

# Can Infrared Facial Thermography Disclose Mental Workload in Indoor Thermal Environments?

Xi Wang University  
of Michigan Ann  
Arbor, Michigan  
wangix@umich.edu

Carol C. Menassa  
University of Michigan  
Ann Arbor, Michigan  
menassa@umich.edu

Da Li University of  
Michigan Ann Arbor,  
Michigan  
dliseren@umich.edu

Vineet R. Kamat  
University of Michigan  
Ann Arbor, Michigan  
vkamat@umich.edu

## ABSTRACT

Mental workload represents the mental resources an individual devotes to a task. In a building environment, understanding how ambient thermal conditions affect occupants' mental workload offers an opportunity to achieve optimal thermal settings for the heating, ventilation, and air conditioning (HVAC) systems. However, directly measuring mental workload on a large and continuous scale requires occupants to perform subjective tests or wear electroencephalogram (EEG) or similar devices, which is impractical. This paper assesses the feasibility of using infrared facial thermography captured by a low-cost thermal camera to disclose mental workload. An experiment was conducted to measure the facial skin temperature while subjects performed cognitive tasks in three different thermal environments, representing occupants' thermal sensation of slightly cool, neutral, and slightly warm. Mental workload was measured using an EEG headset to eliminate subjective bias. The correlations between facial temperature and mental workload vary with different individuals and thermal conditions. Relatively strong correlations are found in the neutral environment and in the regions of ears, mouth, and neck. The results also suggest that future work should collect data under extended experiment duration. This is because it was observed that the response of facial skin temperature to mental workload varies with task type; thus, increasing the repetitiveness for each type of task or using more challenging tasks in the experiment could potentially lead to more insights on this relationship.

## CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; • **General and reference** → *Validation*; Experimentation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*UrbSys'19, November 13–14, 2019, New York, NY, USA*

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7014-1/19/11...\$15.00

<https://doi.org/10.1145/3363459.3363528>

## KEYWORDS

mental workload, facial thermography, thermal comfort, electroencephalography (EEG)

### ACM Reference Format:

Xi Wang, Da Li, Carol C. Menassa, and Vineet R. Kamat. 2019. Can Infrared Facial Thermography Disclose Mental Workload in Indoor Thermal Environments?. In *1st ACM International Workshop on Urban Building Energy Sensing, Controls, Big Data Analysis, and Visualization (UrbSys'19), November 13–14, 2019, New York, NY, USA*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3363459.3363528>

## 1 INTRODUCTION

Building indoor environment plays an important role in people's well-being. In the U.S. and Canada, people on the average spend more than 90% of their time indoors [21]. Several researchers studied the impact of indoor environmental quality (IEQ) factors on building occupants, including lighting [18], acoustic conditions [46], and thermal comfort [23]. Good IEQ can improve occupants' satisfaction, health, and work performance [7]. Therefore, in recent years, much effort has been placed to evaluate and improve the IEQ in the workplace [37].

Thermal comfort, as one of the most important IEQ factors, can significantly affect the building energy consumption, as well as occupants' well-being and satisfaction with the built environment [8, 22]. As the largest energy-consuming sector, the heating, ventilation, and air conditioning (HVAC) system accounts for approximately 48% of building energy, primarily for providing occupants with comfortable working and living spaces [30]. In spite of the significant energy footprint to maintain desired thermal comfort, studies show that up to 43% occupants are not satisfied with their workplace thermal environments [19]. Dissatisfied thermal environment can lead to sick building syndrome symptoms, such as skin dryness, eye irritation, and headache [46]. In addition, poor thermal conditions can increase the number of complaints, absenteeism and reduce occupants' productivity [14].

Studies have suggested that thermal environments have an effect on occupants' mental workload, which is also related to occupants' work performance and health [13]. Mental workload is defined as "the mental resources devoted to the task of an individual" [12]. It reflects the subjects' personal efforts when performing tasks under a specific condition and describes "an overall efficiency of human performance" [20]. Changes in mental workload would ultimately

cause changes in occupants' performance when it becomes high and approaches to the mental resource capacity of an individual [13]. In relation to thermal environments, Wang et al. (2019) found that the slightly warm indoor environment (28.6°C/83.5°F) results in higher mental workload demand in occupants [42]. Continuous high mental workload will result in subjects' mental fatigue and cause mistakes, efficiency and alertness decline, and effort disinclination [47]. In the long run, it would result in detrimental health effects on occupants [14]. Therefore, thermal environment should be controlled to yield an appropriate amount of mental workload.

To investigate the effect of thermal environments on occupants' mental workload, Wang et al. (2019) analyzed the electroencephalogram (EEG) signals of occupants performing standardized tasks under different room temperatures [42, 43]. EEG is a technique to monitor and record the brain electrical activity through the electrodes placed on the scalp surface [15, 16]. Although EEG is sensitive to mental workload as it directly captures central brain activities and subjects' cognitive states, it can cause significant intrusiveness on the occupants (e.g. headaches if worn for a long time, difficulties to set up for subjects with thick hair, etc.), which prevents its applications in real operational built environments. Thus, developing a non-intrusive method that can dynamically reflect the mental workload induced by thermal environments is critical to achieving a healthy and productive built environment.

In this research, we explore the use of facial infrared thermography as a proxy to measuring mental workload reported through EEG signals. Li et al. (2019) proposed a thermal camera network that can simultaneously collect facial temperature from multiple occupants and predict their thermal comfort [26]. Several researchers explored the relationship between facial thermography collected from advanced thermal camera models and the subjective-rated mental workload [27, 32]. However, whether the facial thermography captured by a low-cost thermal camera is feasible to be used as the proxy to predict mental workload is still in question. In this study, the authors conducted experiments in three different thermal environments, representing the thermal sensation of slightly cool, neutral, and slightly warm, to evaluate the relationship between the temperature of different facial regions and mental workload measured by EEG. The paper is organized to first review the existing methods for mental workload measurement in Section 2, followed by discussions of their main limitations and the formulation of our hypothesis to bridge the research gaps. Next, the research methods and experimental protocol are presented in Section 3. Then, the results are explained and discussed in detail in Section 4. Finally, the main conclusions of this study are highlighted in Section 5.

## 2 BACKGROUND

In this section, three categories of methods to measure mental workload and their limitations are discussed. The first category of methods measures mental workload through subjective ratings by asking subjects a series of questions about their self-perception. On the other hand, the second category assesses mental workload by using EEG to measure central nervous activities. In the third category, peripheral physiological data are examined to reflect mental workload, where the facial thermography-based method was discussed in detail.

