Bachelor Thesis

August 22, 2021

Emotion Sensing and Self-Awareness

Remy Egloff

of Wettingen, Switzerland (17-705-823)

supervised by

Prof. Dr. Thomas Fritz
Roy Rutishauser
Dr. André Meyer

University of Zurich[UZH]

HASEL

Bachelor Thesis

Emotion Sensing and Self-Awareness

Remy Egloff

University of Zurich[UZH]

Bachelor Thesis

Author: Remy Egloff, remy.egloff@bf.uzh.ch

Project period: 22.02.2021 - 22.08.2021

Human Aspects of Software Engineering Lab
Department of Informatics, University of Zurich

Acknowledgements

I would like to thank all members of the Human Aspects of Software Engineering Lab (HASEL) at the University of Zurich for their support during this thesis. Especially to Roy Rutishauser, Dr. André Meyer and Prof. Dr. Thomas Fritz for their advice. Their continuous feedback and new inputs were a great help throughout the process of this thesis. Moreover, many thanks to all participants of the user study for testing the prototype and providing valuable insights. Lastly, I want to thank Kevin Chow from University of British Columbia for continuing this project together with people from HASEL. I'm excited to see where it will lead.

Abstract

Working long hours in front of a computer can negatively affect the health of knowledge workers. In this regard, emotions can play a significant role. Emotional awareness can help people to recognize their emotions, to make sense out of them, and to (pro-)actively regulate them. For example, by switching to another, more exciting task or taking a break when being in a negative emotional state. However, recognizing and being aware of one's own emotions can be challenging. We propose the EmotionalAwareness-Tool that tries to sense the emotional and cognitive state of a person sitting in front of a computer through a regular webcam. We aim to increase the users' awareness by visualizing the sensed data in real-time on a glanceable, always on top display. A user study revealed that even though the sensing accuracy of the EmotionalAwareness-Tool is very limited, it helped users to be more aware of their emotional and cognitive state. Participants of the study perceived the accuracy of the tool as reasonably good, possibly because the glanceable display was biasing them.

Zusammenfassung

Langes Arbeiten vor dem Computer kann die Gesundheit von Wissensarbeitenden beeinträchtigen. Emotionen spielen in dieser Hinsicht eine bedeutende Rolle. Emotionales Bewusstsein hilft Menschen, Emotionen zu erkennen, aus ihnen Schlüsse zu ziehen und diese (pro-)aktiv zu regulieren. Beispielsweise durch das Übergehen zu einer anderen, spannenderen Aufgabe oder durch eine Pause, wenn man in einem negativen emotionalen Zustand ist. Die eigenen Emotionen zu erkennen und sich derer bewusst zu werden kann jedoch schwierig sein. Wir präsentieren das EmotionalAwareness-Tool, das versucht, mit einer handelsüblichen Webcam den emotionalen und kognitiven Zustand einer Person, die vor dem Computer sitzt, zu erkennen. Indem wir die gemessenen Daten in Echtzeit in einem immer sichtbaren Fenster visualisieren, versuchen wir, das Bewusstsein der Nutzenden zu erhöhen. Eine Nutzerstudie hat gezeigt, dass auch wenn die Messgenauigkeit des EmotionalAwareness-Tools sehr limitiert ist, es den Nutzenden geholfen hat, sich ihres emotionalen und kognitiven Zustandes bewusst zu werden. Teilnehmende der Studie haben die Genauigkeit des Tools als zufriedenstellend empfunden. Möglicherweise, weil sie durch die Visualisierung beeinflusst wurden.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Knowledge workers spend long hours in front of the computer, working on tasks, planning, writing emails and meeting virtually. However, working the whole time in front of a computer can negatively affect health [DB09]. Emotions can play a significant role in this regard. Being in a positive emotional state promotes health and increases creativity and productivity [DTT19]. For example, positive emotions have a significant effect on the debugging performance of software developers [KBH10]. Contrary, negative emotions over time negatively affect overall job satisfaction and are therefore undesirable [Fis00]. Emotional awareness helps people to recognize their emotions, to reason about them, and to (pro-)actively regulate emotions when needed. For example, by taking a break or switching to another, more exciting task when being in a negative emotional state. However, it can be challenging to recognize and be aware of one's own emotions [BR03]. For this reason, a digital tool that makes knowledge workers more aware of their current emotional and cognitive state could be a helpful aid to increase self-awareness. With the recent advances in automated image analysis, new opportunities arise to detect these states using devices that are already in the workplace, such as webcams, and to provide real-time feedback to knowledge workers.

Various approaches exist that aim to sense the emotional and cognitive state of users and increase awareness. To this end, some approaches visualize the emotional and cognitive state in real-time, for example by using colored light [SMC+15]. Others provide retrospective timeline visualizations [MKK+12]. However, many of these approaches require specialized hardware. Further, many sensing approaches can be perceived as intrusive, since they rely on body-worn sensors and make use of interventions that can affect the user's concentration [CAJ+16]. In addition, some approaches are designed for a therapeutic or social setting, rather than for knowledge workers in the work environment. A more detailed overview of related work is provided in Chapter 2.

