metabolic increases and increased blood flow in the underlying vessels. Based on these observations, we proposed a novel unobtrusive technique for estimating and quantifying cognitive load and possibly other affective states. In addition, we investigated the latency of temperature change and the ability of thermal cameras to capture those changes: we found an average latency of 0.7 ±0.2 seconds.

Therefore, our system was able to unobtrusively estimate changes in cognitive load in close to real-time. The exploration of content types gives rise to thermal-based activity tracking, which can empower new applications in the field of cognition-aware computing: thermal imaging techniques, for example, can be applied in classroom settings with multiple students being monitored in real-time to estimate cognitive load levels and assess current difficulty of content. It could also be used in assistive systems in production environment, where the worker is monitored unobtrusively without interrupting their work flow to estimate the current difficulty of the task in hand. Additionally, our proposed system could be utilized in usability testing help identify user interface features that increase cognitive load. Awareness of cognitive demand allows systems to dynamically adapt to users' current cognitive capacities and either reduce task difficulty to prevent frustration or add complexity to sustain interest and productivity.

8 ACKNOWLEDGMENTS

This work was partly conducted within the Amplify project which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 683008) and received funding from the German Research Foundation within the SimTech Cluster of Excellence (EXC 310/2).

REFERENCES

[1] Yomna Abdelrahman. 2013. *Thermal imaging for interactive surfaces.* Master's thesis.

[2] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3751–3763.

[3] Yomna Abdelrahman, Pascal Knierim, and Albrecht Schmidt. 2017. Snake View: Exploring Thermal Imaging as a Vision Extender in Mountains. In *(UbiMount) ubiquitous computing in the mountains*.

[4] Yomna Abdelrahman, Pascal Knierim, Pawel Wozniak, Niels Henze, and Albrecht Schmidt. 2017. See through the Fire: Evaluating the Augmentation of Visual Perception of Firefighters Using Depth and Thermal Cameras. In *(WAHM) Workshop on Ubiquitous Technologies for Augmenting the Human Mind*.

[5] Yomna Abdelrahman, Alireza Sahami Shirazi, Niels Henze, and Albrecht Schmidt. 2015. Investigation of Material Properties for Thermal Imaging-Based Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 15–18. https://doi.org/10.1145/2702123.2702290

[6] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276.

[7] Kirsch C. (Producer) Moxham J. B. (Producer) Alusi G. (Producer) Berkovitz, B. (Producer) and T. (Producer) Cheesman. 2013. 3D Head and Neck Anatomy with Special Senses and Basic Neuroanatomy. [DVD-ROM]. United Kingdom: Primal Pictures.

[8] Katie Crowley, Aidan Sliney, Ian Pitt, and Dave Murphy. 2010. Evaluating a Brain-Computer Interface to Categorise Human Emotional Response.. In *ICALT*. 276–278.

[9] Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. 2017. ABBAS: An Adaptive Bio-sensors Based Assistive System. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2543–2550.

[10] Douglas C Engelbart. 1962. Augmenting human intellect: a conceptual framework (1962). *Technical Report for the Air Force Office of Scientific Research* (1962).

[11] Veronika Engert, Arcangelo Merla, Joshua A Grant, Daniela Cardone, Anita Tusche, and Tania Singer. 2014. Exploring the use of thermal infrared imaging in human stress research. *PloS one* 9, 3 (2014), e90782.

[12] Stephen H. Fairclough. 2009. Fundamentals of Physiological Computing. *Interact. Comput.* 21, 1-2 (Jan. 2009), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

[13] Ismael Fernández-Cuevas, Joao Carlos Bouzas Marins, Javier Arnáiz Lastras, Pedro María Gómez Carmona, Sergio Piñonosa Cano, Miguel Ángel García-Concepción, and Manuel Sillero-Quintana. 2015. Classification of factors influencing the use of infrared thermography in humans: A review. *Infrared Physics & Technology* 71 (2015), 28–55.

[14] Markus Funk, Tilman Dingler, Jennifer Cooper, and Albrecht Schmidt. 2015. Stop helping me-I'm bored!: why assembly assistance needs to be adaptive. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 1269–1273.

[15] Rikke Gade and Thomas B Moeslund. 2014. Thermal cameras and applications: a survey. *Machine vision and applications* 25, 1 (2014), 245–262.

[16] Marc Germain, Michel Jobin, and Michel Cabanac. 1987. The effect of face fanning during recovery from exercise hyperthermia. *Canadian journal of physiology and pharmacology* 65, 1 (1987), 87–91.

[17] Nitesh Goyal and Susan R. Fussell. 2017. Intelligent Interruption Management using Electro Dermal Activity based Physiological Sensor on Collaborative Sensemaking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 16.