### 2.1 Subjective mental workload measurement

The most common method to assess mental workload is through subjective ratings. Due to its simplicity and high sensitivity [3], subjective ratings are often used as the baseline to evaluate the effectiveness of other approaches to assess mental workload [27, 32, 38]. In existing studies, researchers have proposed different unidimensional rating scales. For example, Brennan (1992) developed the Instantaneous Self-Assessment method for real-time workload assessment, which is a 5-point rating scale from "Excessive workload" to "Under-utilized workload" [2]. Wierwille and Casali (1983) developed the Modified Cooper-Harper Scale, which is a 10-point unidimensional rating scale [44]. It provides a global rating of perceptual, cognitive, and communications workload. However, the subjective ratings are highly biased. Some researchers argued that these unidimensional scales, despite their convenience, are not sufficient to reflect the complex nature of the mental workload [28].

Thus, multidimensional subjective mental workload measurement approaches have been developed. Two most popular ones are the NASA-Task Load Index scale (NASA-TLX) and the Subjective Workload Assessment Technique (SWAT). NASA-TLX uses six subscales to assess mental workload, including mental demand, physical demand, temporal demand, performance, effort, and frustration [11]. Similarly, the SWAT produces an interval rating scale of mental workload from three dimensions, including time load, mental load, and physiological stress load. Each dimension has three levels: low, medium, and high [36].

Compared to unidimensional scales, the multidimensional mental workload measures have the merits of incorporating insights from multiple aspects [28]. However, several deficiencies prevent this approach from being applied in actual working or living settings. First, the rating process is relatively time-consuming (typically 5-10 minutes) [34], which can interrupt subjects' normal work or harm their work performance. Meanwhile, the ratings can be inaccurate since subjects make responses based on their memory of the mental workload during the tasks. Furthermore, the assessment tools are rather complicated and require proper training or experience [31]. The complicated assessment processes can pose extra workload on subjects, which will affect their judgments on mental workload during designated tasks.

### 2.2 EEG mental workload measurement

EEG can monitor subjects' cognitive states by directly capturing the brain electrical activity [37], and thus it is widely adopted by researchers to measure the mental workload [28]. According to the signal frequency, EEG rhythmic activities can be divided into 5 different frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz) and gamma (>25 Hz) [29]. As mental workload increases, the theta band activity of frontal lobe increases and the alpha band activity of parietal lobe decreases [10, 14, 20].

Even though the emergence of low-cost and wireless EEG makes it affordable and easier to use, its application to measure mental workload is still limited to laboratory settings. In previous studies [42], it was found that the tight clamping of the EEG headset can cause headaches after wearing it for about 30-40 minutes. As a result, the headset needs to be removed during the experiment

to relax subjects and reset after the rest, which prevents the continuous measurement of mental workload. In addition, the set-up process and tight contact to subjects' scalp also interfere with their daily work. Second, for a multi-occupancy built environment, an EEG headset is needed for each occupant which can be cost prohibitive and impractical to implement. Third, to obtain accurate EEG measurements, subjects are required to stay in steady states since physical movements will induce noises to EEG signals and greatly reduce the signal-to-noise ratio. However, regular work involves a great number of movements and communications. Thus, keeping subjects in steady states is both impractical and intrusive.

### 2.3 Peripheral physiological mental workload measurement

In addition to subjective ratings and EEG, a variety of peripheral physiological measurements have been used to assess mental workload. For example, heart rate and heart rate variability (HRV, i.e., the beat-to-beat variation in heart rate) have shown inconsistent correlations with mental workload. Hankins and Wilson (1998) and Wilson (2002) found that heart rate increases with increasing mental workload [10, 45]. Jorna (1992) found that higher mental workload cause decreases in HRV [17]. However, some researchers claimed no decrease in HRV was found with increasing mental workload [10, 39]. Others believe that heart rate and HRV can be influenced by multiple physiological and environmental factors, as well as subjects' emotions, and thus are not robust enough to measure mental workload [28]. In terms of other physiological measurements, Recarte and Nunes (2003) discovered a significant increment in pupil diameter with increasing mental workload [35]. Veltman and Gaillard (1998) also observed an increase in eye blink interval and a decrease in eye blink duration during higher mental workload [40]. Nonetheless, researchers claim that eye blinks reflect visual workload instead of cognitive workload [4]. Respiratory rate was also used for mental workload measurement in some studies [9, 39]. However, Marinescu et al. (2017) found the correlation between respiratory rate and mental workload was insignificant for most subjects [27]. These peripheral physiological measurements assess subjects' mental workload objectively through their physical responses to the imposed workload. Nevertheless, these methods suffer from the same limitations as EEG (i.e., cost prohibitive and intrusive to occupants).

As a non-intrusive method, the thermal camera can detect the variation of face temperature caused by the vasoconstriction response of the autonomic nervous system while subjects experience stress or negative emotions [32]. Therefore, the relationship between facial thermography and mental workload have been explored by several researchers. Or and Duffy (2007) observed a significant correlation between variations in nose temperature and mental workload [32]. Abdelrahman et al. (2017) found the facial thermography pattern changes with the Stroop test difficulty and reading text complexity [1]. Marinescu et al. (2017) also discovered that facial temperature could increase the mental workload prediction accuracy by 47.7% [27]. However, two limitations of these studies should be acknowledged. First, they either infer the increase of mental workload by increasing task difficulty or use subjective mental workload ratings, which is not as accurate as

EEG which measures mental workload directly from the central nervous activity. As proposed by the Cognitive Reserve theory, when the subject's neural resource is not sufficient to cope with the demands, mental workload will not increase even though the task difficulty is increasing [13]. Meanwhile, subjective mental workload ratings used as the baseline are biased towards subjects' perception and memory as discussed in Section 2.1 [11]. Second, even though thermal camera can be configured to capture the facial skin temperature of multiple occupants instead of having one sensor for each subject, meaning the number of devices can be greatly reduced in multi-occupancy environments, it still suffers from the cost prohibitive problem because even a single camera used in these studies is costly (\$3,000 - \$15,000).

The objectives of this study are:

- Develop a framework to study whether facial infrared thermography captured by a low-cost thermal camera (\$200) can disclose the mental workload measured by EEG. The low-cost thermal camera network has high scalability potential and flexibility to capture facial infrared thermography in various building environments [26].
- Present preliminary results and develop recommendations for future studies in this area where non-intrusive methods can be used to simultaneously assess the mental workload of all occupants in multi-occupancy building environments.