In this work, we propose a new webcam-based approach to sense the emotional and cognitive state of users. For this purpose, the user's valence, level of fatigue and task engagement are considered. Valence was chosen as it defines the spectrum between positive and negative emotions, and fatigue was selected as it can be closely related to emotions [GQEME15]. Engagement is considered to take also the user's cognitive state into account. In contrast to existing work, our approach tries to eliminate the need for specialized external hardware or sensors, and aims to be as unobtrusive as possible, but still provide real-time awareness on emotional and cognitive states. To achieve this goal, we formulated the following research questions:

- **RQ1: How accurately can we sense emotional and cognitive states, such as the level of valence, fatigue and engagement of a user, in real-time, by using images taken from a regular webcam?**

- **RQ2: How can knowledge workers' sensed emotional and cognitive state be visualized in a glanceable display, and what is the impact on awareness?**
  a. Can the sensed emotional and cognitive state be visualized in a glanceable display?
  b. Does the visualization accurately represent knowledge workers' emotional and cognitive state?
  c. How does the glanceable display influence them and what do they learn from it?
  d. What do users think about the value of a real-time visualization of their emotional and cognitive state in a glanceable display?

In this work, we present a tool, from now on referred to as *EmotionalAwareness-Tool*, which senses the emotional and cognitive state of a person sitting in front of a computer by using a regular webcam. We try to increase the awareness of users by visualizing the sensed data in real-time on a glanceable, always on top display. The tool consists of two parts: A *tracker*, where webcam images are processed, and a *client* that provides the glanceable display. This two-tier approach was chosen to make the tracker reusable for future projects. In addition, the presented tool runs completely locally to preserve the privacy of its users. Chapter 3 provides more information about the chosen approach.

To evaluate the sensing accuracy of the EmotionalAwareness-Tool and to study the applicability of the visualization approach, we performed a four-day user study. During the first three days, the glanceable display was disabled and the user was asked to perform self-reporting at regular intervals. On the last day of the study, only the glanceable display was visible. The study was concluded by an interview with participants to learn about their self-awareness and experience with the EmotionalAwareness-Tool. The study method is presented in Chapter 4.

By analyzing the data collected by the EmotionalAwareness-Tool as well as the feedback we got during the interviews, we found that participants of the user study consider an improved self-awareness useful and desirable. Even though the sensing accuracy of the EmotionalAwareness-Tool is very limited, it helped participants to be more aware of their emotional and cognitive state. The perceived accuracy of the tool was reasonably good, possibly because the glanceable display was biasing participants. Further, the study revealed that people express their emotional and cognitive state differently, which limits the applicability of a heuristic based sensing approach. The results of the user study are provided in Chapter 5 and in Chapter 6, potential limitations of the results are presented.

In Chapter 7, we discuss the results from the user study, explain how our sensing approach could be improved and discuss potential future research. Finally, Chapter 8 concludes the main findings of this work.

The main contributions of this work are the following:

- We introduce the EmotionalAwareness-Tool, an approach that senses valence, fatigue and engagement of a person sitting in front of a computer only using webcam images and visualizes this information in real-time.
- We implemented a reusable tracker that processes webcam images and provides data about raw as well as aggregated facial and posture related features.
- We provide the results of a four-day user study with four participants based on sensed data, a survey and follow-up interviews that show applications and limitations of the real-time sensing and visualization of emotional and cognitive data.

Chapter 2

Related Work

The approach of this work is related to previous studies that sensed a person's valence, fatigue or engagement by using digital tools. In the following sections, related work about each of the three dimensions is presented (Section 2.1 – Section 2.3). Further, related applications that make use of sensed emotional and cognitive data are considered (Section 2.4). Lastly, literature related to glanceable displays is presented (Section 2.5).

2.1 Valence

To describe emotional states, discrete categories like happy, angry and sad are generally used. According to Russell's circumplex model of affect [Rus80], each of these states can be characterized by a dimension of valence and arousal. Valence describes how positive or negative a state is according to the person. Arousal can be seen as the level of activation [VCl+14]. Russell's circumplex model is visualized in Figure 2.1. In this work, we focus on the dimension of valence, to consider the spectrum between positive and negative emotions. More information about the choice of dimensions is provided in Section 3.1.

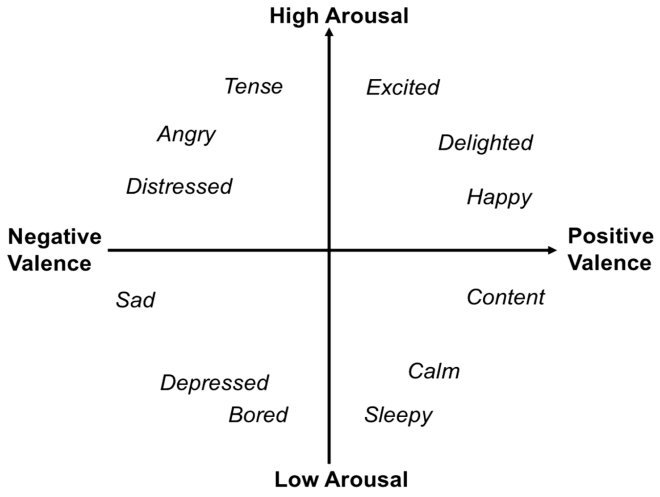

Figure 2.1: Russell's circumplex model of affect [Rus80] as used in Du et al. [DZP+20].

