Classifying Attention Types with Thermal Imaging and Eye Tracking

YOMNA ABDELRAHMAN*, Bundeswehr University Munich, Germany
ANAM AHMAD KHAN*, The University of Melbourne, Australia
JOSHUA NEWN*, The University of Melbourne, Australia
EDUARDO VELLOSO, The University of Melbourne, Australia
SHERINE ASHRAF SAFWAT, German University in Cairo, Egypt
JAMES BAILEY, The University of Melbourne, Australia
ANDREAS BULLING, University of Stuttgart, Germany
FRANK VETERE, The University of Melbourne, Australia
ALBRECHT SCHMIDT, Ludwig Maximilian University Munich, Germany

Despite the importance of attention in user performance, current methods for attention classification do not allow to discriminate between different attention types. We propose a novel method that combines thermal imaging and eye tracking to unobtrusively classify four types of attention: *sustained*, *alternating*, *selective*, and *divided*. We collected a data set in which we stimulate these four attention types in a user study ($N = 22$) using combinations of audio and visual stimuli while measuring users' facial temperature and eye movement. Using a Logistic Regression on features extracted from both sensing technologies, we can classify the four attention types with high AUC scores up to 75.7% for the user independent-condition independent, 87% for the user-independent-condition dependent, and 77.4% for the user-dependent prediction. Our findings not only demonstrate the potential of thermal imaging and eye tracking for unobtrusive classification of different attention types but also pave the way for novel applications for attentive user interfaces and attention-aware computing.

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

Additional Key Words and Phrases: Eye Tracking, Thermal Imaging, Attention Classification, Attention Types

ACM Reference Format:
Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 69 (September 2019), 27 pages. https://doi.org/10.1145/3351227

*Authors contributed equally to this research.

Authors' addresses: Yomna Abdelrahman, yomna.eldin@gmail.com, Bundeswehr University Munich, Munich, Germany; Anam Ahmad Khan, anam.khan@unimelb.edu.au, Computing and Information Systems, The University of Melbourne, Melbourne, Australia; Joshua Newn, joshua.newn@unimelb.edu.au, Computing and Information Systems, The University of Melbourne, Melbourne, Australia; Eduardo Velloso, Computing and Information Systems, The University of Melbourne, Melbourne, Australia; Sherine Ashraf Safwat, German University in Cairo, Cairo, Egypt; James Bailey, Computing and Information Systems, The University of Melbourne, Melbourne, Australia; Andreas Bulling, Institute for Visualisation and Interactive Systems, University of Stuttgart, Stuttgart, Germany; Frank Vetere, Computing and Information Systems, The University of Melbourne, Melbourne, Australia; Albrecht Schmidt, Ludwig Maximilian University Munich, Munich, Germany.

2474-9567/2019/9-ART69 $15.00
https://doi.org/10.1145/3351227

1 INTRODUCTION

A common, though often incorrect, hidden assumption underlying how we currently design interactive systems is that, during the interaction, the user focuses all of their attention on the interaction with the system. As a consequence, considerable effort in the research and development of ubiquitous computing systems has been placed on supporting users while they perform single-focus tasks (e.g., [16, 21]). However, given the multitude of devices and applications constantly fighting for users' attention through interruptions and notifications, single-focus interactions are the exception rather than the rule [69, 70].

This phenomenon has led economists to frame the problem in terms of an *attention economy*, where attentional resources are the currency and actors are competing for consumers' attention [11, 18]. A fundamental concept in this idea is that, similar to other economic resources, *attention is a limited resource*. Further, attention is strongly influenced both by *internal stimuli* (e.g. remembering where you left your keys causes your attention to shift, or feeling motivated to read a book leads to a more focused reading experience) and *external stimuli* (e.g. hearing a dog bark behind you causes you to turn around or writing an essay while being observed by your teacher keeps your eyes on the exercise at hand). Therefore, the context around the interface affects how much attention is paid to the interaction.

To further complicate matters, attention itself is a complex concept, one that even psychologists struggle to conceptualize [64]. Early studies suggested that there are several levels of attention instead of a unitary one, due to its complex nature involving memory, behavior and consciousness [30, 41, 43, 51, 63]. One model that emerged from this literature is Sohlberg and Mateer's *Clinical Model of Attention* [59]. This hierarchical model discriminates between people's ability to maintain attention towards a single stimulus (*sustained* and *focused* attention); to switch attention between different stimuli (*alternating* attention); to pay attention to one stimulus while inhibiting others (*selective* attention), and to pay attention to multiple stimuli simultaneously (*divided* attention). This model highlights two challenges: quantifying attention (how much attention) and qualifying the nature of attention (what type of attention).

Prior work on attention has shown that our well-being is tied strongly to our ability to manage our attention successfully, for example, we know that multitasking hinders performance [36]. Such known issues create opportunities to design interactive systems that monitor and actively help users to manage their attention. The vision of *pervasive attentive user interfaces* encapsulates this well, stating that interfaces could "adapt the amount and type of information based on [users'] current attentional capacity, thereby simultaneously optimizing for information throughput and subtlety. Future interfaces could trade-off information importance with users' current interruptibility level and time the delivery of information appropriately" [8].

To realize this vision, interfaces that attempt to leverage the users' attention must accomplish two tasks: (1) *identify* the locus of attention and (2) *characterize* the nature of the current attentional state. While the locus of attention is typically considered to be equivalent to gaze direction, this is not always the case due to the diverse nature of attention orienting, which is classified as overt or covert [72]. In overt attention, the person selectively attends to a source of information by moving their eyes to point in that direction [50]. However, humans do not necessarily direct their eyes towards their area of focus. During covert attention, a corresponding shift in attention is not followed by a corresponding shift in gaze direction [17], e.g., when a person has a conversation with a friend while looking at their mobile phone [41], or when eavesdropping on a conversation while typing up an email in an open office environment. Therefore, even though eye-tracking data can be very informative, it is essential to understand the limits of the gaze point as a sole representation of the locus of attention.

In this paper, we address the limits of eye-tracking for attention detection by proposing its combination with thermal imaging in order to classify the various attention types by stated in Sohlberg and Matter's *Clinical Model of Attention*. The Clinical Model describes attention as a model based on the degree of focus, consisting of lower fundamental levels and higher levels [59]. The lower level includes *focused* and *sustained* attention, while the higher levels includes *selective*, *alternating*, and *divided* attention [30, 41, 59]. The Clinical Model further describes attention as a multidimensional cognitive capacity, which means that attentive tasks need different levels of cognitive load to be achieved [41]. The findings of recent work in HCI demonstrated the ability to unobtrusively quantify cognitive load using thermal imaging and temperature sensors [2, 65, 75]. Our paper, therefore, builds on the ability to use thermal input as a method of measuring different levels of cognitive loads and the knowledge that these different attention types require different cognitive capacities. By combining this concept with the ability to detect overt attention reliably well through eye-tracking, we explore the novel combination of eye-tracking and thermal imaging for attention classification. To explore this combination, we collected a dataset in a user study designed to elicit different attention types using different stimulus modalities, in controlled and (semi-)naturalistic tasks. We build on the opportunity that eye-tracking can reveal the locus of attention, and thermal imaging can give us an estimate of cognitive load. Together, this allows us to paint a better picture of users' attentional state. We hypothesize that by combining these modalities, we are able to classify different attention types according to the Clinical Model. We present our results regarding the applicability of our approach to differentiate attention types and suggest directions for use cases and future steps towards attention-aware systems. Specifically, in this paper, we contribute the following:

- We propose a novel approach for classifying attention types by monitoring users' gaze and facial temperature.
- We propose a set of features used for classification, along with a variety of classifiers that researchers could adopt to build systems that can differentiate between the attention types.
- We collected empirical data from 22 participants while exposed to various stimuli eliciting four attention types. We evaluated the performance of the built classifiers using the collected data. Our approach classified attention types based on thermal and gaze features and produced AUC up to 87% and an average of 80.3% for all classifiers.

2 RELATED WORK

In the past decades, many scientific fields have been interested in understanding the processes behind human attention, from its measurement to its modeling. A pre-condition for this is the ability to sense and characterize attentional states in near real-time, and prior work has explored the use of various sensors and algorithms in attempts to achieve this. In this section, we discuss the background theory of attention, technology-based approaches for sensing attention, and existing algorithms for classifying attention.

2.1 Models of Attention

The vast body of research on theories of attention can be split loosely into theories of focused attention and theories of divided attention, with few studies attempting to bridge the gap between the two (e.g., [49]). Whereas theories of focused attention are grounded on visual selection and unintentional processing, theories of divided attention revolve around the issue of capacity limits [19, 29]. These differences in theoretical grounding have led to the evolution of different attention models in the field of psychology. In our work, we employ Sohlberg and Mateer's *Clinical Model of Attention* [59] as it has been deemed to be one of the most comprehensive models [4]. The Clinical Model describes attention as a model based on the degree of focus, consisting of lower fundamental levels and higher levels [59]. The lower level includes *focused* and *sustained* attention, while the higher levels includes *selective*, *alternating*, and *divided* attention [30, 41, 59]. In other words, attentive tasks need different levels of cognitive load to be achieved [41]. The attention types introduced in the model are:

Focused attention. The ability to respond discretely to specific visual, auditory, or tactile stimuli.

Sustained Attention. The brain can discretely respond to specific auditory, tactile, or visual stimuli for a prolonged period. Reading a book in a deeply focused state is an example of sustained attention.

Alternating Attention. Happens when we switch focus from one task to another or from one part of the task to another, regardless of different cognitive demands between them. Examples include: listening to a lecture while taking notes, or reading a recipe while cooking.

Selective Attention. The ability of the brain to focus on a specific stimulus while inhibiting others. A prime example of selective attention is called the *Cocktail Party Effect* [33], which describes our ability to selectively attend to the voice of one person while minimizing other voices and noise.

Divided Attention. The brain divides its attention between different stimuli simultaneously. Examples include: playing a mobile game while having a conversation or, one that we do not recommend, texting while driving.

2.2 Current Approaches to Classify Attention

A crucial step in building attentive systems lies in the ability to quantify users' attentional states. However, as changes in these states happen inside users' minds, we can only measure attention indirectly through users' behaviors and physiological signals, leading to the development of technologies potentially offering insights about the users' attention states. These technologies vary in their levels of obtrusiveness.

Previous works have explored a variety of sensors for measuring attentional states, including electroencephalography (EEG) [1, 32, 37, 39], electrooculography (EOG) [32], electrocardiography (ECG) [9], and electromyography (EMG) [52]. These sensors have been shown to provide high accuracy in recognizing states but are obtrusive in nature (requiring users to wear a device or have electrodes attached to their skin), and therefore cumbersome for daily use. For instance, Liu et al. [39] were able to distinguish between attentive and inattentive states with an accuracy of 76.82% using EEG but required the placement of electrodes on participants' heads. On the other hand, researchers have employed less unobtrusive approaches such as functional Magnetic Resonance Imaging (fMRI), commonly used to reveal aberrant brain activity, to measure attentional states [26, 40, 45]. For example, Moisala et al. [45] measured human brain activity during single-tasking and dual-tasking using fMRI, looking for activation in the medial and lateral frontal regions of the brain. Their results highlight the relationship between different attentional demands and levels of brain activity associated with sustained and divided attention. Though able to show differences in attention states, fMRI remains impractical for daily use, in terms of costs and practicality.

Recent work has explored unobtrusive contactless sensing approaches, including eye tracking and temperature sensors. Eye tracking is a common technique to investigate visual attention as we tend to fixate on objects that have drawn our attention or relevant to the task that we are attending to [41, 47, 48]. Through our visual attention, we only 'see' what we are paying attention to, as our cognitive system allocates sufficient resources for visual processing to avoid overloading. Because we receive a large amount of information through our eyes, this mechanism helps us to manage what gets processed. Eye movements are an important part of visual attention and are primarily comprised of fixations (stationary phase) and saccades (rapid, ballistic eye movements phase). Previous works have long explored how eye movement features can help uncover psychological states and recognize activities [60, 68]. Eye tracking is a powerful tool for understanding human attention as it can measure both the frequency of eye movements and the location of the gaze point [5, 6, 14, 20, 41]. While researchers often use gaze point as a proxy for the locus of attention, this is not always the case due to the diverse nature of attention orienting—classified as overt or covert [72]. Therefore, even though eye tracking data can be very informative, it is essential to understand the limits of the gaze point as a sole representation of the locus of attention.