## 3 METHODOLOGY

This study employs a variety of techniques to investigate the capability of facial infrared thermography captured by the low-cost thermal camera to disclose mental workload. Considering the impact of indoor thermal environments on mental workload and facial thermography pattern [24, 42], the authors choose three typical thermal sensations occupants experience in the built environment, namely slightly cool, neutral, and slightly warm. The room temperature and relative humidity of each condition is as shown in Table 1. They are determined based on Fanger's Predicted Mean Vote model, which predict occupants' thermal comfort level using four environment parameters (air temperature, mean radiant temperature, air velocity, and relative humidity) and two occupant-related parameters (metabolic rate and clothing insulation) [6].

An overview of the research methodology and data analysis method is shown in Figure 1. In each thermal condition, mental workload is measured by EEG while subjects perform tasks (see Section 3.1 for task details). In the meantime, facial thermography is measured using a low-cost thermal camera (see Figure 2, the camera was placed beneath the computer screen). After the timestamp calibration of the facial thermography and EEG-measured mental workload (i.e. align them onto the same timestep), the correlation between them is examined for each subject. Based on the correlation result, prediction models are trained and tested to disclose the mental workload from selected facial skin temperature features.

### 3.1 Baseline and Cognitive tasks

At the start of the experiment in each thermal condition, subjects were asked to try their best to relax and keep their mind blank to record a baseline EEG and facial thermography data, which is recognized as the subjects' status with the lowest mental workload.

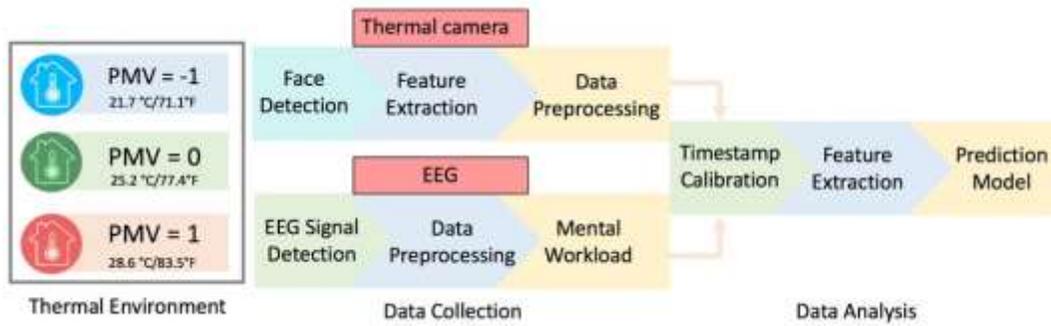


Figure 1: Research Methodology Overview

Table 1: Thermal environment settings

PMV	Thermal Sensation	Air Temperature	Humidity
PMV=-1	Slightly Cool	21.7 °C (71.1 °F)	23% ±1%
PMV=0	Neutral	25.2 °C (77.4 °F)	23% ±1%
PMV=1	Slightly Warm	28.6 °C (83.5 °F)	23% ±1%

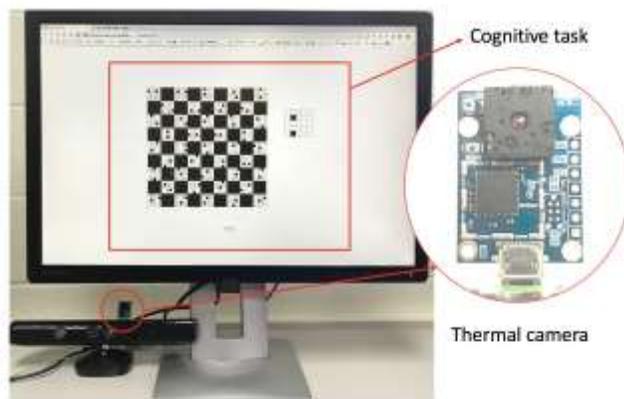


Figure 2: Thermal camera setting

During the experiment, subjects were asked to perform four cognitive tasks, including number addition, digit span, choice reaction, and visual search on a desktop computer. The cognitive tasks were developed using Javascript. No calculator or note-taking was allowed during the task period. The number addition task requires subjects to add up a column of numbers and consists of two difficulty levels (Figure 3a). The digit span task requires subjects to memorize and recall a series of numbers appeared on the computer screen one after another, which also consists of two difficulty levels (Figure 3b). The choice reaction task consists of three parts, including the simple reaction task (easy level), the four-choice reaction task (hard level, requires subjects to press the corresponding key when the target emerge in different boxes), and the Stroop task (Figure 3c). The visual search task only has one difficulty level. It requires subjects to search for the target pattern among given

choices rapidly and accurately (Figure 3d). The detailed description of the four cognitive tasks could be found in [47].

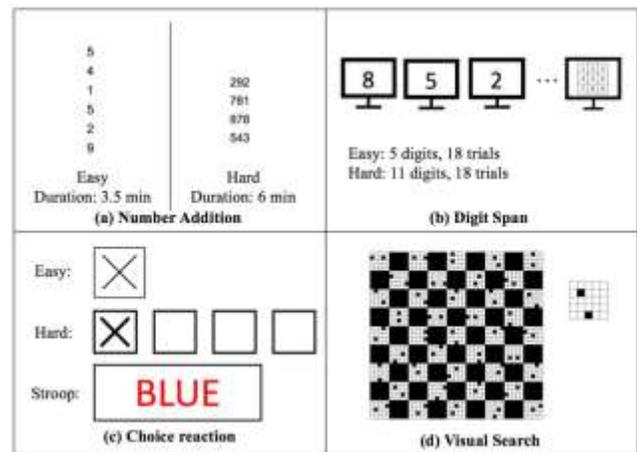


Figure 3: Cognitive task interface

### 3.2 Low-cost thermal camera and facial skin temperature extraction

In this study, the FLIR Lepton 2.5 radiometric thermal camera was used to capture facial thermography. It is a low-cost, uncooled thermal imaging core that provides a factory-calibrated temperature value through long-wave infrared thermography. Compared to the advanced thermal camera models, it has relatively low radiometric accuracy ( $\pm 5^\circ\text{C}$  or  $\pm 0.5\%$  of reading) and low resolution ( $80(\text{h}) \times 60(\text{v})$  pixels). However, its feasibility to disclose occupants' mental workload is still worth investigation because continuous thermal videos consist of multiple frames are adopted in this study instead of a single image frame. As a result, the error of temperature measurement, which follows a Gaussian distribution, could be reduced by removing outliers and averaging data within a period of time. Meanwhile, Li et al. (2018) have successfully used the facial skin temperature data collected by the same thermal camera module to predict occupants' thermal comfort level [24, 25].