In order to estimate valence, McDuff et al. [MKK+12] found that head movements and facial activity were most useful among the various input streams used in their approach. Other researchers sensed valence using other methods: Valenza et al. [VCl+14] analyzed heartbeat measurements, other approaches use electroencephalographic (EEG) signals [LS13] or acoustic features in speech [GKMN07].

With advances in computer vision algorithms, the accuracy of systems that estimate emotional states by analyzing the human face increased. Over the last decade, several approaches for facial expression detection were implemented. Current approaches using convolutional neural networks have accuracies larger than 95% under ideal conditions [LdDOS17]. One way to detect expressions is to sense facial action units (AUs), which describe specific muscle activations. Expressions can then be seen as the combination of multiple AUs. Hernandez et al. [HMO+21] have shown that such an approach is applicable independent of the facial appearance of people.

2.2 Fatigue

The terms fatigue and sleepiness are used interchangeably in literature [SSF+21,GLNS21,PCB+07, QZL04]. Our motivation to sense fatigue/sleepiness is based on a study by Girardi et al. [GLNS21], in which the authors used the term *fatigue* (see Section 3.1). Therefore, we use this term in our work as well.

As for the detection of valence, electrophysiological measurements were studied to sense driver fatigue. Papadelis et al. [PCB+07] came to the conclusion that EEG and electrooculogram (EOG) data are promising indicators in fatigue detection. But also electrocardiographic (ECG) features, like the heart rate, were used successfully to detect fatigue by other researchers [ABD17].

Ji, Zhu & Lan [QZL04] proposed a reliable computer vision based system that monitors the fatigue level of drivers. In contrast to other approaches, they included multiple visual cues into their prediction. Eyelid and pupil movements, like the openness of eyes and the eye-closure speed, were considered. Further, head movements and facial expressions were studied, for example to detect yawning. Other eye-related parameters can be associated with fatigue as well. Schleicher et al. [SGBG08] focused on blinks and saccades as fatigue indicators. They found that along the eye-closure speed, the blink duration, delay of eyelid reopening and the time between blinks can be reliably associated with fatigue.

While all the above-mentioned studies are interested in measuring fatigue in driving scenarios, this work deals with fatigue estimation of people sitting in front of a computer. A similar scenario was analyzed by Divjak & Bischof [DB09]. They used a webcam for blink detection to prevent users from unhealthy behavior. While their main focus was the detection of fatigued eyes to reduce the Computer Vision Syndrome (CVS), they state that the approach could be applied to related use cases as well.

2.3 Engagement

The term *engagement* is not consistently defined in literature and is used to refer to several related concepts [DD18]. Schaufeli et al. [SSGR02] state that engagement can be seen as the opposite of burnout due to a negative correlation in a study performed by them. Further, they define the term as "a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption" [SSGR02]. Dobrian et al. [DSA+11] state that "qualitatively, engagement is a reflection of user involvement and interaction". In this work, we refer to the concept of engagement as defined by Schaufeli et al. [SSGR02].

Engagement can be sensed using various approaches. McDuff et al. [MKK+12] found that posture analysis was most useful to detect engagement in their work. Other researchers used EEG data to gain insights into the engagement level of a person [KM19]. Moreover, Doherty & Doherty [DD18] found out in a literature review, that behavior logging is one of the most frequent measuring methods in recent years. For example, Mark et al. [MICJ14a] analyzed the attentional states of users during online activities in the workplace by logging computer interactions and using experience sampling.

Babaei et al. [BSN+20] investigated how attentional states of knowledge workers are visible through facial cues. While working in front of a computer, participants were repeatedly asked to enter their perceived level of engagement. Meanwhile, a webcam captured images of the user's face. By analyzing the collected data, the researchers found a correlation between the viewing angle and the reported engagement, because disengaged users tended to focus less on the screen. Further, stretched or separated lips positively correlated with engagement in their study, but lips were frequently covered by the users' hands [BSN+20], which can be an indicator for different attentional states [MR11].

2.4 Existing Applications

Previous work not only focused on estimating emotional and cognitive states as precisely as possible but also proposed applications that add value for the user, for example by providing visualizations. In the following, existing applications are presented and compared to the approach of this work.

AffectAura by McDuff et al. [MKK+12] captures the emotional and cognitive state of users with a multitude of sensors. The dimensions valence, arousal and engagement are sensed by using a webcam for detecting facial actions, a Kinect sensor for posture estimation, microphones, a wrist-worn electrodermal activity (EDA) sensor and other metrics. By providing a retrospective timeline visualization, their approach serves as a memory aid and promotes self-reflection. They found out that AffectAura can help to discover patterns related to emotions and productivity and helps to reflect about them. Like AffectAura, the tool presented in this work provides visualizations to help users reflect about their emotional and cognitive state. However, to collect data, it relies only on webcam images. Further, it tries to improve self-awareness by providing real-time visualizations instead of serving as a memory aid and retrospection tool.