Thermal imaging and temperature sensors have been explored as a means of understanding users' mental states [2, 27, 35, 53, 65], for instance, thermal imaging has been used to detect several states including stress, guilt, fear [28]. Our work, however, builds specifically on Abdelrahman et al.'s *Cognitive Heat* [2] and Zhou et al.'s *Cognitive Aid* [75], which demonstrate the relationship between facial temperature and cognitive load estimation, in which the authors employ the use of thermal imaging as a way to unobtrusively detect changes in cognitive load in real-time. To elaborate, the authors found substantial changes in facial temperatures upon the activation of the ANS when exposed to the stimulus, specifically between the nose and forehead regions. This seminal work gave rise to developing thermal-based activity tracking, which further facilitates new applications in the field of cognition-aware computing. Wearable variations have also been developed using the same concept. For example, Tag et al. [65] presented early work on the use of facial temperature to measure attention; demonstrating the ability to measure attention using IR temperature sensors. However, their focus was attention level rather than type. Similarly, Zhou et al. [75] explored the use of thermal sensors to detect mental workload, demonstrating the ability of such sensors to detect when a user is currently performing a task.

Table 1. Summary of existing Work on attention state classification.

Sensors: Features	Classifier	Attention States	Accuracy
Software Interaction: Haar wavelets On/off task entries of participants [44]	Linear SVM	Primarily attentive Primarily attentive-short inattentive periods Primarily inattentive with short attentive Primarily inattentive	82.8%
Eye tracking and inertial sensor: Iris movement and head rotation [10]	G(GA)-SVM	Attentive Non-attentive	93.10%
Kinect: Head angle and displacement, body lean, Face deformation and 2D gaze points [74]	Decision Trees	High attention Medium attention Low attention	75.30%

The variety of sensors discussed above opened the opportunity to use machine learning techniques to classify users' mental states and to build systems that adapt to these states [56]. Table 1 summarizes recent work on attention classification through machine learning, showing that cognitive ability can be estimated by measuring the attention level of users while performing activities. Whereas there have been initial efforts to use machine learning to classify attention primarily into attentive and non-attentive states with a maximum accuracy of 93.10%, no work has attempted to classify attention according to the four types of attention outlined in Sohlberg and Mateer's Clinical Model of Attention [59].

2.3 Summary & Research Direction

In summary, there are two clear limitations from the existing literature on recognizing attention types. First, the sensors employed for measuring attention tend to be obtrusive and therefore not appropriate for the development of interactive systems. Second, works to date employed models that oversimplify attentional processes, as a binary variable or as a one-dimension continuous signal.

In this paper, we address this research gap by using the combination of two unobtrusive sensors—*thermal imaging* and *remote eye tracking*, from which we can build classifiers for recognizing the four types of attention outlined in Sohlberg and Mateer's *Clinical Model of Attention* [59]. To our knowledge, this is the first work that has attempted to differentiate between four attention types (*sustained*, *alternating*, *selective*, *divided*). Our combined approach exploits the fact that each attention type requires different cognitive resources [41] and visual direction [20].

To elaborate, our novel approach leverage two ways in which attentive states are manifested in our physiology to measure and classify attention effectively. First, attention is related to the allocation of cognitive resources [41]. This process strongly correlates with changes in the blood flow, which is reflected in changes in the temperature distribution in our skin [2, 57, 58]. In a seminal work, Abdelrahman et al. [2] explored the use of thermal imaging to measure cognitive load, in which the authors relied on how the activation in the Autonomic Nervous Systems (ANS) due to an increase in cognitive load is reflected in the facial temperature. Using the same ideology, we hypothesize that changes in attentive states will also lead to a change in cognitive load levels that are observable in facial temperature patterns measured with a thermal camera. Our hypothesis is built upon the fact that different attentive states require different levels of cognitive load [41]. Informed by the literature, we estimate cognitive load using the nose-forehead differential [2, 28]. Also, we explored the effect of different attentive states on the user's cheeks, as previous work [28] highlighted the usage of cheeks as state indicator. Second, when engaging in overt attention, the gaze point—which we can easily measure with eye-tracking—is a reasonable estimate of the locus of attention [20]. Further, low-level statistical features of eye movements are also indicative of cognitive load levels, which can be useful for an attention classifier (e.g. [73]).

These physiological properties present an opportunity for the design of pervasive attentive user interfaces. Both eye movements and facial temperature patterns can be unobtrusively captured with remote eye trackers and thermal imaging cameras, particularly considering that the face is the most often exposed part of the user's body. Moreover, recent advances in both eye-tracking and thermal imaging have made it cheaper and more accessible than ever to capture this information without the need to augment the user, but rather the environment.

In the following sections, we present the data collection with a detailed description of the tasks used for attention elicitation (Section 3.1). We hypothesize that the higher-level attention types (selective and divided) will result in a more significant temperature difference. To explore this hypothesis, we conducted a user study to elicit attention types using the combination of audio and video stimuli, while recording the gaze and thermal data (Section 3.2). We then present our methodological approach to analyze the collected data set, including statistical analysis, feature extraction, and classification. In Section 5, we report the results from different classification approaches (user-dependent, user-independent (condition dependent) and user-independent (condition independent), showing the applicability of thermal cameras and eye tracker as unobtrusive sensors to classify attention. Lastly, we discuss how the findings of our work can be applied and present directions in the future work section.

3 DATA COLLECTION

Our goal in this paper is to build a classifier that is able to distinguish between attention types based on facial thermal imaging and eye-tracking data. To train and evaluate this classifier, we collected a dataset in which we recorded the eye movements and the temperature of facial features (nose, forehead, left cheek, and right cheek) of 22 participants as they completed tasks designed to elicit four types of attentional states. The tasks were inspired by the literature on *attention* in psychology. We used a repeated-measures design, where all participants performed four sets of tasks with different stimuli. We counterbalanced the order of the tasks. We created variations of each stimulus to elicit four types of attention in the *Clinical Model of Attention — sustained*, *selective*, *alternating* and *divided*, for a total of 16 tasks (4 attention types × 4 types of stimuli). We did not include focused attention, as we are interested in the attention over prolonged periods. Further, we included a baseline task (detailed in Section 3.3) at the beginning of the experiment. All content used during the study was in English, and all participants recruited were proficient in the English language.

3.1 Tasks

We used a combination of tasks from the attention elicitation literature and developed a series of tasks to elicit different attention types starting with Stroop conditions as a reference task, followed by more naturalistic tasks that involved a combination of visual- and audio-based stimuli. For the baseline task, we asked the participants to relax while listening to white noise. We used the baseline task to capture and record the participants' temperatures at rest, which serves as a point of comparison with the other tasks [2]. Figure 1 illustrates the remaining tasks used in the study. We published a playlist with the stimuli online[1], for reproducibility purposes. We displayed the tasks in full screen for 3 minutes, and conditions without a visual stimulus contained a white background. For consistency, we primarily used selected TED Talks[2] for the content of the tasks. In audio-based tasks, we extracted the audio from the videos, while in the visual tasks, we used the transcripts of the talks.

[1] https://bit.ly/2LyZWay
[2] https://www.ted.com/talks

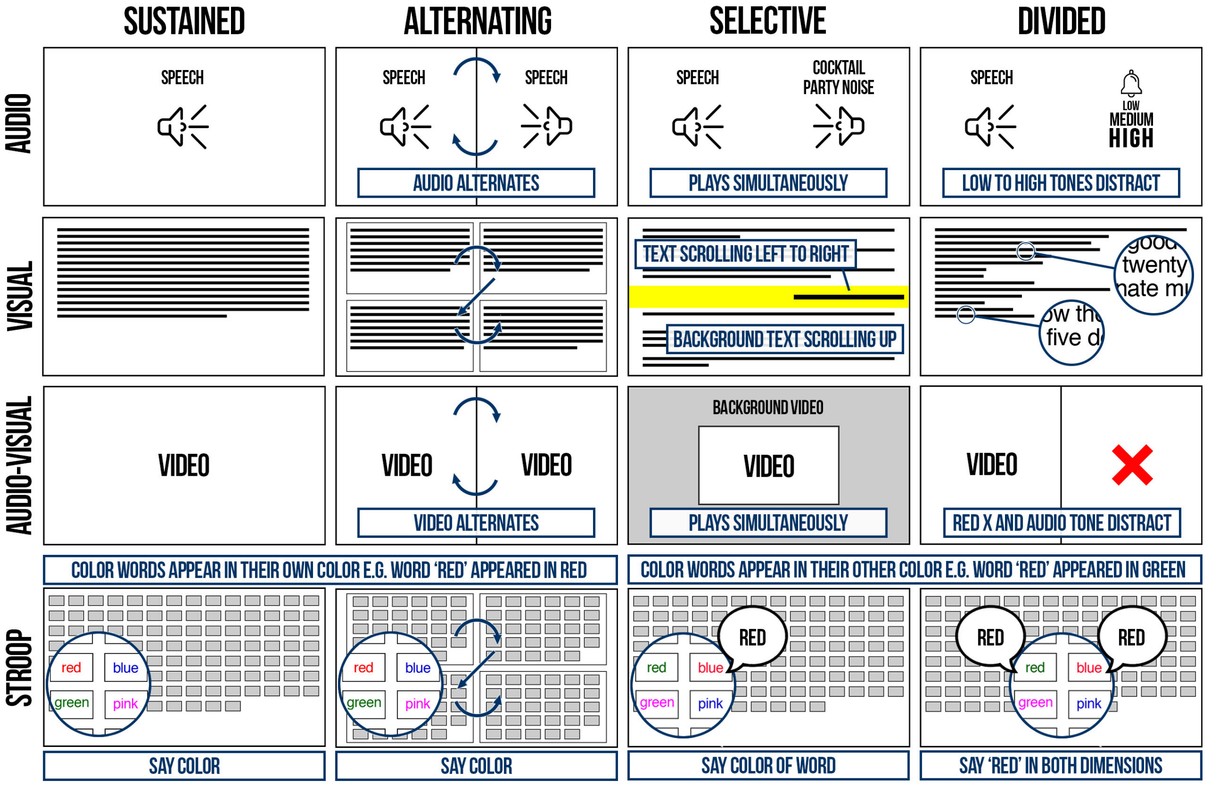

Fig. 1. Illustration of the (conditions) visual, audio-visual and stroop stimuli used to stimulate the four different attention types.

Stroop Tasks

The Stroop test is a classic Psychology task for eliciting selective attention [34, 62]. In the typical test, users are asked to name the color of the font in which words are written. The difficulty of the task lies in the fact that the words displayed correspond to a different color to the one in which they are colored while the user *selectively attends* to the color of the font. For example, in the classic experimental task, the word 'RED' would be colored in blue, and the participant must reply 'Blue' while ignoring the fact that the word itself corresponds to a different color. For our study, we created three variations of the Stroop test to elicit the remaining attention types, described below:

Sustained Stroop: We first created a simplified variation of the Stroop test to elicit sustained attention, where we retained a single source of information. We showed color names written in their own color and asked participants to read it aloud. For example, the word 'green' would appear colored in green, and participants were asked to say 'green'. This effectively removed the challenge of the task allowing the participant to focus on reading the words, therefore maintaining sustained attention.

Alternating Stroop: In this variation of the Stroop test, the display was split into two halves. Each half had a sustained Stroop test variation, and participants had to alternate between the two halves, spending 45 seconds in each half.

Selective Stroop: We used the original Stroop test [62] as the selective Stroop test where text and color are presented differently. For example, the word 'green' would appear colored in blue, and the participant had to say 'blue.' Participants, therefore, have to 'selectively' choose between the two.