In spite of the low resolution of the thermal camera, the contour of interested regions on human faces (e.g. forehead, nose) are preserved in the thermal image (Figure 4). As a result, Haar Cascade algorithm can be applied to detect the existence of faces in thermal images [41]. Haar Cascade algorithm can detect the existence of certain image characteristics (e.g. edges, changes in texture), and thus could be used for frontal and profile face recognition [41]. Based on previous studies, we select forehead, nose, cheeks, ears, mouth, and neck as ROI (Regions of Interest). The face contour, the nose, and the eyes are first detected by the Haar Cascade algorithm directly. The other ROI are inferred during the runtime based on their size and relative location to the already detected face regions (i.e. facial contour, eyes, and nose), which had been tuned on several subjects in advance to ensure correct recognition (Figure 4). For each ROI, we use the average temperature of all pixels within the identified region to represent the corresponding skin temperature. Pixel values that are lower than  $28^{\circ}\text{C}$  or higher than  $38^{\circ}\text{C}$  in each ROI are recognized as outliers or noises and are excluded from the analysis.

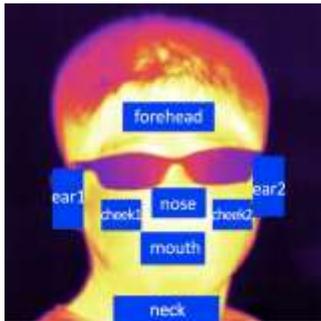


Figure 4: Facial region of interest

In this study, only one subject was tested at a time and the camera was placed at a 0.8-1m distance in front of the subject's face. However, the thermal camera module could be assembled with Microsoft Kinect into a thermal camera network that is capable of extracting facial skin temperature of multiple occupants at the same time from a longer distance (about 2m) [26].

### 3.3 Experiment

A three-day experiment was designed to find the relationship between facial skin temperature and EEG-measured mental workload. The experiment was conducted in a research office with the air quality, lighting, and ventilation condition controlled to be the same during the experiments. The room temperature was controlled by a thermostat connected to the air-conditioning system. The air temperature and relative humidity of subjects' ambient environment were measured using two COZIR sensors placed 0.6m above the floor level. The experiment protocol was approved by the university's Institutional Review Board.

All 15 subjects were recruited from healthy graduate students, who were well-acquainted with computer use. Subjects were required to keep their hair clear, remove makeup, and avoid energy drink or alcohol before coming to the experiment. They were also asked to pay full attention to the tasks and try their best to get

the correct answers. In order to improve consistency and eliminate irrelevant factors, subjects were asked to come to the testbed office at the same time on three consecutive days to avoid the circadian effects. The order of different thermal conditions was randomly shuffled for different subjects. Meanwhile, a good night's sleep and consistent sleep time were recommended on the nights before the experiment. The experiment data of three subjects were excluded from analysis due to connectivity loss or excessive sweating.

On each experiment day, subjects followed the experiment procedures as outlined in Figure 5. In the preparation phase, subjects spent 30 minutes relaxing in the testbed office to adapt to the environment and receive experiment instructions. EEG headset and thermal camera were set up subsequently to record the baseline data, followed by the first cognitive task session that randomly selected two tasks from the four cognitive tasks. After a 15-minute rest with EEG headset removed, EEG and thermal camera were reset and subjects continued to perform the second cognitive task session that consists of the other two tasks. While subjects perform tasks, the authors constantly checked the EEG channel connectivity through the interface of the bundled software to ensure the success of the experiment and the quality of data collected.

### 3.4 Data preprocessing and analysis

Mental workload was measured using a wireless low-cost EEG headset Emotiv EPOC+ [5], which has 14 channels with the sampling frequency of 128 Hz. The absolute power of each frequency band was calculated 8 times per second using the bundled software. The absolute power data was sliced into data segment according to the start and end timestamp of each cognitive task recorded by the system, with each segment containing the EEG data of one difficulty level of a cognitive task. Since the data collected by the thermal camera was recorded once per second, which has fewer recordings than the EEG relative power, and it may lose some frames when it fails to detect the ROI, the facial thermography and EEG data need to be calibrated into an aligned timestep. The method for calibration is: (1) for each frame in facial thermography data, find the data point in the EEG data with the nearest timestamp to that of facial skin temperature; (2) assign the facial skin temperature to the nearest EEG timestamp; (3) use forward filling to fill the nearest facial skin temperature data in the forward to the timestamp that with no facial data assigned; (4) if the first few timestamps in EEG were not assigned facial data (i.e. no forward data can be filled in), the nearest facial data backward was used for these timestamps. This method is feasible because human facial skin temperature does not change rapidly in a very short time period. After calibration, a size 199 median window and a size 101 Savitzky-Golay filter with polynomial order 3 were applied to smooth and remove the outliers of the EEG relative power and facial skin temperature data. The average of left and right cheek ROI temperature is used to represent cheek temperature and the average of left and right ear ROI temperature is used to represent ear temperature.

## 4 RESULTS AND DISCUSSIONS

### 4.1 Mental workload index

By comparing several EEG signal indexes, Holm et al. (2009) found the ratio of frontal theta power and parietal alpha power is more



Figure 5: Experiment Protocol

sensitive in reflecting mental workload [14]. Considering the configuration of the 14-channel EEG headset used in this study, the authors propose a mental workload index calculated using Equation (1), where the relative power of each component was calculated using Equation (2). A large mental workload index represents a high mental workload.

$$\begin{aligned}
 \text{Mental workload index} &= \frac{\text{Average frontal theta power}}{\text{Average parietal alpha power}} \quad (1) \\
 &= \frac{F3\text{Theta RP} + F4\text{Theta RP}}{P7\text{Alpha RP} + P8\text{Alpha RP}}
 \end{aligned}$$

$$\text{Relative power(RP)} = \frac{\text{power of certain frequency band}}{\text{all frequency bands power}} \quad (2)$$

Before calculating the index, the authors filter each relative power component to smooth the data and remove the noise. To validate the capability of the EEG mental workload index, a pairwise comparison among the average mental workload during the baseline, easy tasks (i.e. easy level number addition, digit span, choice reaction), and hard tasks (i.e. hard level number addition, digit span, choice reaction) are conducted. The mental workload comparison is shown in Table 2. For the baseline, the average mental workload of three baseline recordings under three thermal conditions are presented. For the easy and hard tasks, we present the average mental workload values of the easy and hard level number addition, digit span, and choice reaction tasks under three thermal conditions, respectively. Each subject had nine data points (3 thermal conditions × 3 easy/hard tasks) for easy and hard level tasks each. According to the result of the one-way repeated measures ANOVA and the Least Significant Difference test ( $\alpha = 0.05$ ), the mental workload index of the hard tasks is significantly higher than that of the easy tasks. Both of the mental workload index during tasks are significantly higher than the baseline. The result proves that the EEG mental workload index could sensitively assess mental workload and be used to examine the validity of facial skin temperature as a potential mental workload measurement method.