[18] M Haak, S Bos, S 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.

[19] Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on* 6, 2 (2005), 156–166.

[20] Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. 2014. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 51–60.

[21] Sarah M Hosni, Mahmoud E Gadallah, Sayed F Bahgat, and Mohamed S AbdelWahab. 2007. Classification of EEG signals using different feature extraction techniques for mental-task BCI. In *Computer Engineering & Systems, 2007. ICCES'07. International Conference on*. IEEE, 220–226.

[22] Curtis S Ikehara and Martha E Crosby. 2005. Assessing cognitive load with physiological sensors. In *Proceedings of the 38th annual hawaii international conference on system sciences*. IEEE, 295a–295a.

[23] Stephanos Ioannou, Sjoerd Ebisch, Tiziana Aureli, Daniela Bafunno, Helene Alexi Ioannides, Daniela Cardone, Barbara Manini, Gian Luca Romani, Vittorio Gallese, and Arcangelo Merla. 2013. The autonomic signature of guilt in children: a thermal infrared imaging study. *PloS one* 8, 11 (2013), e79440.

[24] Stephanos Ioannou, Vittorio Gallese, and Arcangelo Merla. 2014. Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology* 51, 10 (2014), 951–963.

[25] Giulio Jacucci, Anna Spagnolli, Jonathan Freeman, and Luciano Gamberini. 2014. *Symbiotic Interaction: A Critical Definition and Comparison to other Human-Computer Paradigms*. Springer International Publishing, Cham, 3–20. https://doi.org/10.1007/978-3-319-13500-7_1

[26] 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.

[27] Jihun Kang and Kari Babski-Reeves. 2008. Detecting Mental Workload Fluctuation during Learning of a Novel Task Using Thermography. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 19 (2008), 1527–1531. https://doi.org/10.1177/154193120805201947 arXiv:http://dx.doi.org/10.1177/154193120805201947

[28] J Kang, JA McGinley, G McFadyen, and K Babski-Reeves. 2006. Determining learning level and effective training times using thermography. In *Proceedings of Army Science Conference, Orlando, Florida, USA*.

[29] Hisanori Kataoka, Hiroshi Kano, Hiroaki Yoshida, Atsuo Saijo, Masashi Yasuda, and Masato Osumi. 1998. Development of a skin temperature measuring system for non-contact stress evaluation. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, Vol. 2. IEEE, 940–943.

[30] Koji Kuraoka and Katsuki Nakamura. 2011. The use of nasal skin temperature measurements in studying emotion in macaque monkeys. *Physiology & behavior* 102, 3 (2011), 347–355.

[31] Hindra Kurniawan, Alexey V Maslov, and Mykola Pechenizkiy. 2013. Stress detection from speech and galvanic skin response signals. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*. IEEE, 209–214.

[32] Elise Labbé, Nicholas Schmidt, Jonathan Babin, and Martha Pharr. 2007. Coping with stress: the effectiveness of different types of music. *Applied psychophysiology and biofeedback* 32, 3-4 (2007), 163–168.

[33] Eric Larson, Gabe Cohn, Sidhant Gupta, Xiaofeng Ren, Beverly Harrison, Dieter Fox, and Shwetak Patel. 2011. HeatWave: Thermal Imaging for Surface User Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2565–2574. https://doi.org/10.1145/1978942.1979317

[34] Robert W Levenson. 1988. Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. (1988).

[35] Joseph CR Licklider. 1960. Man-computer symbiosis. *IRE transactions on human factors in electronics* 1 (1960), 4–11.

[36] Barbara Manini, Daniela Cardone, Sjoerd JH Ebisch, Daniela Bafunno, Tiziana Aureli, and Arcangelo Merla. 2013. Mom feels what her child feels: thermal signatures of vicarious autonomic response while watching children in a stressful situation. *Front. Hum. Neurosci* 7 (2013).

[37] TV McCaffrey, RD McCook, and RD Wurster. 1975. Effect of head skin temperature on tympanic and oral temperature in man. *Journal of Applied Physiology* 39, 1 (1975), 114–118.

[38] Arcangelo Merla and Gian Luca Romani. 2007. Thermal signatures of emotional arousal: a functional infrared imaging study. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 247–249.

[39] Takakazu Oka, Kae Oka, and Tetsuro Hori. 2001. Mechanisms and mediators of psychological stress-induced rise in core temperature. *Psychosomatic medicine* 63, 3 (2001), 476–486.

[40] 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.

[41] Avinash Parnandi, Beena Ahmed, Eva Shipp, and Ricardo Gutierrez-Osuna. 2013. Chill-Out: Relaxation training through respiratory biofeedback in a mobile casual game. In *Mobile Computing, Applications, and Services*. Springer, 252–260.