Divided Stroop: We reused the Stroop test variation introduced by Eidels et al. [15] to elicit divided attention where participants are directed to attend to both word and color. The task included all four combinations of the words, RED and GREEN, and the ink colors, red and green. The participants are asked to respond to 'redness' in the Stroop stimulus, regardless of whether the 'redness' comes in the word (RED), the color (red), or both (RED in red). Hence, the participant must attend to the color and to the word (i.e. divide attention across the Stroop stimulus components).

Audio Tasks

Sustained Audio: This task used a single audio file of a TED talk speech to which participants were asked to listen attentively.

Alternating Audio: To simulate a group conversation, we used two audio sources, which alternated between being on and off every 45 sec. We used the same topic to mimic the real-life example of a group conversation.

Selective Audio: We simulated the *Cocktail Party effect* [33], where we combined audio of a speaker with the audio of a cocktail party. Participants were asked to attend to the speaker selectively.

Divided Audio: Inspired by Gardiner et al. [22] to elicit divided attention, participants were asked to listen to a presentation talk while listening to and reporting high, low, or medium tone sequences by saying the tone level out loud.

Visual Tasks

Sustained Visual: A single panel of text was displayed, and participants were asked to read the text as it appeared.

Alternating Visual: We divided the screen into two panels, which we further subdivided into two parts. The text first appeared in the left panel for 45 sec and then alternated to the right panel, the text then alternated back to the bottom half of the first panel and lastly alternated to the bottom half of right panel. Participants are asked to read the text displayed in the active panel.

Selective Visual: A stream of text was displayed in a highlighted region in the middle section of the screen from left to right. A stream of text also flowed upwards in the background. Participants were asked to read the text in the highlighted region selectively. This task was inspired by a news ticker (also called slide) that typically appears at the bottom of TV channels.

Divided Visual: In this task, we augmented numbers into a text transcript. We asked participants to read the text while performing mental addition on the numeric values that appeared in the text, e.g. twenty, five, etc. This forced the participants to divide their attention between the text itself and the mental arithmetic task.

Audio-Visual Tasks

Sustained Audio-Visual: For this task, we had a single video running from a selected top TED talk.

Alternating Audio-Visual: Similar to the Alternating Visual task, the screen was divided into two panels and two videos played alternatively in the two panels. The first video played for 45 seconds and alternated to the second panel. This alternating process repeats twice.

Selective Audio-Visual: Two videos were displayed, one embedded in the other, as shown in Figure 1. Participants were asked to selectively attend to the video with the talk that was displayed in the middle of the screen. The larger video acted as a cocktail party like noise [22].

Divided Audio-Visual: Inspired by Gardiner et al. [22] to elicit divided attention and similar to the Divided Audio task, participants were asked to watch a video while listening to and reporting high, low, or medium tone sequences. Additionally, we added an appearing "X" in Red, and the user was asked to say "X" out loud when the symbol appears to elicit attention on divided audio-visual type stimuli.

3.2 Experimental Setup

Figure 2-Left illustrates our experimental setup, consisting of a commercial Tobii EyeX eye tracker[3] operating with frequency of approximately 55 Hz, connected via USB. The eye tracker provided the gaze x- and y-coordinates on

[3] https://tobiigaming.com (recent firmware upgrade enabled increased frequency to 70Hz)

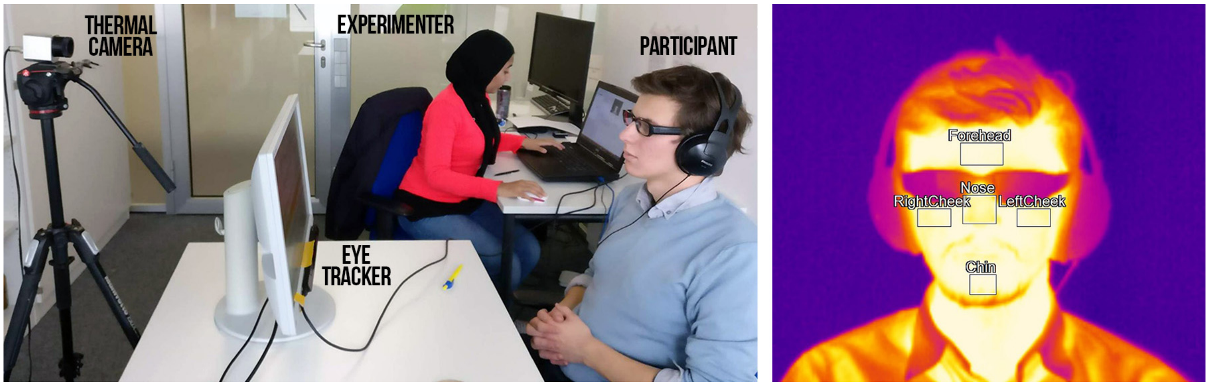

Fig. 2. Left: Experimental setup consisting of thermal camera and eye tracker facing the participant. Right: The facial features being monitored by the thermal camera (forehead, nose, left cheek, right cheek and chin).

the screen. We attached the eye tracker to a 24" screen and placed an Optris PI450 thermal camera[4] mounted on a tripod 1m away from the participant behind the screen. The camera has an optical resolution of 382×288 pixels, has a frame rate of 80 Hz, and measures temperatures between -20°C and 900°C, with a thermal sensitivity of 0.04°C. Further, the camera captured wavelengths in the spectral range between 7.5μm and 13μm with a 38°× 29° field of view. The output of the camera encodes temperature information with 16-bit color values. Further, we developed a system to display the stimuli (tasks) for each test in a counterbalanced order using Latin square that records both streams of data. The Optris PI connect software [5] used with the camera has a built-in annotation function, using the so-called measure areas of 10×20 pixels. We annotated the regions of interest including forehead, nose and cheeks, as depicted in Figure 2-Right. Additionally, the Optris PI connect has a built-in save option, that stores the mean temperature values of the annotated regions in CSV files.

3.3 Participants & Procedure

We recruited a total of 24 participants, and discarded 2 participants due problem with eye-tracking calibration. The remaining 22 participants in our final data set consisted of 14 Males and 8 Females with an average age of 20.45 years ($SD = 1.14$), recruited through university mailing lists. Upon arrival, participants were asked to sign a consent form and were informed about the aim of the study. We first asked participants to relax for 5 minutes while listening to white noise (relaxing sound of ocean waves) as the baseline task. This allowed us to collect their physiological data in a state of relaxation. Following, we presented the different tasks, 16 tasks in total for 3 minutes each. We explained each task to the participants before starting the task. The order of the tasks was counterbalanced using Latin squares. After each task, we asked participants to complete a NASA-TLX [24] questionnaire to assess the perceived cognitive load. The study lasted approximately 85 minutes ($SD = 10.25$). During the entire experiment, we recorded the facial temperature and eye gaze coordinates of the participant. The study was recorded using an RGB video camera (further described in the next section). We maintained the room temperature at 23°C, and participants were compensated with 10 EUR upon completion.

[4] http://www.optris.com/thermal-imager-pi450
[5] https://www.optris.com/

4 METHOD

In this section, we describe our step-by-step process in which we use to evaluate the combination of thermal imaging and eye-tracking for attention classification. First, we statistically analyzed the results to evaluate objective and subjective measures. Second, we extracted the features required for classification. Third, we built and tested different classifiers based on these features. We then measured the best performing classification model on our different classifiers before diving down into the performance of the combination.

4.1 Statistical Analysis

To validate our attention elicitation, we analyzed the effect of the attention types on both the subjective cognitive load from the NASA-TLX reported by the participants and the cognitive load inferred from the recorded facial temperature. We used three metrics as our dependent variables: the NASA-TLX score, forehead-nasal temperature, and cheeks temperature (detailed in Section 5.1). We statistically analyzed the data using a repeated measures ANOVA (with Greenhouse-Geisser correction if sphericity was violated). This was followed by post-hoc pairwise comparisons using Bonferroni-corrected t-tests.

4.2 Feature Extraction

To train our classifiers, we derived a feature set (14 features) that best describe the various attention types from both the gaze and thermal data (see Table 2 below). Below, we explain our reasoning behind our choices of features. The details of how we trained our classifier can be found in Section 4.3.

Table 2. Selected feature set used for classification.

Type	Subcategory	Feature
Gaze	Stimulus-dependent	Number of gaze transitions between pairs of AOI. We had up to 3 AOI so total features of 6 (2*3). AOI where maximum fixation lies in a window of 45 sec. Our task is designed for 180s so total features of 4 (45 * 4=180s).
Gaze	Stimulus-independent	Number of fixations. Mean fixation duration.
Thermal	Both	Mean forehead and nose temperature difference from the baseline. Mean temperature change in the cheeks from the baseline.

4.2.1 Gaze Features. Stimulus-dependent features are those that involve the knowledge of the AOI of the interface, whereas stimulus-independent features are statistical measures computed from eye movements. We pre-processed the gaze data by removing outliers and by clustering gaze points into fixations. We identified fixations using the Dispersion-Threshold Identification algorithm [55], as it produces accurate results in real-time using only two parameters, dispersion, and duration threshold (set to 20 and 100, respectively). From this data, we computed low-level statistical features, such as the number of fixations and mean fixation duration, as shown in Table 2.

As a representative example, Figure 6 shows the gaze plots for all combinations of task and attention type for one participant. For our purposes, the meaning of the area under the gaze point in regards to the task at hand is an important factor in determining the attention state. For example, consider a system that monitors a student while they watch a video lecture. Two similar fixation patterns will be indicative of attentive or inattentive states depending on whether it falls inside or outside the video player. Therefore, as suggested by Toker et al. [66], in

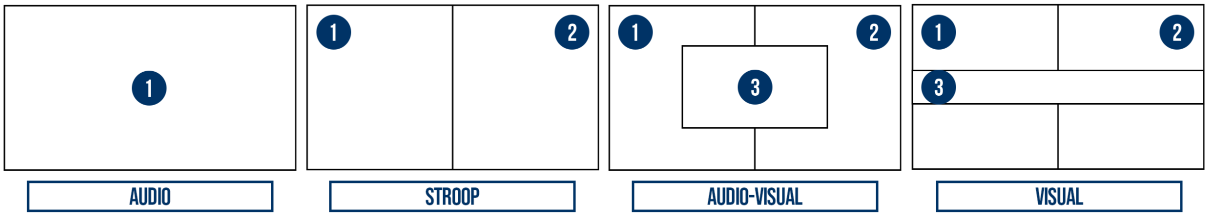

Fig. 3. Area of Interest (AOI). For the visual and audio-visual task and two AOI for the Stroop task. For audio-only stimuli, we only have one AOI as there is no visual stimuli.

addition to the stimulus-independent features, it is important also to compute stimulus-dependent features that encode the meaning behind different AOI. Hence, we divided the task interface into different numbers of AOI depending on the stimuli (see Figure 3). The stimuli-dependent features extracted were the number of fixations in an AOI for every 45 seconds and the number of gaze transitions between pairs of AOI. To compute the gaze pattern, we used the number of fixations in each AOI to identify the area with the highest number of fixations fixation every 45 seconds. Though in our experimental setup, we manually created the AOI, in a real system implementation, they could be set by the UI implementation framework used for its development.

4.2.2 Thermal Features. Previous works on thermal imaging for users' mental state detection [2, 27, 65], build upon the fact that changes in the internal states influence the blood flow [58]. Because our blood carries heat, changes in the blood flow influence our skin temperature [27, 28, 57, 58]. Therefore, monitoring changes in facial temperature can give us an insight into the changes in mental states. Researchers explored multiple regions of interest in the human body e.g., mouth, nose, and hand [28]. In particular, the face showed potential in detecting changes in states, as it is exposed and easy to capture by thermal cameras. Furthermore, it has a thin tissue layer, making temperature changes more observable. Therefore, in this work we explore how facial temperature fluctuations can give us an insight into changes in cognitive load caused by the experienced different attention types. We computed the temperature difference of the cheeks, forehead, and nose from their mean baseline temperature, similar to previous works [2, 27, 65].