Through preliminary observation of the data, the authors find the EEG mental workload data fluctuates during the task period while the facial skin temperature data is relatively stable. One possible reason for this is that EEG is very sensitive to subjects' central neural activity, and thus varies significantly while the subject was reading the question, figuring out the answer, and typing in the result. However, the facial skin temperature does not have rapid and quick responses to subjects' cognitive status as a peripheral physiological measurement. As a result, facial skin temperature could not reflect the real-time mental workload at each timestamp. Nevertheless, the facial skin temperature still has the potential to

Table 2: Mental workload comparison of baseline, easy, and hard tasks for each subject

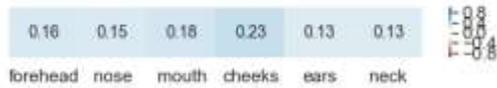
	Baseline	Easy	Hard
Subject 1	0.417	1.305	1.389
Subject 2	0.385	0.940	1.097
Subject 3	0.359	0.965	1.186
Subject 4	0.564	1.052	1.323
Subject 5	0.403	0.980	1.185
Subject 6	0.457	1.240	1.376
Subject 7	0.373	0.796	1.062
Subject 8	0.387	0.737	1.052
Subject 9	0.421	0.836	1.118
Subject 10	0.389	0.955	1.200
Subject 11	0.376	1.620	1.856
Subject 12	0.395	0.995	1.122

disclose the mental workload during a task period (3 minutes - 9.5 minutes), as other related studies found. In this study, we use the average mental workload index of each data segment to represent the mental workload for a subject to perform a specific task and evaluate its relationship with the average facial skin temperature during the task period. Nine tasks of different types and difficulty levels are studied in each thermal condition.

## 4.2 Facial skin temperature and mental workload

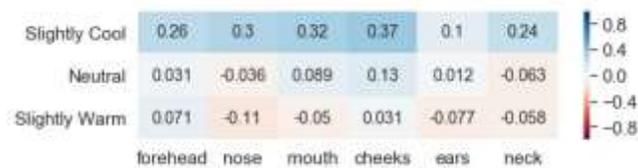
After obtaining each subject's facial skin temperature and the mental workload for each of the tasks, their relationships are studied in three different scales using the Pearson correlation. We first consider the data from all subjects in all the three thermal environments at the same time. The correlation results are shown in Figure 6. The numbers in the cells represent the Pearson correlation coefficient between the corresponding average facial skin temperature and the average mental workload of each task for the 324 (3 environments × 9 tasks × 12 subjects) data points for all subjects. The correlation is found to be poor when considering all thermal environments at the same time. A possible reason is that the facial skin temperature increases significantly with room temperature [24]; however, no significant difference in mental workload is found between the slightly cool and neutral environment. In addition, the change of mental workload in different thermal environments varies significantly within individuals [42]. For example, some subjects demonstrate lower mental workload in the cooler environment, while vice versa for the other subjects. Therefore, rather than having a general model for all thermal conditions, the relationship between facial

thermography and mental workload under each thermal condition should be studied separately.



**Figure 6: Pearson correlation coefficients between facial skin temperature of each region and mental workload (considering all thermal environments and all subjects)**

Next, the correlations between facial skin temperature and mental workload are examined for each thermal condition separately, as shown in Figure 7. The numbers in the cells represent the Pearson correlation coefficient between the corresponding facial skin temperature and mental workload for the 108 (9 tasks × 12 subjects) data points for all subjects in each thermal condition. Even though the facial skin temperature in the slightly cool environment demonstrates higher correlations with mental workload than the other two, the relationships are still weak even if we consider each thermal condition separately.



**Figure 7: Pearson correlation coefficients between facial skin temperature and mental workload (considering all subjects in each thermal environment)**

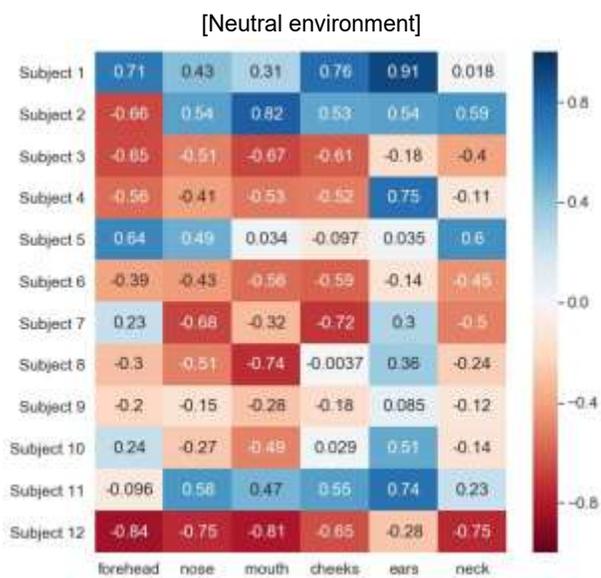
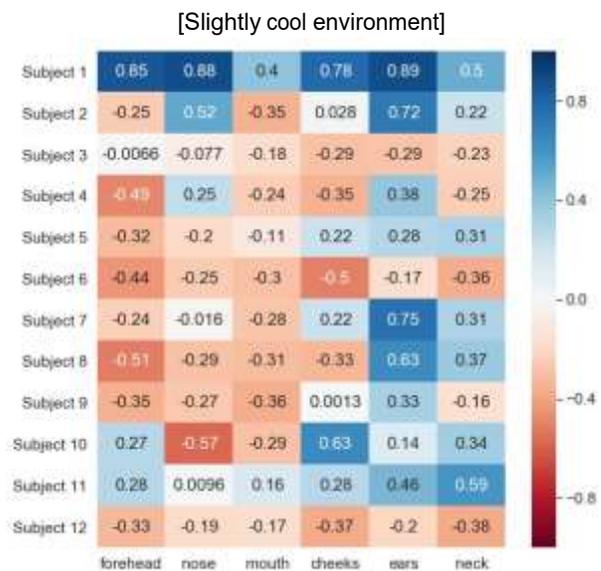
In order to examine whether the poor correlation in each thermal environment is caused by individual differences, the authors evaluate the within-subject correlation between facial skin temperature and mental workload in each thermal condition for the 9 tasks, as shown in Figure 8. It is found that the correlation varies significantly within individuals. While some subjects' facial skin temperature is positively correlated with mental workload (e.g. Subject 11), others' facial skin temperature decreases with mental workload (e.g. Subject 12). Meanwhile, while some subjects' facial skin temperature has relatively strong correlations with mental workload (e.g. Subject 12), the correlations for some subjects are weak (e.g. Subject 5). Even for the same subject, the correlation of a specific facial region could be positive in one environment but negative in another, such as the correlation between neck temperature and mental workload for Subject 2. The individual difference in the relationship between facial thermal patterns and subjective mental workload ratings was also observed by [38]. The neutral environment demonstrated stronger relationships between facial skin temperature and mental workload, possibly because the vasoconstriction and vasodilation responses are weak when the air temperature is cool or warm that attenuates the response of facial temperature to mental workload.