An approach that makes use of real-time feedback was chosen by Snyder et al. [SMC+15] in *MoodLight*. By using an EDA sensor, they were able to sense the arousal level of participants and visualized it with colored light. While the researchers found that ambient light helps to increase awareness in social contexts, the participants had to hold the EDA-sensor between thumb and forefingers during the entire lab study, making the approach less feasible for everyday applications.

Both above-mentioned studies as well as this work are based on the idea that users gain useful insights about themselves by analyzing collected data. Kersten-van Dijk et al. [KvDWBI16] critically examined this hypothesis by performing a literature review in the personal informatics domain. They concluded that almost all considered articles point out positive insights for the participants. It has to be noted though that this new knowledge about behavioral patterns did not necessarily lead to further actions by people.

An approach that tries to circumvent this limitation is *AttentivU* by Kosmyna & Maes [KM19]. AttentivU senses engagement with an EEG headband and provides subtle haptic feedback over a necklace to nudge participants when their engagement is low. Further, in a desktop application, engagement scores are visualized in real-time by a graph. Participants of a user study pointed out that the use of AttentivU increased their productivity and that they see the potential benefit

of the tool. However, in contrast to our work, the approach chosen by Kosmyna & Maes requires specialized hardware and the authors state that the form factor of the device could limit its social acceptability. Due to the ubiquity of webcams in the workplace, our approach does not face those issues.

Bialoskorski, Westerink & van den Broek [BWvdB09] also made use of colored light to visualize emotions. They created an interactive light installation called *Mood Swings* that consists of light bulbs, each equipped with an accelerometer. Users can interact with the installation by moving the bulbs according to their emotions. Based on the movements, an assessment regarding the valence and arousal of the user is made and visualized using different colors. Mood Swings was designed as an art installation, our approach on the other hand focuses on a working context.

To sum up, in contrast to existing work, our approach tries to eliminate the need for specialized external hardware by sensing the emotional and cognitive state of users only with a regular webcam. Further, unlike existing approaches, we aim to be as unobtrusive as possible, but still provide real-time awareness on emotional and cognitive states by making use of a glanceable display.

2.5 Glanceable Displays

Approaches that include screen-based visualizations of emotional or cognitive states usually make use of time-series graphs, which makes them unsuitable for real-time reflection. To overcome this limitation, Umair, Hamza Latif & Sas [ULS18] studied wrist-worn displays that continuously represent the arousal level of users. While some people liked that the glanceable display triggered awareness regarding their stress level, others felt distracted and had privacy concerns, since a wristband can be seen by everyone around. Our screen-based approach is suitable for real-time reflection but the provided visualization should not raise privacy concerns, because an application window is much less exposed than a wrist-worn display.

UbiFit Garden by Consolvo et al. [CMT+08] made use of a screen-based glanceable display to constantly depict reached physical activity goals as flowers in a virtual garden. Another study used a glanceable display to visualize transportation habits to promote sustainable mobility. Users noted that the approach increased their awareness and supported self-reflection [FDK+09]. Obermair et al. [ORM+08] used a digital picture frame on office desks to give users continuous feedback on their sitting position. The approach tried to promote healthier sitting habits while working in front of the computer. Users reported that the approach helped them to reflect about own habits, and that their productivity and concentration was not reduced by the glanceable display.

Chapter 3

Approach

To answer our research questions, we developed a privacy-preserving, locally run approach called *EmotionalAwareness-Tool* that senses the emotional and cognitive state of a person sitting in front of a computer through a regular webcam. In this regard, the user's valence, level of fatigue and engagement are sensed. We aim to increase the users' awareness of their cognitive and emotional state by visualizing the sensed data on a glanceable, always on top display using a Chernoff face [Che73]. By being more aware, users can sense negative states earlier and take countermeasures, such as taking a break or switching to another task. Furthermore, knowing when positive states occur can give valuable insights into the workday, like being more positive at a specific daytime [MKK+12]. In this chapter, we present why the dimensions valence, fatigue and engagement were chosen (Section 3.1), how they are sensed (Section 3.2) and how the tool visualizes the computed data (Section 3.3). In addition, we explain the architecture of our developed prototype (Section 3.4).

3.1 Chosen Dimensions

Valence. The overarching goal of our approach is to improve the users' self-awareness towards their emotional and cognitive state. Emotions affect ourselves in various ways. Positive emotions promote health, support teamwork and increase creativity and productivity [DTT19]. For example, research has shown a significant effect of emotions on the debugging performance of software developers [KBH10]. On the other hand, being in a negative emotional state over time can negatively affect overall job satisfaction [Fis00]. Therefore, to make users more aware of positive and negative emotions could be valuable. The spectrum between very positive and very negative emotions is defined as valence. This is why the dimension *valence* is considered in our approach. When studying emotions, usually arousal is measured alongside valence (see Chapter 2.1). As the spectrum between positive and negative emotions seems more important than the user's level of arousal for our purposes, we omit this dimension.