4.3 Classification Approach

The goal of our classifier is to map a feature vector computed from a window of data to one of four classes corresponding to the type of attention the user was engaged as per the Clinical Model of Attention. To do this, we first built a *user-dependent, condition- independent* classifier, which was trained on the data from the same participant but different condition on which it was evaluated. This was followed by a *user-independent* classifier which was trained on the data from different participants on which it was evaluated. We then further evaluated the user-independent classifier in two ways—*condition-dependent* and *condition-independent.* To put simply, we trained the *condition dependent* variant on the data from other participants in the same condition (leave-one-out-cross-validation on participant), while the *condition-independent* classifier is trained on a different set of conditions and users to the dataset on which it was evaluated, e.g., trained on the Stroop, Audio and Visual datasets and evaluated on the Audio-Visual (leave-one-out-cross validation on participant and task). We provide more details on the three distinct classifiers in the remainder of this section. As different classification models will generate different levels of performance, we compared three different classifiers for all three classifiers: Support Vector Machines (SVM), K-Nearest Neighbor (KNN) and Logistic Regression (LR). For the SVM classification model, we used the two hyper-parameters C=5 and gamma=0.01 with

RBF kernel, while the KNN model was trained with k=5 neighbors. We used the scikit-learn package[6] for machine learning in Python for feature extraction and classification and PyCharm[7] as a development environment.

4.3.1 *User-Dependent Classifier.*

We built a user-dependent, condition-independent classifier by training the data on the same participant but different condition for the four tasks. This allows us to evaluate the performance of our approach of a system that is trained on its own user (e.g. by having a calibration phase). To do this, we trained and evaluated the classifier 22 times, using all 14 features, each time for a specific participant for the remaining conditions. For example, we trained the classifier on the data of a participant of Stroop, Visual and Audio task and evaluated the classifier on the data of the same participant but the Audio-Visual task).

4.3.2 *User-Independent Classifiers.*

As the results of a user-dependent classifier can potentially be optimistic (as the training and testing data are different, but not entirely independent), we next built a user-independent classifier. Being independent of the user, we can obtain a more robust and generalized classifier. To further avoid overfitting the user-independent classifier to the particular tasks we have chosen for our experiment, we further split the user independent classifier in *condition-dependent* and *condition-independent* variants.

Condition-Dependent. We evaluated the classification performance of the condition-dependent classifier on the data from the same condition on which it was trained, but from a different participant. We conducted separate evaluations for each task (Stroop, Audio, Audio-Visual, Visual), building and evaluating the classifiers using leave-one-participant-out cross-validation. We trained the classifier 22 times, each time training on the data of 21 participants and evaluating it on the remaining one participant.

Condition-Independent. We evaluated the condition independent classifier by training it 22 times using leave-one-participant-out cross-validation four times, one for each condition. Each time, we trained it on the data of the 21 participants for three conditions and evaluated it on the data of the fourth condition from the last participant. The reported results, in the next sections, are averaged by participant but split by the task on which it was evaluated.

5 RESULTS

5.1 Statistical Analysis

Below we present the effect of the attention types of the different conditions on the facial temperature as opposed to the baseline.

5.1.1 *Cognitive Load: NASA-TLX.*

To confirm that each attention type requires different cognitive resources [41], we first analyzed the effect of the different attention types on the reported cognitive load via the NASA-TLX. We tested the effect of the different attention types from different conditions on the overall cognitive load.

Condition-Independent NASA-TLX:. We first analyzed the mean NASA-TLX Score from all conditions (Stroop, Audio, Visual, and Audio-visual) for the four Attention Types. As depicted in Figure 4, the sustained attention had the lowest load with an average score of 23.50 (SD = 12.22), followed by the alternating attention with an average of 27.81 (SD = 14.59), selective attention with an average of 39.17 (SD = 15.88) and the highest was divided attention with an average score of 52.72 (SD = 18.49). We tested the effect of the Attention Type (4 types) on the overall NASA-TLX Score with a one-way ANOVA. Mauchly's test showed a violation of sphericity against difficulty ($0.29, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.65$) values. We found a significant large effect of Attention Type on the NASA-TLX score ($F_{2.09,41.72} = 72.29, p < .001, ges = 0.38$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$).

[6] https://scikit-learn.org/stable/

[7] https://www.jetbrains.com/pycharm/

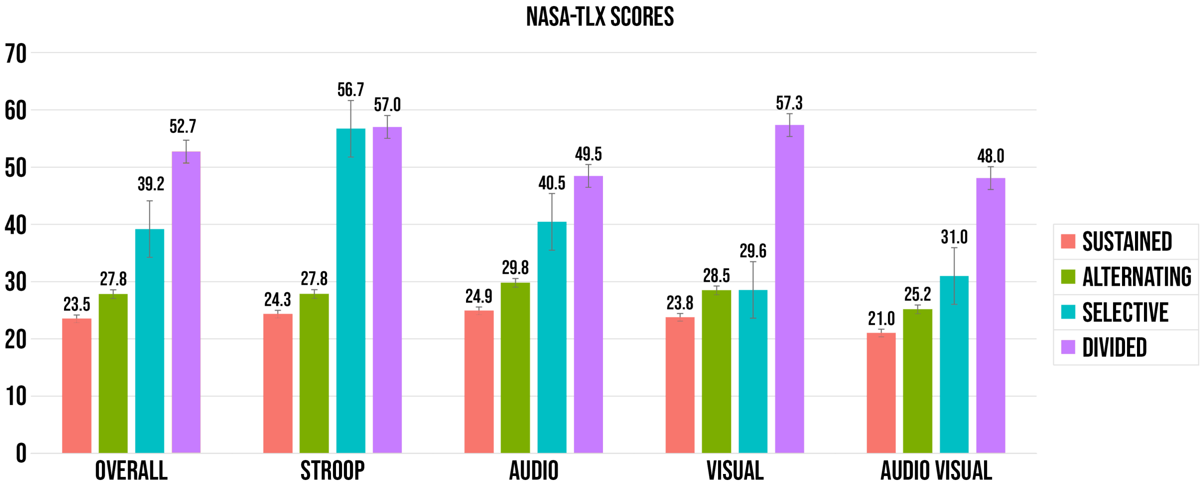

Fig. 4. The average cognitive load perceived by the participants when experiencing different attention types from different tasks. Error bars depict the standard error.

Stroop NASA-TLX: . We further analyzed the NASA-TLX score from the Stroop condition. The sustained attention had the lowest load with an average score of 24.32 (SD = 18.23), followed by the alternating attention with an average of 27.84 (SD = 22.51), selective attention with an average of 56.70 (SD = 24.03) and the highest was divided attention with an average score of 57.01 (SD = 22.16). We tested the effect of the attention Type (4 types) on the NASA-TLX Score with a one-way ANOVA. Mauchly's test showed a violation of sphericity against difficulty ($0.45, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.73$) values. We found a significant large effect of attention type on the NASA-TLX score ($F_{2.30,46.02} = 30.16, p < .001, ges = 0.34$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the selective and the divided attention.

Audio NASA-TLX: . In the audio condition, the sustained attention had the lowest load with an average score of 24.89 (SD = 18.23), followed by the alternating attention with an average of 29.78 (SD = 17.03), selective attention with an average of 40.45 (SD = 16.63) and the highest was divided attention with an average score of 48.45 (SD = 20.57). ANOVA revealed a significant effect of attention type on the NASA-TLX score ($F_{3,60} = 13.04, p < .001, ges = .21$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the selective and the divided attention.

Visual NASA-TLX: . Again, the sustained attention had the lowest load with an average score of 23.77 (SD = 13.24), followed by the alternating attention with an average of 28.48 (SD = 17.56), selective attention with an average of 28.55 (SD = 15.72) and the highest was divided attention with an average score of 57.34 (SD = 18.18). Mauchly's test showed a violation of sphericity against difficulty ($0.51, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.75$) values. ANOVA revealed a significant effect of attention type on the NASA-TLX score ($F_{3,60} = 57.96, p < .001, ges = .41$). However, Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, between the sustained and the selective, and between the alternating and the selective attention.

Audio Visual NASA-TLX: . Lastly, for the audio-visual condition, the sustained attention had the lowest load with an average score of 21.02 (SD = 13.29), followed by the alternating attention with an average of 25.15 (SD = 14.78), selective attention with an average of 30.98 (SD = 19.88) and the highest was divided attention with an average score of 48.07 (SD = 19.80). ANOVA revealed a significant effect of Attentio Type on the NASA-TLX score ($F_{3,60} = 24.88, p < .001, ges = .27$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the alternating and the selective attention.

In summary, the different attention types exhibited different NASA-TLX, where sustained attention showed the lowest NASA-TLX score, followed by the alternating, then selective, and the highest score was observed in the divided attention. Additionally, we found a significant difference in the NASA-TLX score.

5.1.2 Cognitive Load: Facial Temperature. Informed by the literature, cognitive load could be assessed by monitoring the facial temperature [2, 75], namely the difference between forehead and nose temperature (difference to the baseline). Other work [28], also investigated the temperature of the cheeks. In this work, we analyzed the effect of the attention type on the Forehead-Nasal temperature and the Cheeks temperatures.

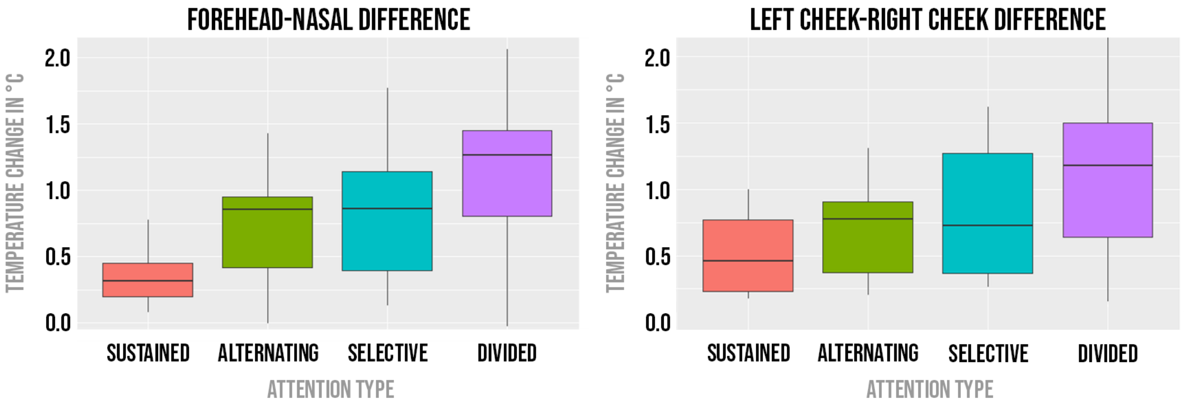

Fig. 5. Mean Temperature change between the baseline and the different attention types.

Effect of Stroop tasks on Facial Temperature.

Forehead-Nose. We tested the effect on the total change in the forehead and nose temperature. Mauchly's test showed a violation of sphericity against attention type ($0.47, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.55$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{2.3,46.4} = 39.22, p < .001, ges = 0.63$) was found. Bonferroni-corrected post-hoc tests shows significant differences between all types of attention ($p < .05$), except between the sustained and the alternating, and between the selective and divided attention.

Cheeks. For the cheeks temperature, Mauchly's test showed a violation of sphericity against attention type ($0.10, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.54$) values. A large significant effect of attention type on the cheeks temperature increase ($F_{1.6,33.9} = 50.36, p < .001, ges = 0.49$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and alternating attention.

Effect of Audio task on Facial Temperature.

Forehead-Nose. Mauchly's test showed a violation of sphericity against attention type ($0.24, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.53$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{1.6,34.7} = 29.08, p < .001, ges = 0.85$) was found. Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating attention, and between the selective and divided attention.

Cheeks. Mauchly's test showed a violation of sphericity against attention type ($0.31, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.57$) values. A large significant effect of attention type on the cheeks temperature increase ($F_{1.7,36.1} = 25.59, p < .001, ges = 0.64$) was found. Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the alternating and the selective attention.

Effect of Visual task on Facial Temperature.

Forehead-Nasal Temperature. we tested the effect of different attention type of visual tasks on the temperature metrics. Mauchly's test showed a violation of sphericity against attention type ($0.36, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.51$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{1.9,38.9} = 50.5, p < .001, ges = 0.78$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention.

