Faces of Focus: A Study on the Facial Cues of Attentional States

Ebrahim Babaei
The University of Melbourne
Melbourne, Australia
ebabaei@student.unimelb.edu.au

Namrata Srivastava
The University of Melbourne
Melbourne, Australia
namratas@unimelb.edu.au

Joshua Newn
The University of Melbourne
Melbourne, Australia
newnj@unimelb.edu.au

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

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

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

ABSTRACT
Automatically detecting attentional states is a prerequisite for designing interventions to manage attention—knowledge workers' most critical resource. As a first step towards this goal, it is necessary to understand how different attentional states are made discernible through visible cues in knowledge workers. In this paper, we demonstrate the important facial cues to detect attentional states by evaluating a data set of 15 participants that we tracked over a whole workday, which included their challenge and engagement levels. Our evaluation shows that gaze, pitch, and lips part action units are indicators of engaged work; while pitch, gaze movements, gaze angle, and upper-lid raiser action units are indicators of challenging work. These findings reveal a significant relationship between facial cues and both engagement and challenge levels experienced by our tracked participants. Our work contributes to the design of future studies to detect attentional states based on facial cues.

Author Keywords
Attentional state; facial expression; engagement; challenge; focus

CCS Concepts
•**Human-centered computing** → **User models;**

INTRODUCTION
In their book *The Attention Economy*, Davenport and Beck argue that "understanding and managing attention is now the single most important determinant of business success" [11]. The authors further emphasize that attention is a scarce and valuable asset, which, when in deficit, may cause serious psychological and organizational consequences. Therefore, managing and making the optimal use of this limited resource is vital, especially when it comes to users such as knowledge workers. Knowledge workers, as defined by Janz et al., are *"high-level employees who apply theoretical and analytical knowledge, acquired through formal education"* [25]. They are described as the most valuable asset for companies in the modern world and are key to economic growth [18]. As they are the most expensive type of workers to employ, it is therefore vital to devise approaches to help them understand and manage their own attention.

An approach for managing individuals' attention resource is to allocate the right amount of attention to each task and direct focused attention to most *profitable* tasks [11]. The first step in applying this approach is to detect the perceived level of attention by knowledge workers. For each unique task that a knowledge worker performs, they require a certain level of attention. As the awareness about attentional states can help attention management systems find opportune moments for each task, these systems can potentially save parts of their focused attention resource. The most straightforward way to measure attention is to ask individuals directly [11]. However, this method is limited to the number and quality of individuals' responses and can cause interruptions. Therefore, finding new ways to continuously measure attentional states unobtrusively is an important issue, warranting further exploration.

The emergence of physiological sensors has enabled researchers to gain a better understanding of people's mental and attentional states. In the last decade, technological advances have made these sensors more available and affordable. In the current HCI literature, the detection of facial expressions from such sensors has been the subject of interest for detecting mental states of users [13, 22, 46]. Lately, facial expression recognition technology and applications have become increasingly available, some examples include: OMRON [1], FaceReader [2],

ACM ISBN 978-1-4503-6708-0/20/04 ..$15.00.
http://dx.doi.org/10.1145/3313831.3376566

[1] https://www.components.omron.com/mobile/hvc_p2/
[2] https://www.noldus.com/human-behavior-research/products/facereader

OpenFace [6], AFFDEX [37] and IntraFace [12]. These products can record a wide range of features for analysis from snapshots of an ordinary web camera, including facial expressions, emotions, action units, pupil size, and facial landmarks. Their affordability, availability, and unobtrusiveness make them a promising technology for in-situ recognition of users' mental states. Given this potential, we wanted to explore whether the same tools could be used to find cues that hint at knowledge workers' attentional states. Since knowledge work covers a broad range of jobs, we recruited academic researchers as our use case because of the diversity of their tasks, availability, and importance of attention management in their job due to the high workload they experience [24, 27, 49].

The preliminary challenge in using facial cues for automated attentional state detection lies in the selection of potential predictors. Hence, this paper is aimed at narrowing down the set of potential features by investigating the correlations between facial cues and attentional states. We describe the data collection of 15 recruited researchers who worked for a day in a real workplace while cameras captured their facial cues. We investigate the facial cues in relation to their attentional states reported. Our results highlight potential indicators to guide future works on building classifiers for classifying attentional states using facial cues.

RELATED WORK

In this section, we present three keys areas in which we have situated our research: (1) attentional state theories, (2) attentional state measurement with physiological sensors, (3) and recognition of attentional states using facial expressions.

Understanding Attentional States

The literature contains a wealth of work on topics related to attentional states, such as user engagement [42], mindfulness [29], cognitive absorption [3], and flow [10]. Many theories have attempted to provide definitions for attentional states, but a comprehensive definition covering all of its diverse aspects is still lacking [39]. Schaufeli et al. [47] performed a study to define engagement. The authors defined engagement as the opposite of burnout and studied their hypothesis on two groups of participants, undergraduate students, and company employees. Based on their results, engagement is negatively correlated with burnout, and considered this negative correlation as the verification for their definition.

Mindfulness is another well-known concept that describes attentional state, and is defined as adjusting attention to focus on and awareness of the current activity [7]. Mindfulness is characterized by three axioms: intention, attention, and attitude [55]. Mindfulness approaches focus on *mental training* to reduce the cognitive costs of distractions.

Among the theories, *flow* proposed by Csikszentmihalyi is the most well-known theory related to attentional states [10]. This theory explains how a balance between skill level and task difficulty can place a person in an optimal state for work, which leads to enjoyment and absorption in a task on hand. Nakamura et al. claimed that entering and staying in the optimal state of flow requires users' attention, which is not considered

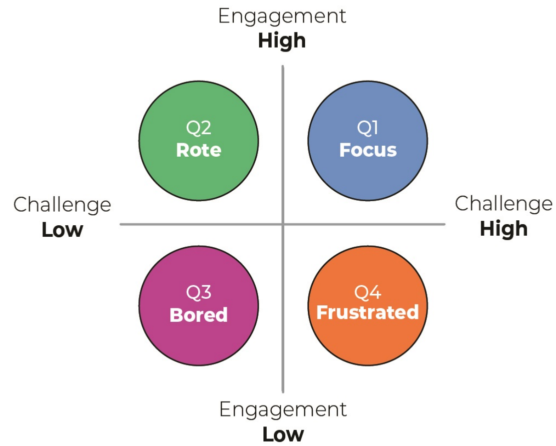

Figure 1: Mark et al.'s framework used in this work to operationalise attentional states.

in flow theory and the lack of it may lead to apathy, anxiety, and boredom [38].

Mark et al. [34] proposed a theoretical framework for attentional states which addressed this problem and took engagement into account. Their proposed a framework describes attentional states by two factors: *engagement* and *challenge* [34]. Based on these two factors, their theoretical framework divides attentional states in the workplace into four quadrants—*Focus*, *Rote*, *Frustrated* and *Bored* (Figure 1). They define challenge as *"the amount of mental effort that one must exert to perform an activity."*, and for engagement, they used the definition proposed by Schaufeli et al. [47]: *"a positive fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption."*.

O'Brien and Cairns [39] discussed the relationship between flow, engagement, and challenge. Whereas earlier studies considered engagement as a *subset of flow* [44, 52], they held the idea that flow is a *subset of engagement*. They concluded this based on a study, in which they investigated consequences of flow in engaged users and observed that engagement does not necessarily come with symptoms of flow such as feelings of pleasure or being challenged. They described flow as a positive mental state of which engagement is a necessary prerequisite, while the opposite is not always true.