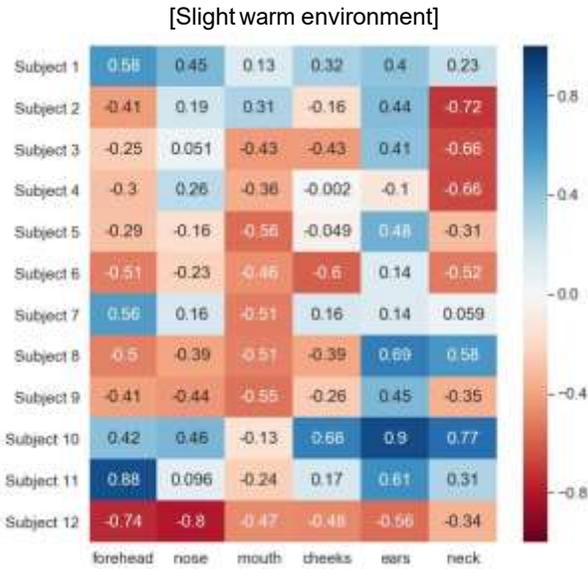
For most subjects, the forehead and mouth temperature is found to be negatively correlated with the mental workload, while the

ear temperature is positively correlated with the mental workload. Despite the fact that several studies found a significant decrease of the nose temperature with an increasing subjective mental workload rating or task difficulty [32], the correlation between nose temperature with EEG-measured mental workload is found to be weak or moderate for most subjects in this study.

### 4.3 Facial skin temperature variation and mental workload

Instead of using the absolute facial skin temperature of each task as shown in Section 4.1, the variation of facial skin temperature is studied in this section. The facial temperature variation is calculated by subtracting the average temperature during the baseline period while the subjects relax from the average temperature of





**Figure 8: Pearson correlation coefficients between facial skin temperature and mental workload in each thermal environment**

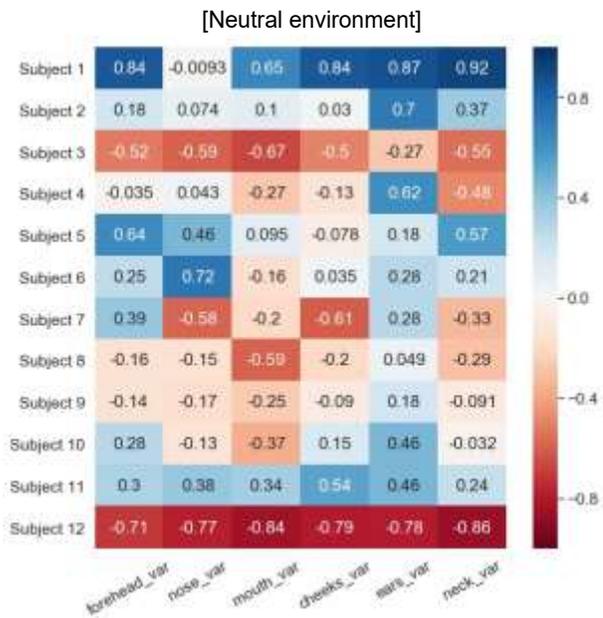
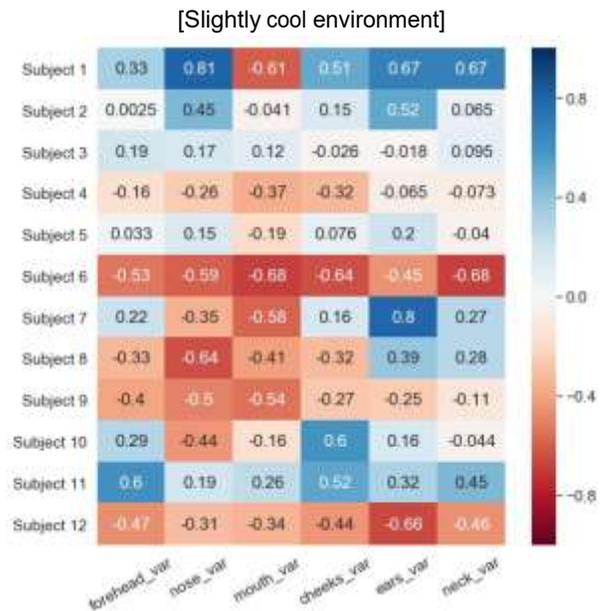
the corresponding ROI during different cognitive task periods in a thermal environment. The correlations between the facial temperature variation and mental workload are examined using Pearson Correlation, as shown in Figure 9. Similar to the absolute facial skin temperature, the correlations between facial temperature variation and mental workload are inconsistent among different subjects. The correlations also vary among different thermal conditions. Despite the correlations for some subjects are relatively strong, such as Subject 1, the correlations between facial skin temperature variation and mental workload are poor in general.

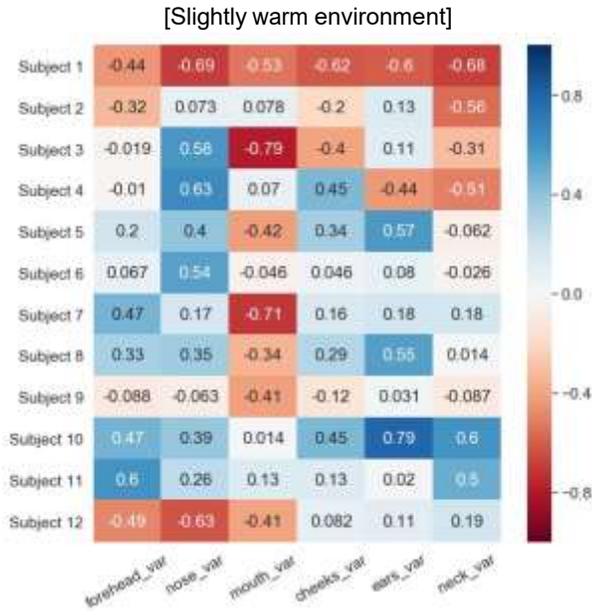
#### 4.4 Mental workload classification using absolute facial skin temperature

From the results of Section 4.2 and Section 4.3, it could be observed that mental workload has stronger correlations with the absolute facial skin temperature than with the facial skin temperature variation. Despite the differences among different subjects, the absolute skin temperature of some facial regions shows moderate or strong correlations with mental workload. Therefore, the feasibility of using the absolute facial skin temperature to disclose mental workload is examined using the Random Forest classifier.