Cheeks. Mauchly's test showed a violation of sphericity against attention type ($0.30, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.58$) values. A large significant effect of attention type on the cheeks temperature increase ($F_{1.8,36.7} = 30.05, p < .001, ges = 0.67$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and the alternating, and between the alternating and the selective attention.

Effect of Audio-Visual task on Facial Temperature.

Forehead-Nasal Temperature. Lastly, we tested the effect of different attention type of combination of audio-visual tasks on the temperature metrics. Mauchly's test showed a violation of sphericity against attention type ($0.39, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.39$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{1.6,33.9} = 56.15, p < .001, ges = 0.88$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the alternating and the selective attention.

Cheeks Temperature. We tested the effect on the cheeks temperature with a one-way ANOVA. A large significant effect of attention type on the cheeks temperature increase ($F_{2.2,46.9} = 20.00, p < .001, ges = 0.59$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and the alternating, and between the alternating and the selective attention.

In summary, our findings from the statistical analysis validate the correlation between attention types and cognitive load, deduced from the temperature changes in the selected region of interest. However, not all tasks exhibited significant difference between the alternating and selective attention types. Further, these findings are also aligned with the results from our subjective measure of perceived workload (NASA-TLX).

5.2 Classification Performance

To measure the performance of the classifiers, we computed the accuracy and Area Under the Curve (AUC), which aggregates precision and recall into one metric. We investigated the effect of the features used (gaze-only, thermal-only and gaze+thermal) as well as the usage of user-dependent and user-independent classifiers (condition-dependent and condition-independent) on the classification of attention types.

5.2.1 Comparison of Different Classification Models. We first compared the performance of the classifiers for the attention types on the three different models: SVM, KNN and Logistic Regression. Table 3 shows the performance of the user-dependent and user-independent classifiers using the AUC score for the three classification models. The AUC score reported in the table is the average AUC for all the four task. As shown, overall the Logistic Regression model outperforms both SVM and KNN for all three feature sets (gaze-only, thermal-only and gaze+thermal). The reason being that KNN is an example of a lazy learner [3] classifier which memorizes the training data rather than learning discriminative function and its performance is highly dependent on the selection of k values passed as an input parameter [23]. Similarly, SVM classification results highly depend on the kernel and hyper parameters chosen. As for the Logistic Regression model, it has less generalization error than KNN and is easier to build compared to an SVM model [13], and for our purpose, it gives the best classification performance overall. Due to this reason, for the remainder of our analysis we have chosen to explore our results using the Logistic Regression Classification model.

Table 3. Classification Results. Average AUC for all four task of the three classification models (SVM, KNN and Logistic Regression) for all classifiers.

Classifier	Classification Model	Gaze Features	Thermal Features	Gaze+Thermal Features
User Dependent	SVM	57.2 ± 3.1%	68.8 ± 3.8%	52.4 ± 2.5%
(Condition Independent)	KNN	58.4 ± 1.7%	66.6 ± 3.4%	52.2 ± 1.3%
	Logistic Regression	**58.9 ± 2.0%**	**70.7 ± 2.3%**	**77.4 ± 2.6%**
User Independent	SVM	59.6 ± 2.2%	70.6 ± 2.7%	61.1 ± 1.8%
(Condition Dependent)	KNN	61.3 ± 2.4%	71.0 ± 3.7%	61.3 ± 2.4%
	Logistic Regression	**76.5 ± 2.2%**	**69.7 ± 3.1%**	**86.9 ± 1.8%**
User Independent	SVM	53.1 ± 1.9%	72.3 ± 4.1%	54.5 ± 2.5%
(Condition Independent)	KNN	54.1 ± 2.1%	72.5 ± 3.4%	59.9 ± 2.2%
	Logistic Regression	**56.9 ± 0.9%**	**72.7 ± 2.6%**	**75.7 ± 1.8%**

5.2.2 Comparison of Different Classifiers (using Logistic Regression). Table 4 shows the overall performance for classification according to tasks for all classifiers. Overall, the user-independent, condition-dependent classifier performs the best compared to the other two classifiers with an average AUC score of 86.9%. In practice, this would be a classifier that is built-into the application, working for only for that application, but for any user. The high performance in this condition is expected due to the fact that this classifier is trained and evaluated on the same condition hence giving a higher performance for the same condition but not necessarily generalizing to other conditions. To build a more generalized classifier we built two other classifiers which are independent of the condition — user-dependent (condition independent) and user-independent (condition independent). We found the performance results to be comparable, obtained an average AUC score of 77.4% and 75.7% respectively. We note that these scores only decreased slightly when compared to the condition-dependent classifier, suggesting the validity of the general approach of using gaze and thermal imaging for attention classification. The user-dependent (condition-dependent) classifier is expected to perform slightly better as it will be trained and evaluated on the same user in the same condition. However, in the context of this work, we did not have enough data to train such a classifier.

Further, the results show that the accuracy of thermal-based classifier remained almost the same across all tasks. This means that the performance of the thermal-based classifier is largely independent of the task being performed by the user. Moreover, our findings showed that sustained attention required the least cognitive load

Table 4. Logistic Regression Classification Performance for All Classifiers (All Stimuli)

		Gaze Features		**Thermal Features**		**Gaze+Thermal Features**	
Classifier	Task	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
User Dependent (Condition Independent)	Stroop	44.3 ± 2.8%	62.8 ± 1.5%	57.9 ± 3.5%	71.9 ± 3.1%	68.3 ± 3.1%	78.8 ± 3.3%
	Audio	25.2 ± 1.8%	51.1 ± 1.0%	57.3 ± 3.1%	71.4 ± 3.3%	58.2 ± 1.2%	71.8 ± 1.2%
	Visual	46.2 ± 2.9%	52.1 ± 2.8%	54.8 ± 2.5%	68.1 ± 1.3%	70.4 ± 3.5%	80.3 ± 2.3%
	Audio Visual	54.4 ± 4.2%	69.7 ± 2.8%	57.9 ± 1.2%	71.4 ± 1.3%	68.8 ± 2.7%	78.8 ± 3.8%
User Independent (Condition Dependent)	Stroop	63.6 ± 4.6%	75.8 ± 3.0%	54.5 ± 4.5%	69.7 ± 3.0%	81.8 ± 3.5%	87.9 ± 2.2%
	Audio	26.1 ± 1.1%	50.8 ± 0.8%	53.4 ± 4.7%	69.7 ± 3.0%	48.9 ± 3.0%	65.9 ± 2.0%
	Visual	78.4 ± 4.1%	85.6 ± 2.8%	54.4 ± 2.4%	69.7 ± 3.8%	95.5 ± 2.7%	97.0 ± 1.8%
	Audio Visual	90.9 ± 3.1%	93.9 ± 2.1%	54.5 ± 4.2%	69.7 ± 2.8%	95.5 ± 2.1%	97.0 ± 1.4%
User Independent (Condition Independent)	Stroop	45.5 ± 4.1%	63.6 ± 1.0%	59.1 ± 3.2%	72.7 ±2.1%	67.8 ± 1.6%	78.8 ± 1.1%
	Audio	25.0 ± 0.0%	50.0 ± 0.0%	59.1 ± 5.5%	72.7 ± 3.7%	54.1 ± 2.7%	69.4 ± 1.8%
	Visual	40.1 ± 1.7%	58.7 ± 1.1%	59.1 ± 4.2%	72.7 ± 3.3%	70.1 ± 2.8%	80.4 ± 1.0%
	Audio Visual	36.0 ± 1.7%	55.3 ± 1.4%	58.2 ± 5.0%	72.7 ± 1.3%	61.8 ± 2.8%	74.0 ± 3.1%

followed by alternating, selective and divided attention, as reflected in the thermal features, and as suggested by our subjective measures. One important finding we observed was that the attention types are most accurately classified with an accuracy of (95.45%) when the participant is performing the visual and least accurately classified when the task being performed is audio. We also observed the same trend when comparing the performance of the user-independent (condition-dependent) classifier trained on just the gaze features. From our results, we observe that when classifying audio-only tasks, the thermal features alone worked as a better predictor than the classifiers that both gaze and thermal features. The reason being that the audio-only tasks lacks any visual stimuli, hence, the gaze features does not hold any significance for classifying attention in audio-only task—effectively working as noise. With reference to Figure 6, we can see that all tasks in each attention type have a distinct pattern, for example, in the alternating attention tasks, we can see a clear pattern the left and right AOIs. As for the audio-only tasks, the gaze patterns appear to be random with the participant, either focusing randomly around the screen or at a focused point with random saccades around the screen. Due to this reason, the average for the condition-independent classifiers for all tasks does not perform as well compared to the condition-dependent classifier as their training set includes the insignificant gaze features of the audio-only task, which decrease the classification accuracy for the gaze-only and gaze+thermal feature sets.

To measure the effect of removing the task which lacks visual stimuli (i.e. audio-only tasks) on the condition-independent and user-dependent classifier, we retrained our classifiers by only considering the task with visual stimuli (Audio Visual, Visual and Stroop). This is so the gaze features in the training data set would remain meaningful in the classification process.

We evaluated the *user-independent, condition-independent* classifier by training it 22 times using leave-one-out-cross-validation (LOOCV) three times, one for each condition. Each time, we trained it on the data of the 21 participants for two conditions and evaluated it on the data of the third condition from the last participant. As for the user-dependent, condition-independent, classification, we trained the classifier 22 times using cross-validation three times, one for each condition. Each time, we trained it on the data of the single participant for two conditions and evaluated it on the data of the third condition from the same participant. The results of the user-independent, condition-independent and the user-dependent, condition-independent classification that only considers tasks

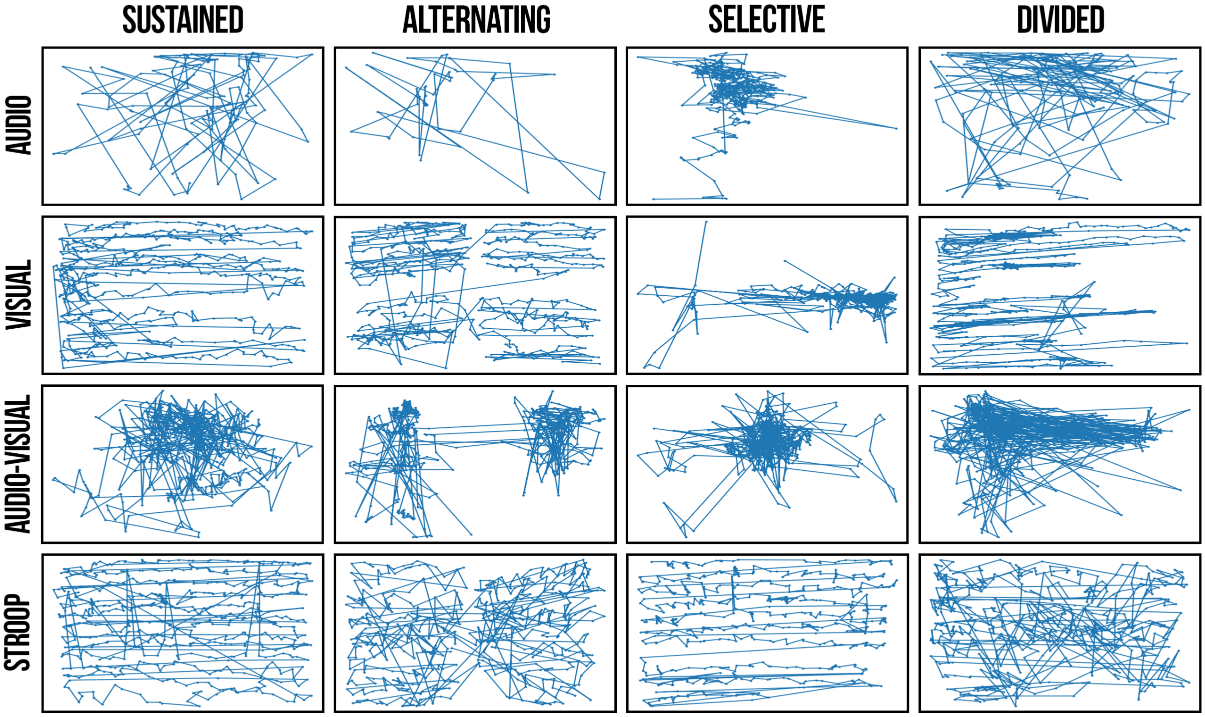

Fig. 6. Gaze Plots, highlighting the patterns in each task and how the audio task has no observed unique patterns.