[42] Avinash Parnandi, Youngpyo Son, and Ricardo Gutierrez-Osuna. 2013. A control-theoretic approach to adaptive physiological games. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 7–12.

[43] Ioannis Pavlidis and James Levine. 2002. Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine* 21, 6 (2002), 56–64.

[44] Ioannis Pavlidis, Panagiotis Tsiamyrtzis, Dvijesh Shastri, Avinash Wesley, Yan Zhou, Peggy Lindner, Pradeep Buddharaju, Rohan Joseph, Anitha Mandapati, Brian Dunkin, et al. 2012. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports* 2 (2012).
[45] Michael Kai Petersen, Carsten Stahlhut, Arkadiusz Stopczynski, Jakob Eg Larsen, and Lars Kai Hansen. 2011. Smartphones get emotional: mind reading images and reconstructing the neural sources. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 578–587.
[46] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA. https://doi.org/10.1145/2858036.2858117
[47] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.
[48] Colin Puri, Leslie Olson, Ioannis Pavlidis, James Levine, and Justin Starren. 2005. StressCam: non-contact measurement of users' emotional states through thermal imaging. In *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 1725–1728.
[49] B. A. Rajoub and R. Zwiggelaar. 2014. Thermal Facial Analysis for Deception Detection. *IEEE Transactions on Information Forensics and Security* 9, 6 (June 2014), 1015–1023. https://doi.org/10.1109/TIFS.2014.2317309
[50] Alireza Sahami Shirazi, Yomna Abdelrahman, Niels Henze, Stefan Schneegass, Mohammadreza Khalilbeigi, and Albrecht Schmidt. 2014. Exploiting Thermal Reflection for Interactive Systems. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3483–3492. https://doi.org/10.1145/2556288.2557208
[51] Albrecht Schmidt. 2017. Augmenting Human Intellect and Amplifying Perception and Cognition. *IEEE Pervasive Computing* 16, 1 (2017), 6–10.
[52] Stefan Schneegass, Bastian Pfleging, Nora Broy, Frederik Heinrich, and Albrecht Schmidt. 2013. A data set of real world driving to assess driver workload. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 150–157.
[53] Nandita Sharma and Tom Gedeon. 2012. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine* 108, 3 (2012), 1287–1301.
[54] Dvijesh Shastri, Arcangelo Merla, Panagiotis Tsiamyrtzis, and Ioannis Pavlidis. 2009. Imaging facial signs of neurophysiological responses. *IEEE Transactions on Biomedical Engineering* 56, 2 (2009), 477–484.
[55] Dvijesh Shastri, Manos Papadakis, Panagiotis Tsiamyrtzis, Barbara Bass, and Ioannis Pavlidis. 2012. Perinasal imaging of physiological stress and its affective potential. *IEEE Transactions on Affective Computing* 3, 3 (2012), 366–378.
[56] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2651–2656.
[57] Alireza Sahami Shirazi, Markus Funk, Florian Pfleiderer, Hendrik Glück, and Albrecht Schmidt. 2012. MediaBrain: Annotating Videos based on Brain-Computer Interaction.. In *Mensch & Computer*. 263–272.
[58] Alireza Sahami Shirazi, Mariam Hassib, Niels Henze, Albrecht Schmidt, and Kai Kunze. 2014. What's on your mind?: mental task awareness using single electrode brain computer interfaces. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 45.
[59] Rajita Sinha, William R Lovallo, and Oscar A Parsons. 1992. Cardiovascular differentiation of emotions. *Psychosomatic Medicine* 54, 4 (1992), 422–435.
[60] 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.
[61] J Ridley Stroop. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General* 121, 1 (1935), 15.
[62] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. *Measuring Cognitive Load*. Springer New York, New York, NY, 71–85. https://doi.org/10.1007/978-1-4419-8126-4_6
[63] Boris M Velichkovsky and John Paulin Hansen. 1996. New technological windows into mind: There is more in eyes and brains for human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 496–503.
[64] Paul Viola and Michael J Jones. 2004. Robust real-time face detection. *International journal of computer vision* 57, 2 (2004), 137–154.
[65] Zixuan Wang, Jinyun Yan, and Hamid Aghajan. 2012. A framework of personal assistant for computer users by analyzing video stream. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. ACM, 14.
[66] Mark Weiser. 1991. The computer for the 21st century. *Scientific american* 265, 3 (1991), 94–104.
[67] 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.
[68] Zhen Zhu, Panagiotis Tsiamyrtzis, and Ioannis Pavlidis. 2007. Forehead thermal signature extraction in lie detection. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 243–246.

Received February 2017; revised May 2017; accepted June 2017