In this work, we based our study design on Mark et al.'s framework [34]. We chose this framework because it describes attentional states as a continuum, which includes positive parts (e.g focus) and negative parts (e.g. frustration and boredom). In contrast, theories such as flow and mindfulness only focus on the positive side of attentional state. This continuous operationalization provides us with more insight into users' attentional states. We can consider the discrepancy between users' skill levels and task difficulty to infer the amount of mental effort required in a task. This amount of effort is defined as challenge in this model similarly to challenge in flow theory. Finally, because our study involved participants' self-reports, the intuitive and straightforward definitions of attentional states was another contributing factor in our choice.

Attentional State Measurement

Current approaches for measuring attentional states can be divided into two categories. The first category is based on subjective methods such as questionnaires and experience sampling [9]. The second is based on physiological sensing and observation. As we employ both categories in our work, we will review them accordingly in this section.

Subjective Methods

Mark et al. [33] considered engagement to be a dimension of mood and tried to measure the effect of interpersonal interaction, task switching, social media using, and email checking on engagement. To measure this, they employed experience sampling [9] in the form of pop-up questions on participants' screen. They used the same method to measure challenge [34].

O'Brien et al. [41] proposed the user engagement scale (UES), a 31-item questionnaire which measures engagement based on 6 factors: *aesthetic appeal*, *focused attention*, *perceived usability*, *endurability*, *novelty*, and *felt involvement*. In the following evaluation of their questionnaire [40], they claimed that engagement is influenced more by the first three factors and combined the last three factors into a single one (reward), then revised their questionnaire to a shorter 12-item form. However, the applicability of this model to other areas of HCI, particularly for in-situ studies, is questionable for two reasons. First, their definitions are domain-specific for each factor and are not applicable to other areas. Second, they ask a large number of questions to measure engagement, and answering them is too disruptive for in-situ studies.

Subjective methods are easy to use to measure attentional states, but their accuracy is limited to their obtrusiveness nature and users' self-assessments. To overcome these limitations, researchers delved into automated approaches to detect attentional states using physiological signals.

Physiological Sensing-Based Methods

In this section, we first review previous works that used physiological signals to measure attentional states and then present a review on works that measured other related concepts with physiological signals. To the best of our knowledge, research on measuring attentional states, engagement, and challenge using physiological signals in the workplace is scarce. There is, however, a broad range of research on measuring other factors related to attentional states using physiological sensing such as affective states, interruptions, and stress in learning environments and the use of technology.

McDuff et. al [36] proposed a system called *AffectAura* to measure engagement using physiological signals. They used a sensor setup including a webcam, a Microsoft Kinect, a microphone, and an electrodermal activity (EDA) sensor to measure engagement. They packaged their model as a toolkit that can measure engagement continuously. Lalmas et al. [28] reviewed physiological signals that have been used to measure engagement in everyday tasks such as reading the news. Based on their results, eye tracking and mouse activity are the most effective modalities to measure engagement with. DeFalco et al. [14] studied engagement and affective states based on participants' posture in a serious game for military training and compared it with a method based on users' interactions with the system such as mouse and keyboard actions. Their results showed that interaction-based methods outperformed posture-based methods in measuring engagement.

Physiological signals have also been used to measure cognitive load [2, 35, 54]. For instance, McDuff et al. [35] employed the use of bio-markers such as respiration rate, heart rate, heart rate variability, and facial features to measure the cognitive load of participants while participants were exposed to a task with high cognitive demand. They found a correlation between facial expressions and other measured bio-signals. Similarly, Abdelrahman et al. [2] used thermal imagining to measure cognitive load, showing that the difference between the temperature of the forehead and nose to be an indicator of cognitive load. In another study, they combined the thermal camera with eye-tracker to detect four types of attention, namely sustained, alternating, selective, and divided attention [1].

Most of the previous works were conducted in lab environments, where the participant movements were restricted and the tasks were controlled. However, there are only a few studies that took place in real workspaces. For example, Zuger et al. [56] studied interruptibility of knowledge workers in their real workspaces. They conducted a two-week field study with 13 professional software developers and measured their heart rate, heart rate variability, physical activity, and sleep during the day alongside their interaction with personal devices. They found that interaction data can give a better prediction for interruptibility than physiological sensors. Similarly, Di Lascio et al. [15] conducted another study in-situ, which aimed at estimating students' emotional engagement based on their EDA signal. They gathered self-reports and EDA measurements of 24 students over 41 lectures and built a classifier for emotional engagement based on that.

Facial Expression Analysis

Facial expressions have been proven to be particularly useful for investigating affect, largely because of the ubiquity of facial expressions in human experience and the unobtrusiveness of video recording [17, 26]. In many studies, the Facial Action Coding System (FACS), which enumerates possible movements of the human face, is used to manually annotate facial movements that comprise expressions of emotion. FACS is a taxonomy proposed by Ekman and Friesen [19], which describes all the visible movements on the face. It describes each movement as components called *action units* (AU). In total there are 44 AUs, each representing the movement of one or more muscles on the face, eyes or neck that change the appearance or orientation of the face or the eyes (30 related to the face, 14 related to the eyes and the orientation of the face) [19]. Action Units happen both individually and in groups. For instance, happiness includes the presence of two basic AUs—Cheek Raiser (AU6) and Lip Corner Puller (AU12).

Advances in machine learning and computer vision made automatic AU recognition possible and several toolkits have been validated for detecting AUs from frames of an ordinary webcam. OpenFace [6] is one of such application, which is able to detect 18 action units, head poses and gaze angles from face images. It firsts detects faces in a frame and extracts landmarks and appearances of the detected faces. Then, using a classifier and these two sets of features, it predicts the action units and other facial cues. Openface was independently validated on the FERA 2015 challenge [51] dataset and outperformed two baseline approaches in estimating both the intensity and the occurrence of action units [5]. In the remainder of this section, we provide a brief review of previous works that measured factors related to attentional state detection based on facial expressions.

Whitehill et al. [53] studied the correlation of engagement and facial expressions in an online learning environment. They analyzed a set of video clips of students' faces and labeled them with four levels of engagement. They subsequently investigated the correlation between these labels and students' facial expressions. Based on their results, the rotation of face (roll), the up/down tilting of the face (pitch), AU01 (inner brow raiser), AU10 (upper lip raiser), and AU45 (blink) have the highest correlations with engagement. Grafsgaard et al. [21] used facial expressions to automatically predict engagement and frustration. Based on their results, AU04 (brow lowerer) correlates with frustration, and AU01 (inner browse raiser) and AU04 (brow lowerer) are two action units that correlate with endurability, which is an aspect of engagement.

Littlewort et al. [31] investigated spontaneous facial expression features of children during problem-solving. Based on their results, facial expressions during the latency of a response to a problem is an indicator of success in solving that problem. Bosch et. al [8] found expressions related to student engagement, boredom, and frustration in a real classroom. Based on their findings, frustration in the classroom was manifested by motion, AU01 (inner brow raiser) and AU10 (upper lip raiser). They reported changes in distance from the screen, AU17 (chin raiser) and head movements as signs of boredom. For engagement, they reported AU18 (lip pucker) as the only correlating factor in facial expressions.