Table 5. Logistic Regression Classifier Performance without audio-only tasks

		Gaze Features		Thermal Features		Gaze+Thermal Features	
Classifier	Task	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
User Independent (Condition Independent)	Stroop Visual Audio Visual	46.5 ± 3.1% 44.1 ± 3.6% 37.5 ± 2.7%	65.2 ± 2.7% 62.9 ± 2.4% 58.3 ± 1.8%	48.9 ± 3.0% 59.1 ± 4.2% 57.9 ± 5.0%	68.2 ±3.5% 72.7 ± 3.3% 71.9 ± 3.3%	77.3 ± 4.0% 79.5 ± 3.5% 62.5 ± 3.10%	84.8 ± 2.7% 86.4 ± 2.0% 75.0 ± 2.3%
User Dependent (Condition Independent)	Stroop Visual Audio Visual	48.0 ± 2.5% 48.9 ± 1.9% 57.9 ± 3.8%	68.3 ± 1.6% 65.9 ± 1.3% 71.9 ± 2.5%	62.5 ± 3.2% 60.3 ± 4.2% 68.1 ± 4.9%	75.0 ± 3.8% 73.5 ± 3.5% 78.8 ± 3.3%	77.8 ± 3.6% 80.7 ± 3.2% 76.1 ± 4.1%	85.1 ± 2.4% 87.1 ± 2.1% 84.1 ± 1.7%

with visual stimuli shows an increased classification accuracy (see Table 5). Hence, observing a better classification accuracy for each task for both type of classifier has increased as now the gaze feature plays a significant role in attention classification. In practice, this means that eye-tracking only helps the classifier when the task involves a visual stimulus. Otherwise (i.e. as in audio-only tasks), it tends to harm the classification results.

6 DISCUSSION

The results of our study and a review of prior work revealed that different attentive states could be distinguished by the locus of visual attention and estimated cognitive load. On this basis, in the previous sections, we presented the results from collecting, analyzing and classifying gaze and thermal data of different attention types, which we summarize and discuss grouped by the most important observations in the following.

6.1 On Performance

In this work, we discuss a first attempt of combining thermal imaging and eye tracking to discriminate between four types of user attention. Our results show that attention classification is feasible, achieving an accuracy of up to 95.45% when using a condition-dependent prediction (see Table 4). This result is promising as it paves the way for new applications in which classification can be tailored to a particular known condition or task. For example, this could be embedded into an e-learning system to measure student attention during a lecture.

In contrast, the condition-independent classification is more challenging. When comparing the performance with the condition-dependent classifier, we observed a decrease in the accuracy between 62.5% and 79.50% while considering only the tasks with visuals stimuli. Though this performance might be sufficient for some applications and is well-above the 25% baseline, further work is needed to bring performance up to the same level as for condition-dependent prediction. This means that this approach is not yet quite feasible for distinguishing attention types in unknown tasks.

However, in all of our experiments, our user-independent results were strong, suggesting that by training the classifier on one specific task, the classification generalizes well to unseen users.

6.2 On Discriminating Different Attention Types

Based on our review of the literature, we hypothesized that different attention types require different cognitive load levels, which would lead to a change in the participants' facial temperature patterns. From previous work, we know that regions on the face such as forehead, nose, and cheeks are often visible points and are feasible for temperature measurement [2, 28] We tested the effects of different tasks and their attention types on temperature changes in these points. We elicited different attention types through set of tasks with different stimuli and found a significant difference in the metrics used across the different attention types. We confirmed the validity of our findings, where the same pattern of facial temperature changes was observed across the different conditions (Stroop, Audio, Visual, and Audio-Visual). Our findings from the statistical analysis validate the correlation of cognitive load, deduced from the selected region of interest, and the different attention types. Although it was not significant across all attention types (e.g. sustained and alternating attention), it could give a hint about the experienced attention type. Furthermore, this highlights the role of gaze data to complement thermal data.

For discriminating attention types using a classifier, we investigated the performance when classifying each attention type separately. Our results show that alternating attention achieved the highest accuracy for the thermal and the eye feature set because the alternating gaze pattern of participants from one AOI to another is a strong indicator of alternating attentional state. For this task, the thermal features do not capture much information of participant attention state as indicated by low performance of the classifier trained on just the thermal features set (see Table 6).

For sustained and divided attention, gaze features did not work as well but we found that thermal features performed well. Temperature variation was considerably different compared to other attentional states as shown in Figure 5. For selective attention, the performance of the classifier was the lowest. This attention type was mostly confused with divided attention, as can be seen from the confusion matrix (see Figure 7). One likely reason is that the thermal features (forehead-nasal and cheeks) change across all attention types. For instance, the change in facial temperature for selective attention overlaps the most with the divided attention (Figure 5).

Table 6. Recognition accuracy of each attention type with the condition-independent classifier based on gaze only, thermal only and gaze and thermal features.

Attention Types	Gaze	Thermal	Gaze & Thermal
Alternating	100%	27.27%	100%
Sustained	13.0%	86.36%	90.9%
Selective	27.0%	36.36%	18.18%
Divided	50.0%	69.7%	81.81%

6.3 On Combining Thermal Imaging and Eye Tracking

Additionally, we found that combining both gaze and thermal features boosted the performance of the classifier as compared to using gaze or thermal only (Figure 7 and Table 5) for the visual tasks. This is because each modality complements the other for the classification of attentional types. For instance, Figure 6 shows that divided and sustained attention present very similar gaze patterns but elicit very different levels of cognitive load, which is reflected in the thermal features (Figure 5). In contrast, alternating attention presents itself somewhere in between sustained and selective attention in terms of facial temperature but exhibits very distinct eye movement patterns as reflected in the figure. This highlights the importance and potential of using thermal imaging and eye tracking in combination to classify attention types. Interestingly, we observed that thermal features exhibit the same performance for different conditions. This validates that different attention types allocate different cognitive load, regardless of the stimuli (see Table 4). In contrast to the gaze features, because we rely on the AOI, the features are influenced by the task and stimuli. This is reflected in classification accuracy using only the gaze features. For instance, as shown in Table 4 the fixations obtained in the audio task for various attention types were arbitrary (see Figure 6), and we did not observe any unique patterns of gaze transition for different attention type as the participant was asked to just attend to the playing audio.

6.4 On Different Conditions

We observed a decrease in classification accuracy for the audio condition when using gaze and thermal as opposed to using thermal features only. This is because in the audio condition, participants' eye movements were arbitrary, due to the lack of visual stimuli. Hence, training the classifier with the audio task gaze data, would mean training the classifier with confusing data. In other words, including gaze data of the audio only condition, would then yield to reduced performance. Further exploring the confusion matrix (see Figure 9) of the classifiers trained on the thermal feature only, we can conclude that for an audio-only task attention could be classified into sustained and divided attention more accurately then the selective and alternating type. Based on this observation, we suggest using only thermal features to classify attention types for audio-only tasks.

6.5 On Attention Type-Aware System Development Approaches

We observed that the average classification accuracy for the user-dependent, condition independent classifier in all four tasks and three feature sets is higher than the user-independent classifiers (see Table 4 and Table 5). This means that the user-dependent classifier was able to predict the attention type of a specific user more accurately when trained on the data of the same user (user-dependent) rather than training it on features of all users (user-independent). Similar results on discriminating attention types were achieved for a user-dependent classifier (see Figure 8) with alternating attention achieving the highest accuracy, and selective attention achieving the lowest accuracy for thermal and gaze feature set. We found that similar to the results obtained for user-independent classifier, the performance of user-dependent classifier is also boosted by combining both gaze and thermal feature set for the Stroop, Visual and Audio-Visual task. One obvious limitation of user-dependent classifier would be

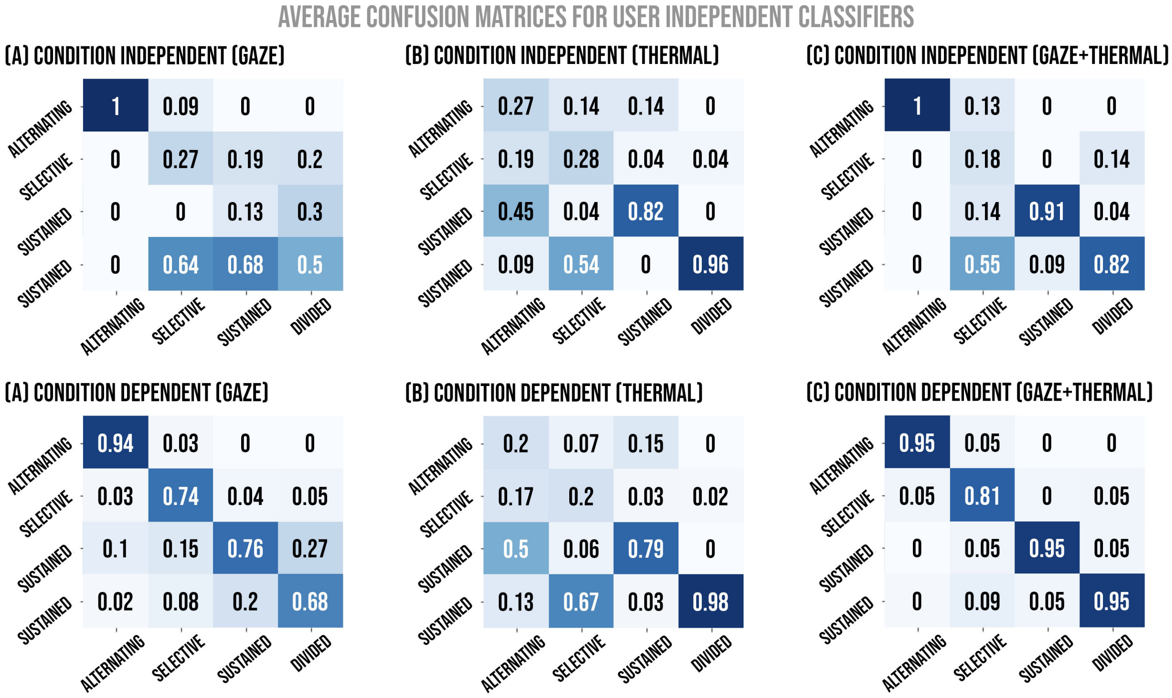

Fig. 7. User Independent Average Confusion Matrices. The results show the classification for all tasks expect Audio-only tasks.

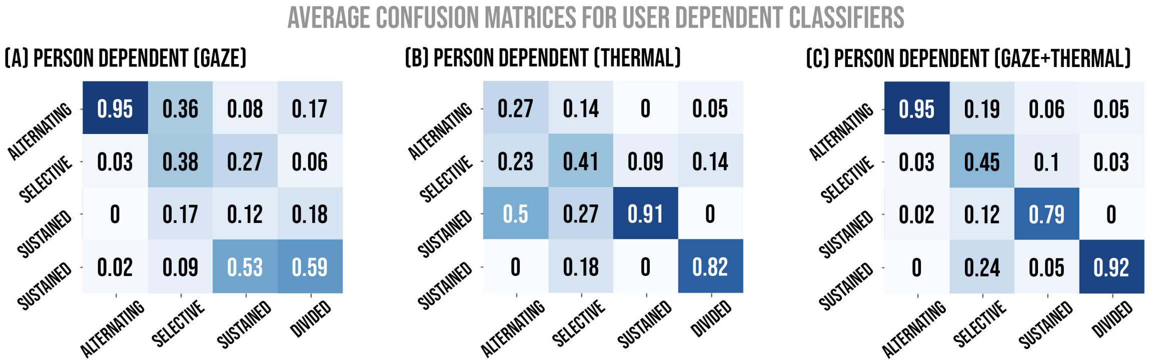

Fig. 8. User Dependent Average Confusion Matrices. The results show the classification for all tasks expect Audio-only tasks.

that it would not be generalized for different users. In practical terms, this means that if the system requires user calibration prior to use. However, if a real-time attention classification system is required, which could classify attention of any user without being trained every time for a new user, then a user-independent classification approach would be more suited. Therefore, the types of classification approach taken would highly dependent on the type of application the system is used in.