Fatigue. Fatigue is closely related to our emotions. In a study, software developers reported that fatigue is one of the most frequent triggers for negative emotions [GLNS21]. Further, mental fatigue can impair emotion regulation [GQEME15]. Girardi et al. [GLNS21] studied the influence of emotions on the perceived productivity of software developers. They noted, that fatigue might play a mediating role between productivity and emotions. In their study, valence was stronger correlated to perceived productivity in the afternoons, when people might be more tired. Because fatigue can be closely related to emotions, we include this dimension in our approach.

Engagement. Besides the emotional state, we are also interested in the cognitive state of users. The cognitive resources of knowledge workers are limited, therefore using them efficiently is important [DDB01]. Approaches that help knowledge workers to understand and manage their cognitive resources are thus desirable [BSN+20]. As related work studied the user's degree of engagement alongside valence [MKK+12, MICJ14b], we decided to sense engagement in our approach, to be able to set findings in context.

3.2 Sensing the Emotional and Cognitive State of Users

To make estimations regarding valence, fatigue and engagement, we created a basic heuristic for each dimension. By examining related literature, we selected facial and posture related features that might provide information on one of the three dimensions. Those input features were then combined using simple algorithms. To estimate valence, we consider the facial expression of the user. For fatigue, we measure blink frequency and consider sensed yawning. Engagement is estimated by head orientation, by detecting separated lips, the approximated distance of the user to the screen, and the presence or absence of wrists and elbows. Due to the simplicity of the heuristics, we expect that the accuracy of our approach is limited, but in a related study by Züger et al. [ZCM+17], even a simple algorithm with limited accuracy already provided value to the user. We are planning to refine the heuristics in a future iteration of the algorithms based on user feedback and collected data from a user study (Chapter 4).

In the following sections, we present how those facial and posture related features are extracted from webcam images and how they are combined to estimate valence, fatigue and engagement.

3.2.1 Sensing Valence

To sense valence, we use a pre-trained machine learning model that detects emotions based on facial expressions in image data. The model is publicly available as a JavaScript API called *face-api*[1] and is built on top of Tensorflow.js[2]. By relying on Tensorflow.js, machine learning models can be run entirely on the user's computer, which reduces privacy concerns. The API takes an image as input and returns probabilities for seven emotions (*angry*, *disgusted*, *fearful*, *happy*, *neutral*, *sad*, and *surprised*). We run the model every second and the emotion with the highest probability is stored. Because we focus on valence, we only consider *happy* and *sad* for further calculations. Those emotions are closest to the axis of valence in Russell's circumplex model of affect [Rus80] (see Figure 2.1).

By testing face-api, we recognized that most of the time, the API returns *neutral* as the most probable emotion during regular work. Further, *happy* is returned when the user is smiling, *sad* when the user is frowning. Since it seems unlikely that a user has to be smiling all the time to be in a state of high valance, we came up with the following basic algorithm visualized in Figure 3.1.

We decided to divide the dimension of valence into five states, depicted in the five nodes $v1$ to $v5$ in the figure. A high number represents high valence. To change between states, the facial expressions of the last 3 minutes are considered. The counter c_h represents how many previous minutes contained at least one *happy* emotion, c_s represents how many minutes contained at least one *sad* emotion. The counter values that lead to specific states can be inferred from Figure 3.1. Based on this algorithm, the current valence of the user is estimated every minute.

[1] `https://github.com/vladmandic/face-api`, verified 02.08.2021
[2] `https://github.com/tensorflow/tfjs`, verified 02.08.2021

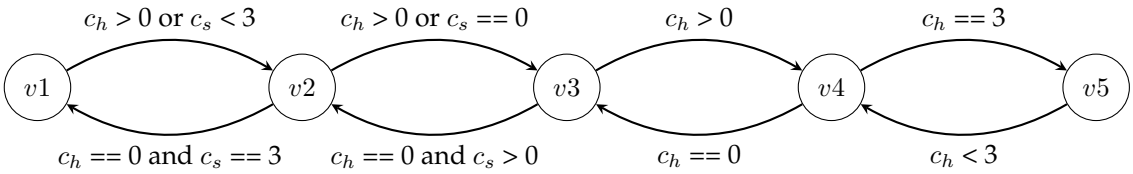

Figure 3.1: Algorithm to estimate valence.

3.2.2 Sensing Fatigue

As presented in Section 2.2, various eye-related features can help to sense fatigue [QZL04,SGBG08]. Due to the potentially limited frame rate and resolution of standard webcams, we decided to use the user's blink frequency to sense fatigue and did not consider features like eye closing speed or delay of eyelid reopening. According to Schleicher et al. [SGBG08], an increased blink frequency indicates fatigue. In order to detect the user's eye blinking, we included a Tensorflow model for facial landmark detection [SHZ+18] in our approach. The output of the model is a three-dimensional mesh of the human face, containing 468 vertices. Based on those vertices, we apply an algorithm proposed by Soukupová & Čech [SC16] to distinguish between open and closed eyes. According to them, an eye blink happens within 100 to 400 ms. To be sure to capture the closed eye, we run the model at 20 Hz. However, the frame rate of the webcam might impair the detection of eye blinks. Based on the sensed blinks, we store the added up number of blinks per minute.