D'Mello et al. [17] studied positive and negative predictors of boredom and frustration on the face. They employed labelers to rate boredom and frustration in videos from an online tutoring system and used them alongside the self-reported rates to find related expressions to boredom and frustration. Based on their findings, AU01 (inner brow raiser), AU02 (outer brow raiser) and AU14 (dimpler) showed positive correlations with frustration, while AU43 (eye closed) was the only expression correlated with boredom.

The above-mentioned works studied different aspects of attentional states in a learning environment. However, tasks and definitions of engagement and challenge are different in a learning environment compared to other workplaces. Engagement in learning environment associates with lecture, whereas in knowledge work, engagement associates with the task in hand, which can be diverse. On the other hand, research in learning environments mostly focused on specific attentional states such as frustration, but did not address all the attentional states. Therefore, the applicability of these methods in areas other than learning environments such as research tasks is questionable.

METHODOLOGY

The aim of our work is to understand the relationship between facial cues and attentional states of knowledge workers in their place of work unobtrusively. We aimed to detect visible signs of different attentional states on the face to provide necessary knowledge for future works to build classifiers for recognizing attentional states.

As a representative sample, we recruited 15 students (7M/8F) aged between 26 and 34 (M=30.4, SD=2.89) from our university. They were undertaking their doctoral research degree from a range of disciplines, including law, civil engineering, computer science, and environmental science.

From our review of the literature, we have chosen Mark et al.'s [34] theoretical framework, as it was clear and had easy-to-follow definitions for our participants and has been demonstrated to be effective for workplace environments. The model describes attentional states in two dimensions, namely **engagement** and **challenge**, divided into four quadrants: *focus* (highly engaged and challenged), *rote* (engaged but not challenged), *frustrated* (not engaged but highly challenged), and *bored* (neither engaged nor challenged). We report the results of an observational study in which we captured participants' facial snapshots and their self-reported attentional states throughout the day. For each self-reported attentional state, the participants reported their current engagement and challenge level.

To have awareness about tasks our participants did during the study, in addition to the data we mentioned above, we also captured the mouse and keyboard activity of our participants through an application we installed on their device with their permission. This application captured mouse movements and clicks and also keyboard strokes and their timestamps. We obtained ethics approval from the University of Melbourne's ethics committee for our data collection.

Experimental Setup

Figure 2 illustrates our experimental setup. We allocated a desk in one of our research offices and placed an external monitor on it attached with an HD webcam (Logitech C930E), to capture snapshots of participants' faces and a small Palette Gear Core display [3] (45x45 mm) for notifying participants to enter their attentional state levels.

On the left side of the desk, we placed two Palette Gear sliders[4] for experience sampling, one for reporting their engagement level and one for reporting their challenge level. Once both sliders were set, participants could submit their results by pushing a button. The scale of the sliders ranges from 0 to 100. We decided not to use a Likert scale as it is uni-dimensional and only gives a few options to the users. Studies by Weijters et al. (2010) and Rohrmann (2003) have shown that when making a judgment about the response category, respondents make use of the meaning of the labels attached to the category, suppressing their true intentions as it is hard for them to orient themselves to one choice. We minimize these effects by

[3] https://store.palettegear.com/products/palette-core-module

[4] https://store.palettegear.com/

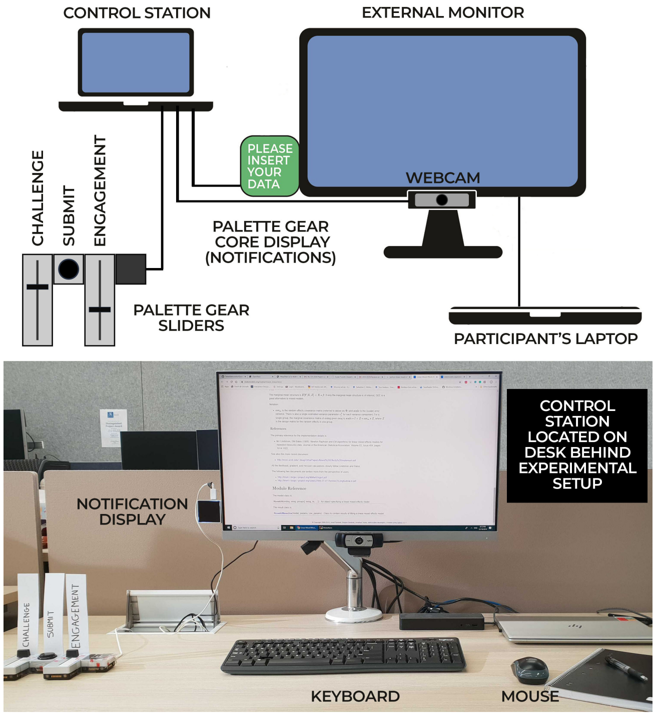

Figure 2: Experimental Setup

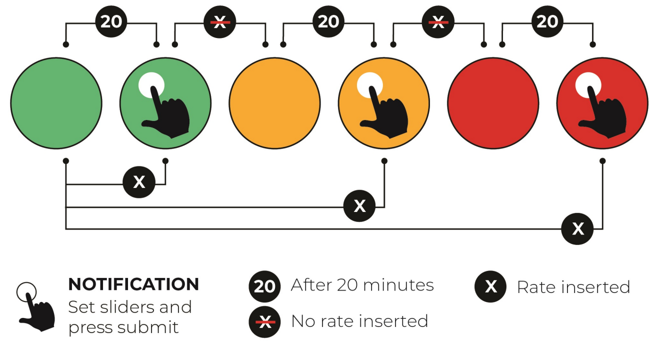

Figure 3: Notification Setup used in our study. Participants were asked to submit their ratings at 20-minutes interval. The color of the notification display changed to yellow, and subsequently red if they forgot to insert the values at these intervals.

training the participants at the beginning of the experiment. We decided to use physical sliders to record the engagement and challenge levels of the participants as this approach interferes less with their daily computer-assisted tasks [48]. We instructed participants to use the value 50 (central position of the sliders) as the threshold between high- and low- engagement/challenge. Also, we found that researchers do some of their activities with materials other than their PC (e.g. reading books or physical papers) and if we used pop-up notifications, notifications were likely to be missed or forced participants to get back to their device to insert their engagement and challenge rates. We connected all the devices to our experimental laptop for control and data logging. As participants were required to bring their laptops for the study, we placed a mouse and keyboard on the desk for them to connect to for their comfort.

Procedure

Upon arrival, we handed participants a plain language statement explaining the purpose of our study with detailed information about the signals we would collect. Then, they signed a consent form and received a short interview for demographics data collection. They also reported their stage and experience in research and general information about how they started their respective day in terms of hours of sleep in the night before, duration of commute to arrive to the office, if they had had breakfast, and any incident that may have had an effect on their mood in the beginning of the day. This information was collected to ensure there are be no confounding factors (due to events like improper sleep, long-commute), which might have affected the mood or fatigue levels of participants.

Then, we showed them a short presentation about the study, in which we introduced the sensors and data we collected and gave detailed information on the framework we used for attentional states and explained the exact definitions of engagement and challenge in that framework to them. To make sure that the participants understood the attentional state model and definitions, we gave them examples of moments a researcher is engaged or challenged and later asked questions about engagement and challenge in different situations to validate their comprehension about these concepts. We repeated this process until participants reported that they fully understood the concept. We also placed a printed material containing definitions of engagement and challenge on the desk in case they needed to refer to them later.