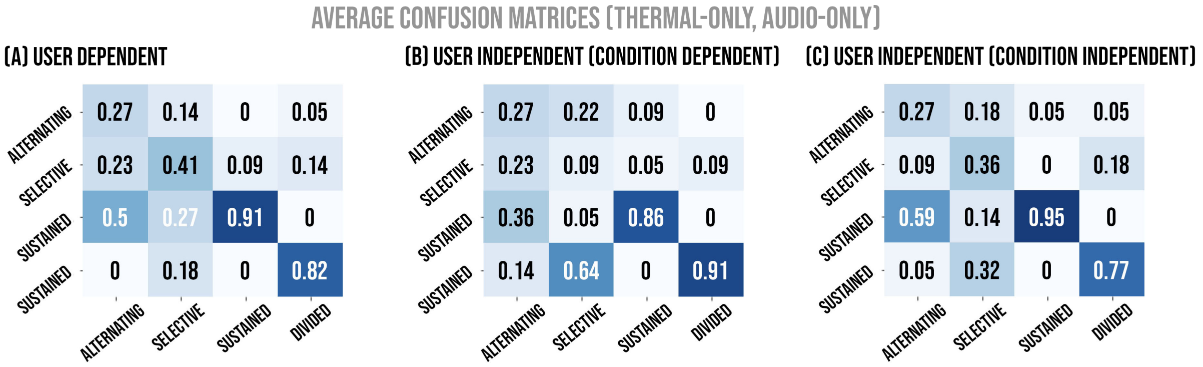

Fig. 9. Average confusion matrices for all classifiers for audio only task on thermal feature set.

7 EXAMPLE USE CASES

Our findings show that the proposed classifier was able to classify attention types unobtrusively. Applications that take into account the attention type can be applied to a broad range of applications ranging from education [25, 54], performance management in the workplace, distraction management [7, 31], to quantified-self applications [12]. Educational applications could monitor students' attention type and adapt accordingly, e.g. assessing if the presented material is “attention-grabbing" so that the students would show sustained attention rather than divided or alternating attention, aiming to better design learning systems [25]. Furthermore, workplaces could benefit from our approach by helping workers manage their attention if he/she are experiencing divided attention during safety-critical task, and to avoid divided attention in dangerous situations (e.g. operating trains) [38, 71]. Additionally, if a user should focus to finish a task, an attentive user interface could help the user to keep their attention sustained on the task [61].

Online distractions are a controversial aspect of our current technology-mediated workplaces. Our approach could be used to manage distractions, especially with users that are more susceptible to social distraction [42]. An interface could block the distraction source (e.g. social media, smartphone) when alternating or divided attention between the task in hand and the distraction source is detected [7]. We also can deploy attention-type detection for quantified-self applications as proposed by Dingler et al. [12]. The system could monitor the attention type patterns throughout the day, aiming to assist users with tracking and managing their attention distribution to enhance their well-being. For instance, stress and frustration occur when there is a mismatch between the accomplished tasks and the planned ones [67] due to the lack of sustained attention on the planned tasks. Furthermore, the high frequency of divided attention may lead to burnout and memory distortion [46].

8 LIMITATIONS AND FUTURE WORK

This work proposes the first steps towards classifying attention using unobtrusive sensors. As such, we required a dataset with clearly labeled attention types for training our classifier.

Despite these promising results, our work has several limitations that we plan to address in future work. First, the controlled task is likely to lead to behavior changes. Similar to studies in Affective Computing, there is a trade-off between the quality of the labels and the naturalness of user behavior. We opted for a controlled setup to increase the quality of the labels at the expense of natural behavior. By demonstrating the feasibility of the approach, our next steps will involve collecting a more naturalistic *in-the-wild* dataset. Second, we labeled the data according to the elicited attentional state. While these tasks were informed by previous work in psychology, it is difficult to guarantee that users were in those states at all times during the tasks. For example, during sustained attention, we cannot guarantee that participants did not momentarily “mind-wandered”. We tried to minimize these effects by keeping our tasks time reasonably short. Third, users’ eye movements are highly dependent on the stimuli used. We attempted to minimize the effects of the particular stimulus by abstracting from the visual layout of the interface , instead, computing features based on Areas of Interest (AOI). This tends to minimize the overfitting due to the visual design of the interface as compared to low-level features only. For example, in our design, the two pieces of text in the alternating attention condition were side-by-side. If we had trained a classifier using saccade directions, a high proportion of large sideways saccades would likely be indicative of alternating attention. However, if the same classifier were applied in an interface where the two texts were displayed one on top of each other, the approach would no longer work. Using AOIs allows us to abstract from the specifics of the interface, but also introduces a new challenge—how to determine which areas are of interest. This limitation can be addressed in many ways. For example, a learning system could specify that the video player is an AOI. A system like *RescueTime* could classify the applications that are part of productive (i.e., attentive) use of time and set it as the AOIs.

Fourth, thermal imaging is influenced by external factors, e.g. changes in room temperature, and internal factors, e.g. changes in affective states. These can be confounds that might affect the performance of the system in the wild. A more naturalistic dataset is required to explore these questions. Additionally, we envision that running an evaluation on participants with more experience in executing focused tasks such as seasoned workers would yield interesting insights, as well as running this over a longer period of time. We also plan to explore the performance of the gaze and thermal classifier by extracting more stimulus-dependent gaze features such as saccade velocity and length from one AOI to another and stimuli dependent feature such as the total fixation rate and mean saccade rate and angle for the individual task. Lastly, our findings open up further research question—how to distinguish between selective and divided attention. This can be explored by augmenting another bio-data e.g., GSR, heart rate, aiming to investigate if they differ in terms of other physiological responses.

9 CONCLUSION

Through our review of related work, we concluded that no prior work explored the use of thermal imaging combined with eye tracking to classify attention types. Consequently, in this work, we began our exploration by identifying gaze and features that could potentially reveal the four attention types—*sustained*, *selective*, *alternating* and *divided* attention. We investigated the effects of using different feature sets (gaze, thermal and the combination of thermal and gaze features) in classifying the four attention types. We used the extracted features to train two categories of classifiers: (1) condition-dependent and (2) condition-independent classifiers. Our classifiers achieved AUC up to 95.45% and 79.5% respectively. Furthermore, we investigated the performance of user dependent and independent classifiers, we had AUC up to 75.7% for the user independent-condition independent, 87% user-independent-condition dependent, and 77.4% for the user-dependent classifier). We additionally found that there is an increase in the classification accuracy when using the combination of gaze and thermal features as opposed to using gaze or thermal features alone. In this work, we were able to classify attention types unobtrusively, using a thermal camera and a remote eye tracker. This enables novel opportunities in the field of attention-aware computing: our approach, for example, can be applied in different research areas, e.g., education, adaptive and assistive systems. It could also be used to track and give feedback to the user, to increase the user’s awareness of their attention patterns.

In summary, our results reveal the feasibility of building an attention classifier based on facial temperature and eye movements. Hence, we envision that our work can serve as an initial building block to understanding the human mind and the influence of different attention types. We hope that developers of attention-aware and adaptive systems can use our results to build enhanced adaptive systems with a diverse set of application to benefit users in everyday usage.

ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG), Grant No. AL 1899/2-1, and 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). Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Researcher Award (Project Number: DE180100315) funded by the Australian Government. Lastly, Anam Ahmad Khan and Joshua Newn are both supported by scholarships under the Australian Commonwealth Government Research Training Program.

REFERENCES

[1] Yomna Abdelrahman, Mariam Hassib, Maria Guinea Marquez, Markus Funk, and Albrecht Schmidt. 2015. Implicit Engagement Detection for Interactive Museums Using Brain-Computer Interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*. ACM, New York, NY, USA, 838–845. https://doi.org/10.1145/2786567.2793709

[2] 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 (Sept. 2017), 20 pages. https://doi.org/10.1145/3130898

[3] David W. Aha. 1997. Lazy Learning. Kluwer Academic Publishers, Norwell, MA, USA, Chapter Lazy Learning, 7–10. http://dl.acm.org/citation.cfm?id=273530.273534

[4] James Rowland Angell. 1910. *Psychology: An Introductory Study of the Structure and Function of Human Consciousness.* H. Holt.

[5] T Armstrong and BO Olatunji. 2009. What They See Is What You Get: Eye Tracking of Attention in the Anxiety Disorders. *Psychological Science Agenda* 23, 3 (2009).

[6] Othman Asiry, Haifeng Shen, and Paul Calder. 2015. Extending Attention Span of ADHD Children Through an Eye Tracker Directed Adaptive User Interface. In *Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference (ASWEC ' 15 Vol. II)*. ACM, New York, NY, USA, 149–152. https://doi.org/10.1145/2811681.2824997

[7] Jonas Auda, Dominik Weber, Alexandra Voit, and Stefan Schneegass. 2018. Understanding User Preferences Towards Rule-based Notification Deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW584, 6 pages. https://doi.org/10.1145/3170427.3188688

[8] Andreas Bulling. 2016. Pervasive Attentive User Interfaces. *IEEE Computer* 49, 1 (2016), 94–98. https://doi.org/10.1109/MC.2016.32

[9] Carlos Carreiras, André Lourenço, Helena Aidos, Hugo Plácido da Silva, and Ana L. N. Fred. 2016. *Unsupervised Analysis of Morphological ECG Features for Attention Detection.* Springer International Publishing, Cham, 437–453. https://doi.org/10.1007/978-3-319-23392-5_24

[10] O. T. . Chen, P. Chen, and Y. Tsai. 2017. Attention Estimation System via Smart Glasses. (Aug 2017), 1–5. https://doi.org/10.1109/CIBCB.2017.8058565

[11] Thomas H. Davenport and John C. Beck. 2001. The Attention Economy. *Ubiquity* 2001, May, Article 6 (May 2001). https://doi.org/10.1145/376625.376626

[12] Tilman Dingler, Albrecht Schmidt, and Tonja Machulla. 2017. Building Cognition-Aware Systems: A Mobile Toolkit for Extracting Time-of-Day Fluctuations of Cognitive Performance. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 47 (Sept. 2017), 15 pages. https://doi.org/10.1145/3132025

[13] Stephan Dreiseitl and Lucila Ohno-Machado. 2002. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics* 35, 5 (2002), 352 – 359. https://doi.org/10.1016/S1532-0464(03)00034-0

[14] Albert Hoang Duc, Paul Bays, and Masud Husain. 2008. Chapter 5.5 - Eye movements as a Probe of Attention. In *Using Eye Movements as an Experimental Probe of Brain Function*, Christopher Kennard and R. John Leigh (Eds.). Progress in Brain Research, Vol. 171. Elsevier, 403 – 411. https://doi.org/10.1016/S0079-6123(08)00659-6

[15] Ami Eidels, James T. Townsend, and Daniel Algom. 2010. Comparing perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition* 114, 2 (2010), 129 – 150. https://doi.org/10.1016/j.cognition.2009.08.008

[16] 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 (CHI EA '17)*. ACM, New York, NY, USA, 2543–2550. https://doi.org/10.1145/3027063.3053179

[17] Charles W Eriksen and James E Hoffman. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics* 12, 2 (1972), 201–204.

[18] Josef Falkinger. [n.d.]. Limited Attention as a Scarce Resource in Information-Rich Economies. *The Economic Journal* 118, 532 ([n. d.]), 1596–1620. https://doi.org/10.1111/j.1468-0297.2008.02182.x

[19] Josef Falkinger. 2007. Attention economies. *Journal of Economic Theory* 133, 1 (2007), 266–294.