Besides eye related features, other researchers used yawning detection for computer vision based fatigue estimation [QZL04, FYS07]. While testing our algorithms, we discovered that the face-api model reliably returns *surprised* when the user is yawning. This makes sense since the mouth is wide open when yawning. We assume that a surprised face with a mouth that is wide open rarely occurs during normal workdays. Therefore, we included yawning into our fatigue estimation by looking at the facial expression returned by the face-api model. As described in Section 3.2.1, the facial expression of the user is estimated every second.

Based on the blink frequency and the detected yawns, we created a basic algorithm to estimate the current level of fatigue of the user. The algorithm is visualized in Figure 3.2. For this first version of the algorithm, we decided to divide the dimension of fatigue into three states ($f1$ to $f3$), before extending it to more states if the approach works successfully. A high number represents a high level of fatigue. To include yawning in the algorithm, data of the last 3 minutes before the estimation is considered. The counter c_y is used to represent the number of minutes that contain at least one *surprised* emotion. While testing the sensing of eye blinks, we discovered that the blinks per minute over time are subject to high fluctuations. This observation is consistent with previous findings [SGBG08]. To reduce the effect of fluctuations in the number of blinks per minute, we consider the last 15 minutes before the estimation divided into three 5 minutes blocks. A time window of 5 minutes was also used by Schleicher et al. [SGBG08]. For each block, we calculate the mean blink count and compare the three values. As a higher blink frequency indicates fatigue, in our algorithm, a continuous increase in the means of the three 5 minutes time frames lead to a higher state of fatigue, a decrease to a lower state. How yawns and blinking are combined in the algorithm can be inferred from Figure 3.2.

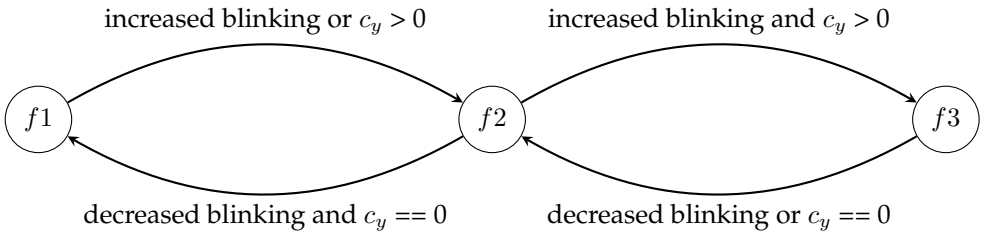

Figure 3.2: Algorithm to estimate fatigue.

3.2.3 Sensing Engagement

To sense engagement, we use facial features proposed by Babaei et al. [BSN+20]. They found that disengaged users tend to focus less on the computer screen. Therefore, head orientation could be a valuable parameter for detecting the current task engagement of users. Head orientation is provided as values for pitch, roll and yaw by the face-api model. Because the camera position might vary between users, those measures have to be related to a baseline, in which the user is looking directly at the screen. In our approach, the baseline is calculated during the first 3 minutes of sensing. We assume that in this early stage of the workday, the user is still sitting upright and is looking directly at the screen. The deviation from the baseline is then calculated by summing up the differences between sensed pitch, roll and yaw and their corresponding baselines. The head orientation is evaluated every 2.9 seconds. We chose odd sampling rates to smooth CPU activity.

In the study by Babaei et al. [BSN+20], separated lips correlated with engagement. To extract this feature from webcam images, we use the model presented in Section 3.2.2 that estimates facial landmarks. From the vertices returned by the model, we chose the ones that represent the edge of the mouth. As used for the detection of open and closed eyes, we took the algorithm by Soukupová & Čech [SC16] and modified it slightly to detect the level of openness of the user's mouth. The algorithm is executed alongside the calculations for blink detection at 20 Hz, because the polygon mesh is generated anyway. However, only the mean mouth openness over 5 seconds is stored into the database to reduce memory consumption. To reduce the influence of anatomical differences, we subtract a baseline value calculated during the first 3 minutes of sensing from the stored values.

McDuff et al. [MKK+12] used depth information captured by a Kinect sensor to detect the user's direction of lean, which can be valuable to detect engagement. Regular webcams don't provide this type of information. To approximate the distance of the user's face to the screen, we use the area of the face related to a baseline computed in upright posture. To calculate this area, we use again the model for facial landmark detection [SHZ+18] and extract the vertices corresponding to the outline of the face. Then, we apply Gauss's area formula to get an approximation for the facial area. As for mouth openness, calculations are made at 20 Hz and the mean over 5 seconds is stored. This value is then compared to a baseline, created during the first 3 minutes of sensing.