Participants were guided to the allocated work desk, where the researcher showed them the setup, followed by the use of the physical sliders. Following, the researcher assisted with connecting their own laptop to the devices we provided and were asked to commence their work once the researcher had verified that the system is receiving the data. Participants were allowed to leave the study for attending meetings, having breaks such as lunch, coffee, etc. throughout the day and could leave the study at any point.

Every 20 minutes, the notification display prompted participants to report their engagement and challenge level at the moment using the sliders and buttons. The Prompt was in the form of a blinking light and we guided participants in the presentation before the study that they must insert their engagement and challenge level at the moment. They were free to ignore the notifications, which would subsequently disappear after one minute. We labeled these sliders and buttons to make sure they could not be mistakenly perceived as each other. The color of the display would change from green to yellow and then to red if participants did not enter their input. The screen will become green again after they entered the levels, and continue with the next notification. The flow of notification display is depicted in Figure 3.

At the end of the study, we conducted a second interview. In this interview, we asked them questions about how they spent the day (e.g. the tasks they worked on or incidents that may

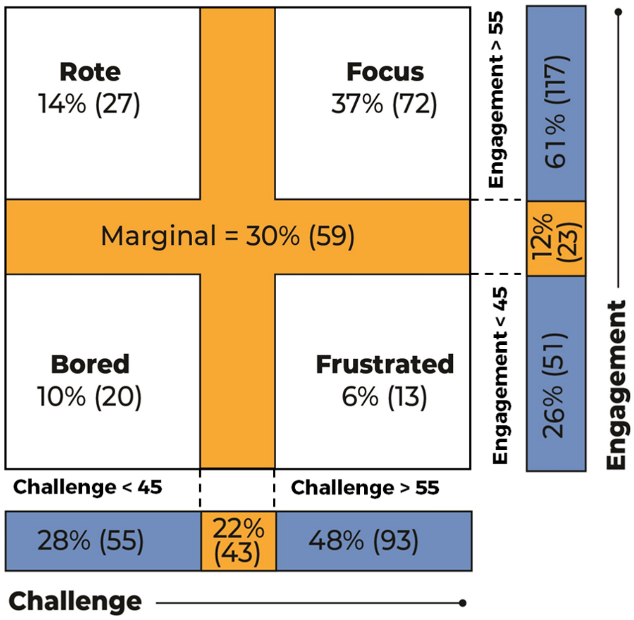

Figure 4: Distribution of self-reported attentional state rates. The region shaded in yellow was discarded for analysis to reduce the misclassification rate among the 4 states.

have affected their attentional state). One complete session of this study lasted approximately eight hours and the starting and finishing time was flexible based on participants' preferences and work routine. We compensated participants with a AUD$40 gift card for their time at the end of the day.

ANALYSIS

We collected 79 hours of face snapshots of participants captured by the webcam while doing their research tasks. In total, participants reported 191 engagement and challenge values (Mean = 12.7, STD = 5.8). On the scales of 0 to 100, engagement rates had an average value of 65 and STD of 23 and challenge rates had an average value of 50 and STD of 25. The distributions are shown in Figure 4.

For our analysis, we divided the self-reported measures of engagement and challenge into two sub-levels—high and low (high engagement: $engagement > 55$, low engagement: $engagement < 45$; high challenge: $challenge > 55$ and low challenge: $challenge < 45$). The $mid\text{-}point - 5$ (45) and $mid\text{-}point + 5$ (55) of the scale was used as the cut-off points to split the participants into these levels. We decided to discard the rates between 45 and 55 for better confidence of the labels. Next, based on Mark et al.'s [34] framework, we labeled our data as focus, rote, bored, and frustrated states. We used these labels as the experienced attentional state of our participants for the self-reported rates.

As the attentional state of a participant might have fluctuated within the 20 minutes, we decided to analyze the data only for the final 60 seconds before each self-reported rate, as we assumed these to better reflect their attentional state. We also discarded the last 15 seconds from the 1-minute observations, considering the amount of time participants took to move toward the slider and set them on the desired rates. To

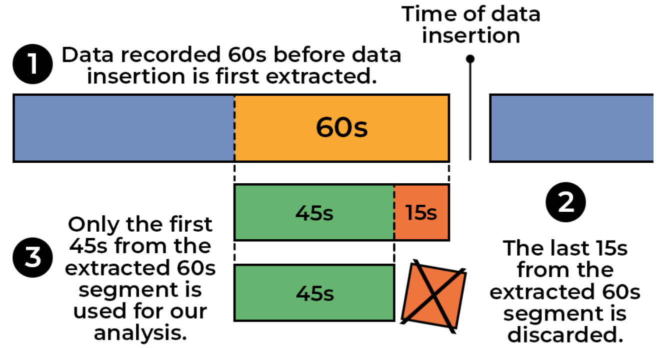

Figure 5: Used and discarded parts of the dataset. For each 20-minutes interval, we used only the initial 45 seconds before the 60 second-interval to perform facial expression analysis.

summarise, we used the initial 45 seconds of the 60 second-interval (before participants submitted their self-reported rate) to perform facial expression analysis. Figure 5 shows the used and discarded parts of our data.

Feature Extraction

We extracted facial expression features from each frame with *OpenFace* [6]. These features include the presence and intensity of Action Units (AUs), gaze angles, head positions, and head rotations (pitch, roll, yaw). We discarded outputs with less than 95% confidence and interpolated the discarded parts with linear interpolation. We used *interp1d*, a Python package to perform the linear interpolations. To remove noise and outliers from our data, we used a median filter with a window size of 61 for outputs of OpenFace. We prepossessed these outputs and computed features for each 45s sample. Table 1 shows all features we extracted from the data. In total, we used 76 features from OpenFace outputs in our dataset.

RESULTS

To understand the relationship between the two dimensions of attentional states in our dataset, we first performed a correlation test between the rates of engagement and challenge. Then, we studied the relationship between each type of attentional state and the facial cues. Finally, we employed a linear mixed model to test the correlations between facial cues and attentional states as categorized in the framework adopted in our study (Figure 1). We present our results in the following subsections corresponding to each test we performed. We discuss our results from each test in next section under matching subheadings.

Engagement and Challenge

First, to understand the correlation between self-reported engagement and challenge rates, we calculated the Pearson correlation coefficient between the ratings, which yielded a statistically significant positive correlation ($r = 0.4, p < .05$).

Engagement and Facial Cues

We considered self-reported challenge and engagement rates in our dataset as ordinal numbers because although higher orders of rates show higher perceived engagement or challenge, the interval between values is not necessarily equally spaced.