In this study, the mental workload prediction model is trained for each subject in each thermal environment to develop thermal condition-specific models using the skin temperature of six facial regions (see column names in Figure 8). The mental workload is classified into two classes. Mental workload with the value higher than the median workload of the corresponding thermal condition was classified to be high, and vice versa. Considering the individual difference in mental workload, the classification of mental workload is conducted within each subject.

After tuning the model hyper-parameters using grid search, leave-one-out cross-validations are used to evaluate the performance of the classification models, which means the facial temperature of one task is treated as the test data each time and the remaining eight data points are used to train the predictive model. After uncertainty analysis, with the confidence interval of 95%, the average prediction accuracy for all subjects under the slightly cool, neutral, and slightly warm environment is  $45\% \pm 9\%$ ,  $57\% \pm 9\%$ , and  $44\% \pm 9\%$ , respectively, which means the facial skin temperature could not predict the EEG-measured mental workload well.





**Figure 9: Pearson correlation coefficients between facial skin temperature variation and mental workload in each thermal environment**

### 4.5 Discussion

Even though several other studies discovered the relationship between facial skin temperature and subjective-rated mental workload, it is found from the extent of the experiment conducted that facial thermography captured by a low-cost thermal camera could not conclusively disclose EEG-measured mental workload. One possible reason is that EEG measures mental workload through brain cognitive activities; however, subjects' self-perception of their mental workload mainly depends on their perception of their own performance and task difficulty, which leads to the differences between the two measurements.

Meanwhile, several limitations of this study could also affect the result. First, the time duration for the experiment might not be long enough to detect significant changes in skin temperature. Except for the visual search task that lasts 9.5 min, all the other tasks last between 2-6 min. While EEG has a rapid response to mental workload, time lags exist in peripheral physiological measurement such as facial skin temperature to respond to the change of mental workload. In the meantime, the noise of the low-cost thermal camera could not be balanced out thoroughly because only a limited number of data points were collected within the short task duration.

Second, the dataset might not be large enough. Even though numerous time-series mental workload and facial thermography data are obtained for each task, it could not be used to develop prediction model because of the high variation of mental workload and low variation of facial skin temperature within each task period. As a result, an average value is used to represent each task, which greatly limited the number of data points we have for each subject under a thermal environment. The correlation analysis relying on

such a few data points could be greatly affected by outliers. Furthermore, only eight data points were used to train the prediction model each time, which also leads to low prediction accuracy.

Third, the task employed in the experiment is in different type. In order to simulate the state of typical building occupants that would engage in a variety of mental activities, four types of cognitive tasks were involved in the experiment that stimulates subjects' cognitive function on thinking, working memory, reaction, and perception. However, it was found that different types of tasks result in various facial thermography patterns [33], which makes it difficult to explore the relationship or make predictions with data combined by different types of tasks.

The limitations found in this study provide us with some directions for future work in this field, which could potentially improve the prediction result. First, the duration of each task could be extended to produce more significant physiological responses. It also allows for an increasing repetitiveness for each type of task such that more data could be collected for correlation analysis and for training the prediction model. Second, increasing the task difficulty to stimulate higher mental workload on subjects could also result in more significant responses on facial skin temperature, which would potentially lead to higher prediction accuracy. Third, we could study the information loss of the low-cost thermal camera compared to the delicate ones that prevents it from accurately reflect mental workload. It would provide us with potential directions on model improvement.

### 5 CONCLUSIONS

A low-cost and non-intrusive method to understand occupants' mental workload offers an ideal opportunity for developing an optimal indoor environment control technique that maximizes its occupants' productivity, health, and quality of life. In this study, the correlation of facial skin temperature and its variation with the EEG-measured mental workload is examined in three different thermal environments. The mental workload fluctuates more widely as tasks go on, while the facial skin temperature is relatively stable. Therefore, the relationships between average facial skin temperature and the average mental workload during each task are studied. The relationships are found to be different with thermal conditions and individuals. In general, the absolute facial temperature has stronger correlations with mental workload than facial temperature variation. Compared to the slightly cool and slightly warm environments, relatively stronger correlations between absolute facial skin temperature and mental workload are found in the neutral environment. The results suggest increasing the repetitiveness of each type of task in future studies since task type could affect the response of facial skin temperature to mental workload. Extending task duration or using more challenging tasks in the experiment could also lead to more insights on the relationship between facial thermography and mental workload, so that non-intrusive data collection methods such as low-cost thermal cameras can be used to simultaneously assess the mental workload of all occupants in multi-occupancy building environments.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support received from the U.S. National Science Foundation (NSF) CBET 1349921 and CBET 1804321. Any opinions and findings in this paper are those of the authors and do not necessarily represent those of the NSF.