Further, if the user leans back and might stretch her or his arms, elbows might appear in the camera frame. To detect the presence of elbows, we use a Tensorflow model for pose detection [PZC+18]. The model takes an image as input and returns the coordinates of 15 pose related keypoints and their confidence score. If a score above 0.5 is returned for the left or right elbow, we assume that an elbow is present in the image. The pose model is run every 5 seconds. We have set this sampling rate lower than the previously mentioned ones because we assume that the user's pose changes less frequently than her or his facial features.

People sometimes hold their hands close to their face as non-verbal behavior. This can be

a sign for interest and thinking, but depending on the way the hand is placed, also for boredom [MR11]. Even though a hand-over-face gesture is not an unambiguous sign of engagement, we included it in our estimation. According to our experience, as well as according to a small study performed by Mahmoud & Robinson [MR11], engagement related hand-over-face gestures occur more frequently than boredom related ones. Of course, this hypothesis has to be critically questioned in the evaluation of the algorithm. As for the detection of elbows, the model for pose detection [PZC+18] is used to detect hands held close to the face. The model returns the position of wrists. If the confidence score of one wrist is higher than 0.5, we assume a hand is present in the camera frame. The chosen sampling rate is 5 seconds.

In order to estimate the current engagement of the user, we designed a basic algorithm. To calculate the estimate, a $score$ is introduced as an intermediate step. The score is initially equal to zero. If the head orientation deviates notably from the baseline value, the score is reduced, as this can be an indicator of disengagement [BSN+20]. Separated lips increase the score because, according to Babaei et al. [BSN+20], this can be a sign of engagement. Leaning towards the screen increases the score, leaning away reduces it. The presence of elbows reduces the score and recently detected wrists increase the score. If the resulting value is positive, the algorithm returns high engagement. If the score is negative, low engagement is returned. How the algorithm calculates the score and how the final estimation is made can be seen in Figure 3.3.

```
score = 0;
if head orientation deviates then
    score -= 1.0
if lips separated then
    score += 0.5
if head close to screen then
    score += 1.5
else if head far from screen then
    score -= 1.5
if elbow recently present then
    score -= 1.0
if wrist recently present then
    score += 1.5

if score > 0 then
    newEngagement = 3
else if score == 0 then
    newEngagement = 2
else
    newEngagement = 1

return newEngagement
```

Figure 3.3: Pseudocode algorithm to estimate engagement.

3.3 Visualization With Chernoff Face

To make users more aware of their current valence, fatigue and engagement, we depict the three dimensions in real-time using a Chernoff Face [Che73]. A Chernoff face is a visualization approach that represents k-dimensional data by a face, whose characteristics depend on the data

point. In our case, we vary different parts of the face depending on the three-dimensional input data. To a certain extent, this approach resembles the visualizations used by Bradley & Lang [BL94] in the *Self-Assessment Manikin (SAM)*, which is frequently used in the literature [GWA15, GLNS21]. Their approach uses manikins to depict different levels of valence, arousal and dominance to facilitate self-reporting of emotions.

Valence is depicted in our approach as well as in the Self-Assessment Manikin by varying the shape of the mouth. If valence is high, the face is smiling, if valence is low, it is frowning. To depict the level of fatigue, the degree of eye closeness is used. Almost closed eyes represent high fatigue, wide open eyes low fatigue. The visualization of engagement is inspired by a finding of Babaei et al. [BSN+20]. In a study conducted by them, disengaged users were focusing less on the computer screen than engaged users. This is why the Chernoff face in our approach is looking straight when engagement is high and to the side if it is low. Two examples of combinations of the three dimensions can be seen in Figure 3.4 and Figure 3.5. By visualizing valence, fatigue and engagement in a single representation, it is possible to get an impression of the three-dimensional data at a single glance. Further, a human-like face visualization seems appropriate for depicting emotional and cognitive data, as we share this information through facial cues [MKK+12]. In addition, a human-like visualization of emotions was already used successfully in the SAM approach.

Figure 3.4: Chernoff face for high valence, neutral fatigue and high engagement.

Figure 3.5: Chernoff face for low valence, high fatigue and low engagement.

Due to the subtle nature of the facial features of the Chernoff face, extreme states could be overlooked. Therefore, extreme states are highlighted by a different background color. If the lowest possible state of valence or engagement is estimated, the background turns red. The same holds if the estimated fatigue is in its highest state. If valence or engagement are at their highest possible level, or fatigue is at its lowest possible level, the background turns green. The algorithm that determines the background color can be seen in Figure 3.6.

```
backgroundColor = black;
if valence == 1 || fatigue == 3 || engagement == 1 then
  | backgroundColor = red
else if valence == 5 || fatigue == 1 || engagement == 3 then
  | backgroundColor = green
```

Figure 3.6: Pseudocode algorithm to determine background color of Chernoff face.

In the prototype we developed, the Chernoff face is included in a glanceable, always on top window and is refreshed every minute based on the current estimates. This approach makes sure that the user can quickly look at her or his estimated emotional and cognitive state without getting distracted by the EmotionalAwareness-Tool. The glanceable window of the prototype is presented in more detail in Section 3.4.3.