OpenFace	
Features	*Metrics*
Gaze Angle	Movements, average, SD
Head Position	Movements, average, SD
Action Units	
AU01 (Inner Brow Raiser)	
AU02 (Outer Brow Raiser)	
AU04 (Brow Lowerer)	
AU05 (Upper Lid Raiser)	
AU06 (Cheek Raiser)	
AU07 (Lip Tightener)	Percentage of presence of each AU, average intensity, SD of intensity
AU09 (Nose Wrinkler)	
AU10 (Upper Lip Raiser)	
AU12 (Lip Corner Puller)	
AU14 (Dimpler)	
AU15 (Lip Corner Depressor)	
AU17 (Chin Raiser)	
AU20 (Lip Stretcher)	
AU23 (Lip Tightener)	
AU25 (Lips part)	
AU26 (Jaw Drop)	
AU28 (Lip Suck)	
AU45 (Blink)	
Head Rotation	Movements, average, median, SD

Table 1: Extracted features from OpenFace data

Feature	*r*	*p*
Gaze Angle (up/down, avg)	-0.32	0.000
Pitch (avg)	-0.27	0.000
AU14 (SD of intensity)	-0.25	0.002
AU25 (avg of intensity)	0.26	0.000
AU20 (avg of intensity)	0.28	0.000

Table 2: Correlation of self-reported engagement rates and facial expressions

For instance, we cannot state that the difference in challenge rates between 60 and 40 is equal to the difference between 40 and 20, as participants report these values. Therefore, we applied the Spearman correlation test for determining monotonic correlations between the self-reported rates and extracted features. To avoid false-positive significant correlations, we corrected our p-values with FDR-BH test. We chose this test over other alternatives (e.g. Bonferroni) because it has greater power to find true positives [16], which is the goal of this exploratory study. We present the highest correlations between engagement and facial expressions in Table 2.

Based on our results, engagement is positively correlated with AU20 (lip stretcher) and AU25 (lips apart), and negatively correlated with gaze angle, pitch, and AU14 (dimpler). Furthermore, based on t-test results, pitch and gaze angle are significantly different in high and low engagement(pitch: $t = 4.2, p < 0.00005$ and gaze angle: $t = 2.8, p < 0.004$). In Figure 8, we present boxplots of pitch and gaze angles for high and low engagement levels.

Feature	*r*	*p*
Gaze Angle (up/down, avg)	-0.34	0.000
Pitch (avg)	-0.25	0.001
AU28 (avg of presence)	-0.25	0.001
Gaze Angle (up/down, movements)	-0.21	0.007
Gaze Angle (right/left, movements)	-0.20	0.007
AU5 (avg of presence)	0.23	0.003

Table 3: Correlation of self-reported challenge rates and facial expressions

Challenge and Facial Cues

We applied the same correlation tests on challenge, and the results are presented in Table 3. Based on our results, challenge is negatively correlated with the gaze angle direction, pitch, and AU28 (lip suck); and positively correlated with AU5 (upper lid raiser).

Attentional States and Facial Cues

As mentioned in the previous section, we divided the self-reported rates to four quadrants based on their engagement and challenge rates (focus, rote, bored, and frustrated). In this part of our analysis, we used these labels as attentional states and evaluated their impact on facial cues. For this purpose, we applied a linear mixed model with participants as the random effect and attentional states as the fixed effect on our features. We adjusted the pvalues of the linear mixed model using FDR-BH test. We discarded the data points in the frustrated quadrant from our analysis because we only observed 13 self-reported rates there. We present the results of this evaluation in Table 4.

Feature	*Atten. State*	*Coeff.*	*SE*	*t*	*p*
AU01 (Avg. Intensity)	Focus	0.13	0.03	4.69	0.000
AU02 (Avg. Intensity)	Rote	-0.13	0.04	-3.62	0.003
	Focus	-0.07	0.02	-3.30	0.009
AU12 (Avg. Intensity)	Rote	-0.11	0.03	-3.30	0.009
	Focus	-0.11	0.02	-5.51	0.000
AU20 (Avg. Intensity)	Focus	0.09	0.02	3.93	0.001
Head Pose (Z direction)	Rote	-0.29	0.09	-3.20	0.013

Table 4: Linear mixed model results on attentional states and facial cues

Based on our results, Focus has a significant effect on AU1 (inner brow raiser), AU2 (outer brow raiser), AU12 (lip corner puller), and AU20 (lip stretcher). Rote has a significant effect on AU2 (outer brow raiser) and AU12 (lip corner puller) and head position (distance between head and display). To justify our result for the relationship of head position and attentional state, we needed to analyze the interactions our participants did with their computer during the study in different attentional states. T-test on mouse movements yielded statistically significant results between rote and focus ($r = 3.03, p < .05$)

and between rote and bored ($r = 3.55, p < .005$). We presented the boxplots of mouse movements and keyboard and mouse streaks in different states in Figure 6.

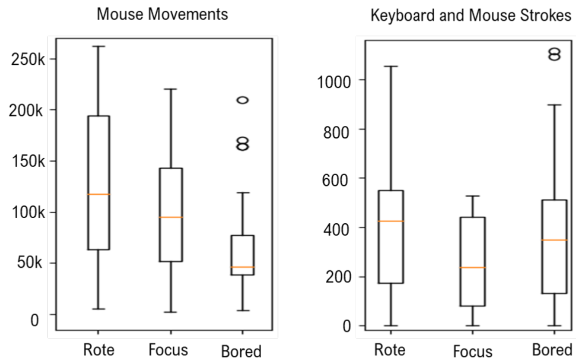

Figure 6: Participants' interactions in different levels of attentional state

DISCUSSION

In this section, we discuss the relationship between facial cues, engagement, challenge, and attentional state in details based on our results and literature.

Engagement and Challenge

As previously discussed in the related work section, the balance between skill and difficulty of a task determines whether an individual will engage in a state of flow. Based on two different views in attentional state theories, engagement can be the result of a state of flow [44, 52] or its pre-requisite [39]. In any case, there should be a positive correlation between engagement and the appropriate balance between skill and difficulty. We argue that this balance can be interpreted as *challenge* in Mark et al.'s framework [34]. For instance, workers should be more engaged to the task on hand when it is challenging enough to require their focus, but not too challenging so that they feel frustrated. Therefore, we expected to see a positive correlation between engagement and challenge when the challenge is not beyond an individual's skill level. As most tasks of researchers are self-initiated and usually not beyond their skills, we expected to see that a rise in challenge would lead to a rise in engagement [38]. Our result of the correlation test between self-reported challenge and engagement supports our expectations.

Engagement and Facial Cues

Grafsgaard et al. [21] found a positive correlation between AU14 (dimpler) and frustration. Because frustration happens in a low engagement state, the negative correlation between engagement and AU14 (dimpler) in our results is consistent with their findings.

The web camera was mounted to the bottom of the screen to capture the faces of participants better. To keep our participants in the field of view of the web camera, we placed the screen a bit higher, according to our participants' face height. Therefore, our participants had to look a bit upward to see the screen, and looking upward leads to negative gaze angles in our data (see Figure 7). On the other hand, because the camera captured participants' faces from the bottom of

Figure 7: Position of cameras in the setup and a frame of a participant looking at the screen

the screen, the rotation of the face in the vertical axis (pitch) was negative when participants looked at the screen, and it was positive when they looked away (downward). Thus, we can explain the negative correlation between the gaze angles, pitch, and engagement with the fact that engaged participants tended to focus more on the screen whereas disengaged participants had a lower tendency to focus on the screen. Therefore, engaged participants showed lower average pitch and gaze angles, which resulted in negative correlations. As the Figure 8 shows, highly engaged participants had lower pitch and gaze angle values than low-engagement participants.