## 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] SD Brennan. 1992. An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. *Portsm. DRA Marit. Command Control Div. DRA Tech. Memo. CAD5 92017* (1992).
- [3] Dick De Waard. 1996. *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands.
- [4] Julia A East. 2000. *Feature selection for predicting pilot mental workload*. Technical Report. AIR FORCE INST OF TECH WRIGHT-PATTERSONAFB OH SCHOOL OF ENGINEERING.
- [5] EMOTIV. 2019. Emotiv EPOC+. <https://www.emotiv.com/epoc>
- [6] Poul O Fanger et al. 1970. Thermal comfort. Analysis and applications in environmental engineering. *Thermal comfort. Analysis and applications in environmental engineering*. (1970).
- [7] William Fisk and Olli Seppanen. 2007. Providing better indoor environmental quality brings economic benefits. (2007).
- [8] William J Fisk. 2000. Health and productivity gains from better indoor environments and their relationship with building energy efficiency. *Annual review of energy and the environment* 25, 1 (2000), 537–566.
- [9] Lisa R Fournier, Glenn F Wilson, and Carolyne R Swain. 1999. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology* 31, 2 (1999), 129–145.
- [10] Thomas C Hankins and Glenn F Wilson. 1998. 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.
- [11] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [12] G Robert J Hockey. 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology* 45, 1-3 (1997), 73–93.
- [13] Chris Hocking, Richard B Silberstein, Wai Man Lau, Con Stough, and Warren Roberts. 2001. Evaluation of cognitive performance in the heat by functional brain imaging and psychometric testing. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 128, 4 (2001), 719–734.
- [14] Anu Holm, Kristian Lukander, Jussi Korpela, Mikael Sallinen, and Kiti MI Müller. 2009. Estimating brain load from the EEG. *The Scientific World Journal* 9 (2009), 639–651.
- [15] Houtan Jebelli, Sungjoo Hwang, and SangHyun Lee. 2017. EEG signal-processing framework to obtain high-quality brain waves from an off-the-shelf wearable EEG device. *Journal of Computing in Civil Engineering* 32, 1 (2017), 04017070.
- [16] Houtan Jebelli, Sungjoo Hwang, and SangHyun Lee. 2018. EEG-based workers' stress recognition at construction sites. *Automation in Construction* 93 (2018), 315–324.
- [17] P.g.a.m. Jorna. 1992. Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology* 34, 2-3 (1992), 237–257.
- [18] Shengxian Kang, Dayi Ou, and Cheuk Ming Mak. 2017. The impact of indoor environmental quality on work productivity in university open-plan research offices. *Building and Environment* 124 (2017), 78–89.
- [19] Caroline Karmann, Stefano Schiavon, and Edward Arens. 2018. Percentage of commercial buildings showing at least 80% occupant satisfied with their thermal comfort. (2018).
- [20] Ying Lean and Fu Shan. 2012. Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries* 22, 3 (2012), 177–187.
- [21] JA Leech, R Burnett, W Nelson, SD Aaron, and M Raizenne. 2000. Outdoor air pollution epidemiologic studies. *American Journal of Respiration and Critical Care Medicine* 161, 3 (2000), A308.
- [22] Da Li, Carol C Menassa, and Vineet R Kamat. 2017. Personalized human comfort in indoor building environments under diverse conditioning modes. *Building and Environment* 126 (2017), 304–317.
- [23] Da Li, Carol C Menassa, and Vineet R Kamat. 2017. A personalized HVAC control smartphone application framework for improved human health and well-being. *Computing in Civil Engineering* 2017 (2017).
- [24] Da Li, Carol C Menassa, and Vineet R Kamat. 2018. Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography. *Energy and Buildings* 176 (2018), 246–261.
- [25] Da Li, Carol C Menassa, and Vineet R Kamat. 2019. Feasibility of Low-Cost Infrared Thermal Imaging to Assess Occupants' Thermal Comfort. In *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience*. American Society of Civil Engineers Reston, VA, 58–65.
- [26] Da Li, Carol C Menassa, and Vineet R Kamat. 2019. Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras. *Applied Energy* 251 (2019), 113336.
- [27] Adrian Cornelius Marinescu, Sarah Sharples, Alastair Campbell Ritchie, Tomas Sánchez López, Michael McDowell, and Hervé P Morvan. 2018. Physiological parameter response to variation of mental workload. *Human factors* 60, 1 (2018), 31–56.
- [28] Sarah Miller. 2001. *Workload measures. National Advanced Driving Simulator. Iowa City, United States* (2001).
- [29] Ernst Niedermeyer and FH Lopes da Silva. 2005. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
- [30] Department of Energy (DOE). 2017. Heating and Cooling. <https://energy.gov/public-services/homes/heating-cooling>
- [31] U.S. Department of Health and human services. 2010. NASA Task Load Index. <https://healthit.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index>
- [32] 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.
- [33] Ioannis Pavlidis, James Levine, and Paulette Baukol. 2000. Thermal imaging for anxiety detection. In *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No. PR00640)*. IEEE, 104–109.
- [34] NIA Rahman, SZM Dawal, and N Yusoff. 2017. Subjective responses of mental workload during real time driving: A pilot field study. In *IOP Conference Series: Materials Science and Engineering*, Vol. 210. IOP Publishing, 012076.
- [35] Miguel A Recarte and Luis M Nunes. 2003. Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of experimental psychology: Applied* 9, 2 (2003), 119.
- [36] Gary B Reid and Thomas E Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*. Vol. 52. Elsevier, 185–218.
- [37] Olli A Seppänen and William Fisk. 2006. Some quantitative relations between indoor environmental quality and work performance or health. *Hvac&R Research* 12, 4 (2006), 957–973.
- [38] John Stemberger, Robert S Allison, and Thomas Schnell. 2010. Thermal imaging as a way to classify cognitive workload. In *2010 Canadian Conference on Computer and Robot Vision*. IEEE, 231–238.
- [39] JA Veltman and AWK Gaillard. 1996. Physiological indices of workload in a simulated flight task. *Biological psychology* 42, 3 (1996), 323–342.
- [40] JA Veltman and AWK Gaillard. 1998. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* 41, 5 (1998), 656–669.
- [41] Paul Viola, Michael Jones, et al. 2001. Rapid object detection using a boosted cascade of simple features. *CVPR (1)* 1 (2001), 511–518.
- [42] Xi Wang, Da Li, Carol C Menassa, and Vineet R Kamat. 2019. Investigating the effect of indoor thermal environment on occupants' mental workload and task performance using electroencephalogram. *Building and Environment* 158 (2019), 120–132.
- [43] Xi Wang, Da Li, Carol C Menassa, and Vineet R Kamat. 2019. Investigating the neurophysiological effect of thermal environment on individuals' performance using electroencephalogram. In *International Conference on Computing in Civil Engineering (i3CE): Smart Cities, Sustainability, and Resilience*.
- [44] Walter W Wierwille and John G Casali. 1983. A validated rating scale for global mental workload measurement applications. In *Proceedings of the Human Factors Society Annual Meeting*, Vol. 27. Sage Publications Sage CA: Los Angeles, CA, 129–133.
- [45] Glenn F Wilson. 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology* 12, 1 (2002), 3–18.
- [46] Thomas Witterseh, David P Wyon, and Geo Clausen. 2004. The effects of moderate heat stress and open-plan office noise distraction on SBS symptoms and on the performance of office work. *Indoor air* 14 (2004), 30–40.
- [47] Chunlin Zhao, Chongxun Zheng, Min Zhao, Yaling Tu, and Jianping Liu. 2011. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Systems with Applications* 38, 3 (2011), 1859–1865.