[20] Maite Frutos-Pascual and Begonya Garcia-Zapirain. 2015. Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games. *Sensors* 15, 5 (2015), 11092–11117. https://doi.org/10.3390/s150511092

[21] Markus Funk, Tilman Dingler, Jennifer Cooper, and Albrecht Schmidt. 2015. Stop Helping Me - I'M Bored!: Why Assembly Assistance Needs to Be Adaptive. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 1269–1273. https://doi.org/10.1145/2800835.2807942

[22] John M. Gardiner and Alan J. Parkin. 1990. Attention and Recollective Experience in Recognition Memory. *Memory & Cognition* 18, 6 (01 Nov 1990), 579–583. https://doi.org/10.3758/BF03197100

[23] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. 2003. KNN Model-Based Approach in Classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Robert Meersman, Zahir Tari, and Douglas C. Schmidt (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 986–996.

[24] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9, 904–908. https://doi.org/10.1177/154193120605000909

[25] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A System for Implicit Audience Engagement Sensing Using Electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5114–5119. https://doi.org/10.1145/3025453.3025669

[26] James E. Hoffman and Baskaran Subramaniam. 1995. The role of visual attention in saccadic eye movements. *Perception & Psychophysics* 57, 6 (01 Jan 1995), 787–795. https://doi.org/10.3758/BF03206794

[27] 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.

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

[29] Daniel Kahneman. 1973. *Attention and Effort*. Vol. 1063. Prentice-Hall Englewood Cliffs, NJ.

[30] Kimberly A. Kerns, Karen Eso, and Jennifer Thomson. 1999. Investigation of a Direct Intervention for Improving Attention in Young Children With ADHD. *Developmental Neuropsychology* 16, 2 (1999), 273–295. https://doi.org/10.1207/S15326942DN1602_9

[31] Thomas Kosch, Yomna Abdelrahman, Markus Funk, and Albrecht Schmidt. 2017. One Size Does Not Fit All: Challenges of Providing Interactive Worker Assistance in Industrial Settings. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp '17)*. ACM, New York, NY, USA, 1006–1011. https://doi.org/10.1145/3123024.3124395

[32] Nataliya Kosmyna, Caitlin Morris, Utkarsh Sarawgi, and Pattie Maes. 2019. AttentivU: A Biofeedback System for Real-time Monitoring and Improvement of Engagement. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article VS07, 2 pages. https://doi.org/10.1145/3290607.3311768

[33] P. Kuyper. 1972. The cocktail party effect. *Audiology* 11, 5-6 (1972), 277–282. https://doi.org/10.3109/00206097209072593

[34] Martijn J. M. Lamers, Ardi Roelofs, and Inge M. Rabeling-Keus. 2010. Selective attention and response set in the Stroop task. *Memory & Cognition* 38, 7 (01 Oct 2010), 893–904. https://doi.org/10.3758/MC.38.7.893

[35] Muhamad Hafiz Abd Latif, Hazlina Md Yusof, S. Naim Sidek, and Nazreen Rusli. 2015. Implementation of GLCM Features in Thermal Imaging for Human Affective State Detection. *Procedia Computer Science* 76 (2015), 308 – 315. https://doi.org/10.1016/j.procs.2015.12.298 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IEEE IRIS2015).

[36] Sophie Leroy. 2009. Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes* 109, 2 (2009), 168 – 181. https://doi.org/10.1016/j.obhdp.2009.04.002

[37] Yongchang Li, Xiaowei Li, Martyn Ratcliffe, Li Liu, Yanbing Qi, and Quanying Liu. 2011. A Real-time EEG-based BCI System for Attention Recognition in Ubiquitous Environment. In *Proceedings of 2011 International Workshop on Ubiquitous Affective Awareness and Intelligent Interaction (UAAII '11)*. ACM, New York, NY, USA, 33–40. https://doi.org/10.1145/2030092.2030099

[38] Lars Lischke, Sven Mayer, Andreas Preikschat, Markus Schweizer, Ba Vu, Paweł W. Woźniak, and Niels Henze. 2018. Understanding Large Display Environments: Contextual Inquiry in a Control Room. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW134, 6 pages. https://doi.org/10.1145/3170427.3188621

[39] Ning-Han Liu, Cheng-Yu Chiang, and Hsuan-Chin Chu. 2013. Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors. *Sensors* 13, 8 (2013), 10273–10286. https://doi.org/10.3390/s130810273

[40] Jun Ma, Du Lei, Xingming Jin, Xiaoxia Du, Fan Jiang, Fei Li, Yiwen Zhang, and Xiaoming Shen. 2012. Compensatory brain activation in children with attention deficit/hyperactivity disorder during a simplified Go/No-go task. *Journal of Neural Transmission* 119, 5 (2012), 613–619.

[41] Matei Mancas, Vincent P Ferrera, Nicolas Riche, and John G Taylor. 2016. *From Human Attention to Computational Attention: A Multidisciplinary Approach*. Vol. 10. Springer.

[42] Gloria Mark, Mary Czerwinski, and Shamsi T. Iqbal. 2018. Effects of Individual Differences in Blocking Workplace Distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 92, 12 pages. https://doi.org/10.1145/3173574.3173666

[43] Allan F. Mirsky, Bruno J. Anthony, Connie C. Duncan, Mary Beth Ahearn, and Sheppard G. Kellam. 1991. Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review* 2, 2 (01 Jun 1991), 109–145. https://doi.org/10.1007/BF01109051

[44] Mahnas Jean Mohammadi-Aragh, John E. Ball, and Donna Jaison. 2016. Using wavelets to categorize student attention patterns. In *2016 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE.2016.7757403

[45] Mona Moisala, Viljami Salmela, Emma Salo, Synnöve Carlson, Virve Vuontela, Oili Salonen, and Kimmo Alho. 2015. Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks. *Frontiers in Human Neuroscience* 9 (2015), 86. https://doi.org/10.3389/fnhum.2015.00086

[46] Moshe Naveh-Benjamin, Jonathan Guez, Yoko Hara, Matthew S. Brubaker, and Iris Lowenschuss-Erlich. 2014. The Effects of Divided Attention on Encoding Processes under Incidental and Intentional Learning Instructions: Underlying Mechanisms? *Quarterly Journal of Experimental Psychology* 67, 9 (2014), 1682–1696. https://doi.org/10.1080/17470218.2013.867517 PMID: 24283628.

[47] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 261, 12 pages. https://doi.org/10.1145/3173574.3173835

[48] Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, New York, NY, USA, 541–552. https://doi.org/10.1145/3116595.3116624

[49] Anat Ninio and Daniel Kahneman. 1974. Reaction time in focused and in divided attention. *Journal of Experimental Psychology* 103, 3 (1974), 394.

[50] Michael I. Posner. 1980. Orienting of Attention. *Quarterly Journal of Experimental Psychology* 32, 1 (1980), 3–25.

[51] Michael I. Posner and Steven E. Petersen. 1990. The Attention System of the Human Brain. *Annual Review of Neuroscience* 13, 1 (1990), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325 PMID: 2183676.

[52] Penelope J Qualls and Peter W Sheehan. 1981. Role of the feedback signal in electromyograph biofeedback: The relevance of attention. *Journal of Experimental Psychology: General* 110, 2 (1981), 204.

[53] Jailan Salah, Yomna Abdelrahman, Yasmeen Abdrabou, Khaled Kassem, and Slim Abdennadher. 2018. Exploring the Usage of Commercial Bio-Sensors for Multitasking Detection. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 265–277. https://doi.org/10.1145/3282894.3282900

[54] Jailan Salah, Yomna Abdelrahman, Ahmed Dakrouni, and Slim Abdennadher. 2018. Judged by the Cover: Investigating the Effect of Adaptive Game Interface on the Learning Experience. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 215–225. https://doi.org/10.1145/3282894.3282930

[55] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

[56] Jerritta Selvaraj, Murugappan M, R Nagarajan, and Khairunizam Wan. 2011. Physiological signals based human emotion Recognition: a review. (03 2011).

[57] 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.

[58] Rajita Sinha, William R Lovallo, and Oscar A Parsons. 1992. Cardiovascular differentiation of emotions. *Psychosomatic Medicine* 54, 4 (1992), 422–435.

[59] McKay Moore Sohlberg and Catherine A. Mateer. 1987. Effectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology* 9, 2 (1987), 117–130. https://doi.org/10.1080/01688638708405352 PMID: 3558744.

[60] Namrata Srivastava, Joshua Newn, and Eduardo Velloso. 2018. Combining Low and Mid-Level Gaze Features for Desktop Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 189 (Dec. 2018), 27 pages. https://doi.org/10.1145/3287067

[61] Julian Steil, Philipp Müller, Yusuke Sugano, and Andreas Bulling. 2018. Forecasting User Attention During Everyday Mobile Interactions Using Device-integrated and Wearable Sensors. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. ACM, New York, NY, USA, Article 1, 13 pages. https://doi.org/10.1145/3229434.3229439

[62] J Ridley Stroop. 1992. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General* 121, 1 (1992), 15.

[63] Walter Sturm, Klaus Willmes, Bernt Orgass, and Wolfgang Hartje. 1997. Do Specific Attention Deficits Need Specific Training? *Neuropsychological Rehabilitation* 7, 2 (1997), 81–103. https://doi.org/10.1080/713755526

[64] Elizabeth A. Styles. 1997. *The Psychology of Attention*. Psychology Press.

[65] Benjamin Tag, Ryan Mannschreck, Kazunori Sugiura, George Chernyshov, Naohisa Ohta, and Kai Kunze. 2017. Facial Thermography for Attention Tracking on Smart Eyewear: An Initial Study. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2959–2966. https://doi.org/10.1145/3027063.3053243

[66] Dereck Toker, Cristina Conati, Ben Steichen, and Giuseppe Carenini. 2013. Individual User Characteristics and Information Visualization: Connecting the Dots Through Eye Tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 295–304. https://doi.org/10.1145/2470654.2470696

[67] Geertje van Daalen, Tineke M Willemsen, Karin Sanders, and Marc JPM van Veldhoven. 2009. Emotional exhaustion and mental health problems among employees doing "people work": The impact of job demands, job resources and family-to-work conflict. *International archives of occupational and environmental health* 82, 3 (2009), 291–303.

[68] Mélodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2012. Wearable eye tracking for mental health monitoring. *Computer Communications* 35, 11 (2012), 1306 – 1311. https://doi.org/10.1016/j.comcom.2011.11.002

[69] Alexandra Voit, Benjamin Poppinga, Dominik Weber, Matthias Böhmer, Niels Henze, Sven Gehring, Tadashi Okoshi, and Veljko Pejovic. 2016. UbiTtention: Smart & Ambient Notification and Attention Management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1520–1523. https://doi.org/10.1145/2968219.2968542

[70] Dominik Weber, Alireza Sahami Shirazi, Sven Gehring, Niels Henze, Benjamin Poppinga, Martin Pielot, and Tadashi Okoshi. 2016. Smarttention, Please!: 2nd Workshop on Intelligent Attention Management on Mobile Devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM, New York, NY, USA, 914–917. https://doi.org/10.1145/2957265.2965025

[71] Paweł W. Wozniak, Lars Lischke, Sven Mayer, Andreas Preikschat, Markus Schweizer, Ba Vu, Carlo von Molo, and Niels Henze. 2017. Understanding Work in Public Transport Management Control Rooms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17 Companion)*. ACM, New York, NY, USA, 339–342. https://doi.org/10.1145/3022198.3026341

[72] Richard D Wright and Lawrence M Ward. 2008. *Orienting of attention.* Oxford University Press.

[73] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2018. Studying Eye Movements As a Basis for Measuring Cognitive Load. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW095, 6 pages. https://doi.org/10.1145/3170427.3188628

[74] Janez Zaletelj and Andrej Košir. 2017. Predicting students' attention in the classroom from Kinect facial and body features. *EURASIP Journal on Image and Video Processing* 2017, 1 (01 Dec 2017), 80. https://doi.org/10.1186/s13640-017-0228-8

[75] Qiushi Zhou, Joshua Newn, Namrata Srivastava, Tilman Dingler, Jorge Goncalves, and Eduardo Velloso. 2019. Cognitive Aid: Task Assistance Based On Mental Workload Estimation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW2315, 6 pages. https://doi.org/10.1145/3290607.3313010