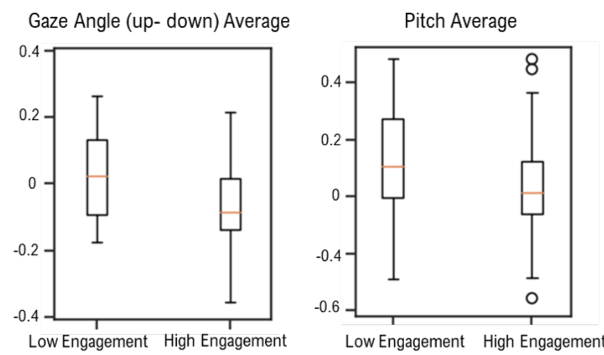

Figure 8: Boxplots of average pitch and gaze angle in high and low engagement

Our results showed that AU25 (lips apart) and AU20 (lip stretcher) were positively correlated with engagement. The positive correlations suggest that these are behaviors indicative of high engagement, which means that these cues are more frequently observed on the faces of engaged people, as they are absorbed in their tasks. Rozin and Cohen [45] previously reported mouth opening action units such as lips parting as signs of concentration. We illustrate a sample of this expression in Figure 9. AU20 (lip stretcher) is another action unit positively correlated with engagement, but it is often impeded by users' hands covering the mouth [32], which frequently occurred in our dataset.

Challenge and Facial Cues

In addition to the negative correlations between challenge and pitch and gaze angles, challenge rates showed negative correlations with gaze angle movements in both directions as well. It shows not only that challenged researchers tend

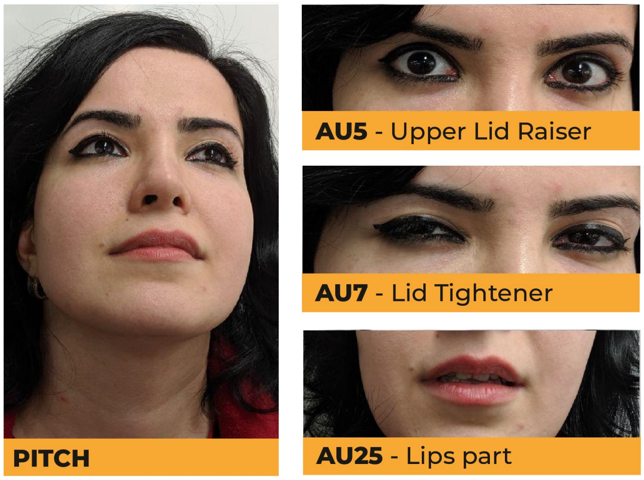

Figure 9: Samples of facial expressions mentioned in our results

to focus on the screen more than researchers who are not challenged, but also that they tend to shift their attention less and fixate their focus on the screen. As challenging tasks require information processing, this finding is consistent with previous studies in eye-tracking [43, 50], which showed that processing difficult information leads to longer fixations and fewer saccade movements. As such, a negative correlation between gaze movements and challenge was expected.

Average presence (frequency) of AU5 (upper lid raiser) showed a significant correlation with challenge. Upper lid raiser is an important facial expression in biology, as this action unit increases the field of view [30] and gives better visual discrimination and sensory advantage to an observer [30, 23]. From this, we can suggest that challenged people perform this action unit to perceive information better, whereas people who are not challenged use this expression less because they experience lower demand for perceiving information.

Attentional States and Facial Cues

Grafsgaard et al. [20] claimed a positive correlation of engagement and AU1 (inner brow raiser). Similar to AU5 (upper lid raiser), this facial cue is an eye-widening type of action unit, which has a function of giving sensory advantage to an observer. From this, we hypothesize that as a focused person must deal with a higher load of information processing, they unconsciously activate this action unit. AU20 (lip stretcher), which previously showed correlations with engagement, is also significantly impacted by focus. This was expected since state of focus happens in high engagement.

Because in our desk setup the camera was at the center of the bottom edge of the screen, we can consider head pose in the Z direction to be an estimate of the distance between participants' heads from the screen (positive Z is away from camera). To understand the reason behind the significant correlation between rote state and head pose in the Z direction, we re-visited the video recordings of the participants during the study. Based on our observations, the interaction with the computer had a major effect on the changes of participants' head distance from the screen. Participants moved closer to the screen whenever they wanted to use the keyboard or mouse to interact with the computer. On the other hand, participants in rote state tended to continuously use their mouse and keyboard, whereas participants in the focus state leaned back on their chair while only intermittently leaning forward and performing some interactions, before leaning back again. Participants in a bored state also tended to demonstrate this behavior. Therefore, participants in rote state interacted more with their computers compared to participants in the other two states. We believe that the reason behind this behavior is that participants in a focused state had to deal with more challenging problems. Therefore, they needed more time to think about the problem, during which they may lean back in a relaxing position.

Conversely, participants in a rote state faced lower levels of challenge which do not require the same amount of time to keep them from continuously performing the task. Bored participants also tended to do fewer interactions because they are not engaged in the task on hand. To confirm our hypothesis, we plotted the interactions that our participants performed with their computers during the study in different attentional states. We present boxplots of mouse movements and keyboard and mouse streaks in different states in Figure 6. We can see the difference between the number of interactions in rote compared to other states, which shows that focused and bored participants had fewer interactions with their computers compared with participants in rote state. We also applied t-test on mouse movements between rote and focus ($r = 3.03, p < .05$) and between rote and bored ($r = 3.55, p < .005$) quadrants which showed a statistically significant difference.

LIMITATIONS

Before this study, it was unrealistic to form a set of hypotheses with attentional states based on any specific set of facial cues. There was not enough support in the literature for any decision on the selection of features from the vast amount of features available, which also limits the statistical power of our exploratory test due to potential Type I errors. Thus, we do not claim any causal relationship based on our significant correlation results. Instead, our results highlight those features as potential indicators to guide future works on building classifiers for attentional states using facial cues.

The accuracy of facial expression recognition is an inherent limitation in this study. The human face is a non-rigid shape with substantial individual differences, which severely affects the performance of toolkits [4]. Further, we performed our study in a workplace where we did not have control over many factors, such as lighting and movement. These factors also influence the accuracy of these tools. Therefore, although we did our best to reduce the noise in our dataset, it may still contain noisy data that may have influenced our results.

The second category of limitation is due to the design of our study. During the instructional stage, before collecting data, we instructed and expected participants to follow their daily routines, which should allow our observation of their natural behaviors. However, participants reported that they felt 'more engaged' during the study than usual. We believe that this is due to the Hawthorne effect of the presence of sensors and the awareness of being observed by them. Consequently, this made our dataset unbalanced as we had few samples in frustration state, which subsequently affected our analysis as we had to discard the data in frustration state as it was not populated enough to be representative of the state. Then again, we did train the participants at the beginning of the study to hold a consistent understanding of engagement and challenge. Nevertheless, their self-reported rates are still dependent on how they interpreted the concepts, and this may have introduced inconsistencies to our labels and affected our results.

Participants rated their attentional states every 20 minutes during the study to record their attentional states. However, fluctuations in attentional states can happen in a matter of seconds. As such, we discarded many of the facial expression samples and only analyzed the data at the time close to self-reported rates. In addition to losing data, not having continuous self-reported rates prevented us from studying immediate facial responses reflecting change of attentional state.

As a sample of knowledge workers, we recruited academic researchers as participants in our study. However, knowledge work covers an extensive range of jobs. Therefore, the fact that academic researchers cannot be representative of all types of knowledge workers is another limitation in this study. Despite these limitations, our work showed very promising results for the estimation of attentional state in-the-wild, only based on available and affordable facial expression recognition tools. The outcomes of our work contribute to in-situ classification of attentional states in the future.

The existing literature on the function of action units is very limited. Previous works mostly investigated the presence of action units, but their absence and intensity variations were unattended. This limits our discussions on facial cues to the presence of a few action units, while ignoring other information that facial cues may convey.

FUTURE WORK

We foresee three areas for future work. First, there is a possibility that the correlation between attentional states and other facial expressions were not detected due to the limited accuracy of the OpenFace software. Hence, it would be interesting to compare the results from other facial expression recognition tools using our data set. Second, with regards to classifying attentional states of researchers for task assistance, we can proceed with two approaches. The first approach is to study other psychological sensors to measure attentional states. This knowledge can lead to stronger predictions for assisting knowledge workers in managing their attention.

We designed our study based on Mark et al.'s [34] attentional state framework. The variance in the number of self-reported rates and the small size of dataset caused lack of data in frustration quadrant of this model. We were not able to analyze facial expressions in this part of the attentional state framework. Therefore, the second approach is to label our data in another way rather than self-reported rates (e.g. inferring attentional state from application usage), from which we will be able to produce analysis on the frustration quadrant as well.

Lastly, another potential direction of future work is to study the Hawthorne bias caused by physiological sensors, which was reported multiple times by our participants. We can investigate the ethical concern of the presence of sensors in the workplace and how that may affect users' mental states.

CONCLUSION

In this paper, we investigated the visible signs of attentional states on researchers' facial expressions in situ. We used Mark et al.'s theoretical framework, which describes attentional states with two main factors: engagement and challenge. We conducted an observational study with 15 researchers who worked a day at our office while a camera captured their faces. Based on our findings, engaged and challenged researchers focus on the screen more, which resulted in a correlation between engagement and facial cues, including gaze angle and vertical rotation of face (pitch). Besides, as an indicator of absorption, engaged researchers express lips-apart behaviors as well. Challenged researchers tend to raise their eyelids as this behavior helps them to bring in more information. They also move their gaze less to concentrate on their tasks. Furthermore, our analysis revealed that the attentional state of rote has a significant effect on the distance between the head and screen. Our findings inform future studies of measuring attentional state with facial cues in the real workplace.

ACKNOWLEDGMENTS

Dr Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Researcher Award (Project Number: DE180100315) funded by the Australian Government. Ebrahim Babaei and Namrata Srivastava are both recipients of the Melbourne Research Scholarship. Joshua Newn and Qiushi Zhou are both supported by scholarships under the Australian Commonwealth Government Research Training Program.

REFERENCES

[1] 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 (Sept. 2019), 27 pages. DOI: http://dx.doi.org/10.1145/3351227

[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. DOI: http://dx.doi.org/10.1145/3130898

[3] Ritu Agarwal and Elena Karahanna. 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. 24, 4 (2000), 665–694. DOI: http://dx.doi.org/10.2307/3250951

[4] Tadas Baltrusaitis. 2014. Automatic facial expression analysis. (2014). DOI: http://dx.doi.org/10.17863/cam.16382

[5] T. Baltrusaitis, M. Mahmoud, and P. Robinson. 2015. Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 06. 1–6. DOI: http://dx.doi.org/10.1109/FG.2015.7284869

[6] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 59–66. DOI:http://dx.doi.org/10.1109/FG.2018.00019

[7] Scott R. Bishop, Mark Lau, Shauna Shapiro, Linda Carlson, Nicole D. Anderson, James Carmody, Zindel V. Segal, Susan Abbey, Michael Speca, Drew Velting, and Gerald Devins. 2004. Mindfulness: A Proposed Operational Definition. *Clinical Psychology: Science and Practice* 11, 3 (2004), 230–241. DOI: http://dx.doi.org/10.1093/clipsy.bph077

[8] Nigel Bosch, Sidney K. D'mello, Jaclyn Ocumpaugh, Ryan S. Baker, and Valerie Shute. 2016. Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms. *ACM Trans. Interact. Intell. Syst.* 6, 2, Article 17 (July 2016), 26 pages. DOI: http://dx.doi.org/10.1145/2946837

[9] S. Consolvo and M. Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (April 2003), 24–31. DOI:http://dx.doi.org/10.1109/MPRV.2003.1203750

[10] Mihaly. Csikszentmihalyi. 1990. *Flow : the psychology of optimal experience / Mihaly Csikszentmihalyi* (1st ed. ed.). Harper & Row New York. xii, 303 p. ; pages.

[11] Thomas H Davenport and John C Beck. 2001. *The attention economy: Understanding the new currency of business*. Harvard Business Press.

[12] F. De la Torre, W. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn. 2015. IntraFace. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 1. 1–8. DOI: http://dx.doi.org/10.1109/FG.2015.7163082

[13] Fiorella de Rosis, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio, and Berardina De Carolis. 2003. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies* 59, 1 (2003), 81 – 118. DOI:http://dx.doi.org/https://doi.org/10.1016/S1071-5819(03)00020-X Applications of Affective Computing in Human-Computer Interaction.

[14] Jeanine A. DeFalco, Jonathan P. Rowe, Luc Paquette, Vasiliki Georgoulas-Sherry, Keith Brawner, Bradford W. Mott, Ryan S. Baker, and James C. Lester. 2018. Detecting and Addressing Frustration in a Serious Game for Military Training. *International Journal of Artificial Intelligence in Education* 28, 2 (01 Jun 2018), 152–193. DOI:http://dx.doi.org/10.1007/s40593-017-0152-1

[15] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive Assessment of Students' Emotional Engagement During Lectures Using Electrodermal Activity Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 103 (Sept. 2018), 21 pages. DOI:http://dx.doi.org/10.1145/3264913

[16] Angel P. Diz, Antonio Carvajal-Rodríguez, and David O. F. Skibinski. 2011. Multiple Hypothesis Testing in Proteomics: A Strategy for Experimental Work. *Molecular & Cellular Proteomics* 10, 3 (2011). DOI: http://dx.doi.org/10.1074/mcp.M110.004374

[17] Sidney K. D'Mello, Scotty D. Craig, and Art C. Graesser. 2009. Multimethod Assessment of Affective Experience and Expression During Deep Learning. *Int. J. Learn. Technol.* 4, 3/4 (Oct. 2009), 165–187. DOI: http://dx.doi.org/10.1504/IJLT.2009.028805

[18] Peter F. Drucker. 1999. Knowledge-Worker Productivity: The Biggest Challenge. *California Management Review* 41, 2 (1999), 79–94. DOI: http://dx.doi.org/10.2307/41165987

[19] Paul Ekman and Wallace V. Friesen. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior* 1, 1 (01 Sep 1976), 56–75. DOI: http://dx.doi.org/10.1007/BF01115465

[20] Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester. 2013a. Automatically Recognizing Facial Expression: Predicting Engagement and Frustration. In *Proceedings of the International Conference on Educational Data Mining (EDM)*. 43–50. http://cise.ufl.edu/research/learndialogue/pdf/LearnDialogue-Grafsgaard-EDM-2013.pdf

[21] Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester. 2013b. Automatically Recognizing Facial Indicators of Frustration: A Learning-centric Analysis. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII '13)*. IEEE Computer Society, Washington, DC, USA, 159–165. DOI: http://dx.doi.org/10.1109/ACII.2013.33

[22] M. Jabon, J. Bailenson, E. Pontikakis, L. Takayama, and C. Nass. 2011. Facial expression analysis for predicting unsafe driving behavior. *IEEE Pervasive Computing* 10, 4 (April 2011), 84–95. DOI: http://dx.doi.org/10.1109/MPRV.2010.46

[23] Rachael E. Jack, Oliver G.B. Garrod, and Philippe G. Schyns. 2014. Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology* 24, 2 (Jan. 2014), 187–192. DOI: http://dx.doi.org/10.1016/j.cub.2013.11.064

[24] Jerry A. Jacobs and Sarah E. Winslow. 2004. Overworked Faculty: Job Stresses and Family Demands. *The ANNALS of the American Academy of Political and Social Science* 596, 1 (2004), 104–129. DOI: http://dx.doi.org/10.1177/0002716204268185

[25] Brian D Janz, Jason A Colquitt, and Raymond A Noe. 1997. Knowledge worker team effectiveness: The role of autonomy, interdependence, team development, and contextual support variables. *Personnel psychology* 50, 4 (1997), 877–904.

[26] Ashish Kapoor, Selene Mota, and Rosalind W. Picard. 2001. Towards a Learning Companion that Recognizes Affect. In *Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium 2001, North Falmouth, MA* (emotional and intelligent ii: the tangled knot of social cognition, aaai fall symposium 2001, north falmouth, ma, 2nd-4th november, 2001 ed.). https://www.microsoft.com/en-us/research/publication/towards-learning-companion-recognizes-affect/

[27] Gail Kinman, Fiona Jones, and Russell Kinman. 2006. The Well-being of the UK Academy, 1998-2004. *Quality in Higher Education* 12, 1 (2006), 15–27. DOI: http://dx.doi.org/10.1080/13538320600685081

[28] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2014. Measuring User Engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 6, 4 (2014), 1–132. DOI: http://dx.doi.org/10.2200/S00605ED1V01Y201410ICR038

[29] Ellen J Langer and Christelle T Ngnoumen. 2017. Mindfulness. In *Positive Psychology*. Routledge, 95–111.

[30] Daniel H. Lee, Joshua M. Susskind, and Adam K. Anderson. 2013. Social Transmission of the Sensory Benefits of Eye Widening in Fear Expressions. *Psychological Science* 24, 6 (April 2013), 957–965. DOI: http://dx.doi.org/10.1177/0956797612464500

[31] G. C. Littlewort, M. S. Bartlett, L. P. Salamanca, and J. Reilly. 2011. Automated measurement of children's facial expressions during problem solving tasks. In *Face and Gesture 2011*. 30–35. DOI: http://dx.doi.org/10.1109/FG.2011.5771418

[32] Marwa Mahmoud and Peter Robinson. 2011. Interpreting Hand-Over-Face Gestures. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 248–255.

[33] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2014a. Capturing the Mood: Facebook and Face-to-face Encounters in the Workplace. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1082–1094. DOI: http://dx.doi.org/10.1145/2531602.2531673

[34] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, and Paul Johns. 2014b. Bored Mondays and Focused Afternoons: The Rhythm of Attention and Online Activity in the Workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3025–3034. DOI: http://dx.doi.org/10.1145/2556288.2557204

[35] D. McDuff, S. Gontarek, and R. Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2957–2960. DOI: http://dx.doi.org/10.1109/EMBC.2014.6944243

[36] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: An Intelligent System for Emotional Memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 849–858. DOI: http://dx.doi.org/10.1145/2207676.2208525

[37] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3723–3726. DOI: http://dx.doi.org/10.1145/2851581.2890247

[38] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. *The Concept of Flow*. Springer Netherlands, Dordrecht, 239–263. DOI: http://dx.doi.org/10.1007/978-94-017-9088-8_16

[39] Heather O'Brien and Paul Cairns. 2016. *Why Engagement Matters: Cross-Disciplinary Perspectives and Innovations on User Engagement with Digital Media* (1st ed.). Springer Publishing Company, Incorporated.

[40] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28 – 39. DOI: http://dx.doi.org/https://doi.org/10.1016/j.ijhcs.2018.01.004

[41] Heather L. O'Brien and Elaine G. Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69. DOI: http://dx.doi.org/10.1002/asi.21229

[42] Christopher Peters, Ginevra Castellano, and Sara de Freitas. 2009. An Exploration of User Engagement in HCI. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots (AFFINE '09)*. ACM, New York, NY, USA, Article 9, 3 pages. DOI: http://dx.doi.org/10.1145/1655260.1655269

[43] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422. DOI: http://dx.doi.org/10.1037/0033-2909.124.3.372

[44] P. Romero and E. Calvillo-Gamez. 2014. An Embodied View of Flow. *Interacting with Computers* 26, 6 (Nov 2014), 513–527. DOI: http://dx.doi.org/10.1093/iwc/iwt051

[45] Paul Rozin and Adam B. Cohen. 2003. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* 3, 1 (2003), 68–75. DOI: http://dx.doi.org/10.1037/1528-3542.3.1.68

[46] Zhanna Sarsenbayeva, Denzil Ferreira, Niels van Berkel, Chu Luo, Mikko Vaisanen, Vassilis Kostakos, and Jorge Goncalves. 2017. Vision-based Happiness Inference: A Feasibility Case-study. 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, 494–499. DOI:http://dx.doi.org/10.1145/3123024.3124438

[47] Wilmar B. Schaufeli, Marisa Salanova, Vicente González-romá, and Arnold B. Bakker. 2002. The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *Journal of Happiness Studies* 3, 1 (01 Mar 2002), 71–92. DOI:http://dx.doi.org/10.1023/A:1015630930326

[48] Namrata Srivastava, Eduardo Velloso, Jason M. Lodge, Sarah Erfani, and James Bailey. 2019. Continuous Evaluation of Video Lectures from Real-Time Difficulty Self-Report. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 586, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300816

[49] Wei Sun, Hui Wu, and Lie Wang. 2011. Occupational Stress and Its Related Factors among University Teachers in China. *Journal of Occupational Health* advpub (2011), 1106160191–1106160191. DOI: http://dx.doi.org/10.1539/joh.10-0058-OA

[50] Sarah Uzzaman and Steve Joordens. 2011. The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition* 20, 4 (Dec. 2011), 1882–1886. DOI: http://dx.doi.org/10.1016/j.concog.2011.09.010

[51] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn. 2015. FERA 2015 - second Facial Expression Recognition and Analysis challenge. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 06. 1–8. DOI: http://dx.doi.org/10.1109/FG.2015.7284874

[52] Jane Webster and Jaspreet S. Ahuja. 2006. Enhancing the Design of Web Navigation Systems: The Influence of User Disorientation on Engagement and Performance. 30, 3 (2006), 661–678. DOI: http://dx.doi.org/10.2307/25148744

[53] J. Whitehill, Z. Serpell, Y. Lin, A. Foster, and J. R. Movellan. 2014. The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions. *IEEE Transactions on Affective Computing* 5, 1 (Jan 2014), 86–98. DOI: http://dx.doi.org/10.1109/TAFFC.2014.2316163

[54] 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. DOI: http://dx.doi.org/10.1145/3290607.3313010

[55] Jon Kabat Zinn. 1994. Wherever you go, there you are: Mindfulness meditation in everyday life. *Hyperion* (1994), 78–80.

[56] Manuela Züger, Sebastian C. Müller, André N. Meyer, and Thomas Fritz. 2018. Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 591, 14 pages. DOI: http://dx.doi.org/10.1145/3173574.3174165