3.4 Prototype

In this section, implementation details of the EmotionalAwareness-Tool prototype are presented. Its overall structure and the most important components are described.

3.4.1 Architecture Overview

The prototype of the EmotionalAwareness-Tool is implemented in TypeScript using the Electron[3] framework, which allows the development of cross-platform desktop applications with a single codebase. The prototype is split into two parts: A *tracker*, where webcam images are processed and the estimates regarding valence, fatigue and engagement are calculated, and a *client* that stores the data received from the tracker and visualizes the estimates. By relying on a two-tier approach, we make the tracker reusable for future projects. The tracker communicates with the client over inter-process communication. For persistency, a SQLite database is used. An overview of the main data flow in the application can be seen in Figure 3.7. The different components are explained in more detail in the following sections.

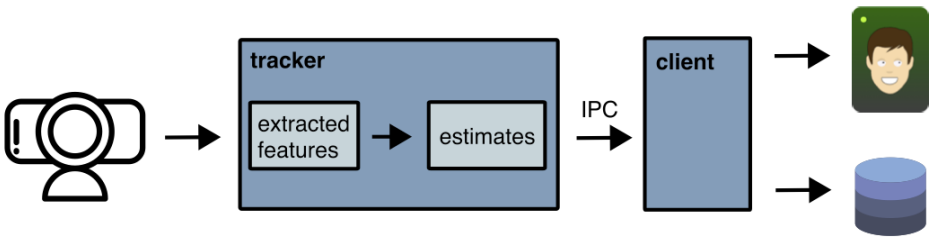

Figure 3.7: Main data flow in the EmotionalAwareness-Tool.

3.4.2 Tracker

The tracker consists of two sub-components. A so called *models* part, where the pre-trained convolutional neural networks (CNN) are run and a *metrics* part, where the estimations regarding valence, fatigue and engagement are calculated. A component diagram of the tracker is provided in Appendix A. The *models* part accesses the webcam stream through the JavaScript MediaDevices interface. The stream is then given as input to the three Tensorflow.js based machine learning models: face-api[4] for facial expression and head orientation detection, face-landmarks-

[3] `https://www.electronjs.org`, verified 03.08.2021
[4] `https://github.com/vladmandic/face-api`, verified 02.08.2021

detection[5] [SHZ+18] to get the coordinates of facial landmarks, and posenet[6] [PZC+18] for pose estimation. These machine learning models run locally on the user's computer. No data is sent to a server, which could raise privacy concerns. Based on the raw data returned by the models, additional features, like the area of the user's head, are calculated (consult Section 3.2.1 to Section 3.2.3 for more details). The raw data returned by the models as well as the additionally calculated features are then sent over inter-process communication to the client.

Some features are given to the *metrics* part and stored there in buffers. From those buffers, data of the first 3 minutes of sensing is used to calculate baseline values. From then on, based on the data stored in the buffers as well as the baseline values, estimations regarding valence, fatigue and engagement are made and sent to the client at an interval of 60 seconds.

The tracker is implemented as an Electron BrowserWindow, consisting mainly out of TypeScript, HTML and CSS files. The code is published as a npm-package[7] to be able to install it in the client.

3.4.3 Client

The client provides the user interface of the EmotionalAwareness-Tool. It receives data from the tracker over inter-process communication, stores the data in a SQLite database and visualizes the estimated dimensions in a glanceable window. Further, the client provides options to start and pause the tracker and shows the tracker's current state in the task bar. In the following, the glanceable window is presented in more detail.

Glanceable Window

As described in Section 3.3, the glanceable window contains the Chernoff face that visualizes the sensed valence, fatigue and engagement of the user. Further, the background color changes depending on the current estimates. The window has a fixed, small size (150×100 pixels) and can be positioned freely. However, to make sure the user is able to see it all the time, the window is not closeable and is always in the foreground. In addition, the glanceable window includes a virtual status LED that indicates the current state of the EmotionalAwareness-Tool. The LED is green when the application is running, orange when it is starting or no user is detected, and red when the application is paused. In Figure 3.8, the glanceable window is shown on top of a browser window.

To make sure the Chernoff face represents the emotional and cognitive state of the user as good as possible, some basic customization options are available. The user can set the skin color, hair length and hair color of the visualization in the application settings. Some possible combinations are depicted in Figure 3.9.

In the following, the drawing process of the Chernoff face is described. For each skin color, the outline of the face is stored as an SVG file. This outline is drawn to screen by using the JavaScript Canvas 2D API. Like the outlines, different hairstyle and hair color combinations are stored as SVG files. Depending on the user's settings, one combination is drawn on top of the outline using the same API. Lastly, the facial features are added. Depending on the estimated valence, fatigue and engagement, canvas points and paths are drawn onto the face.

[5] `https://github.com/tensorflow/tfjs-models/tree/master/face-landmarks-detection`, verified 03.08.2021

[6] `https://github.com/tensorflow/tfjs-models/tree/master/posenet`, verified 03.08.2021

[7] `https://www.npmjs.com/package/pa-egloff`, verified 03.08.2021

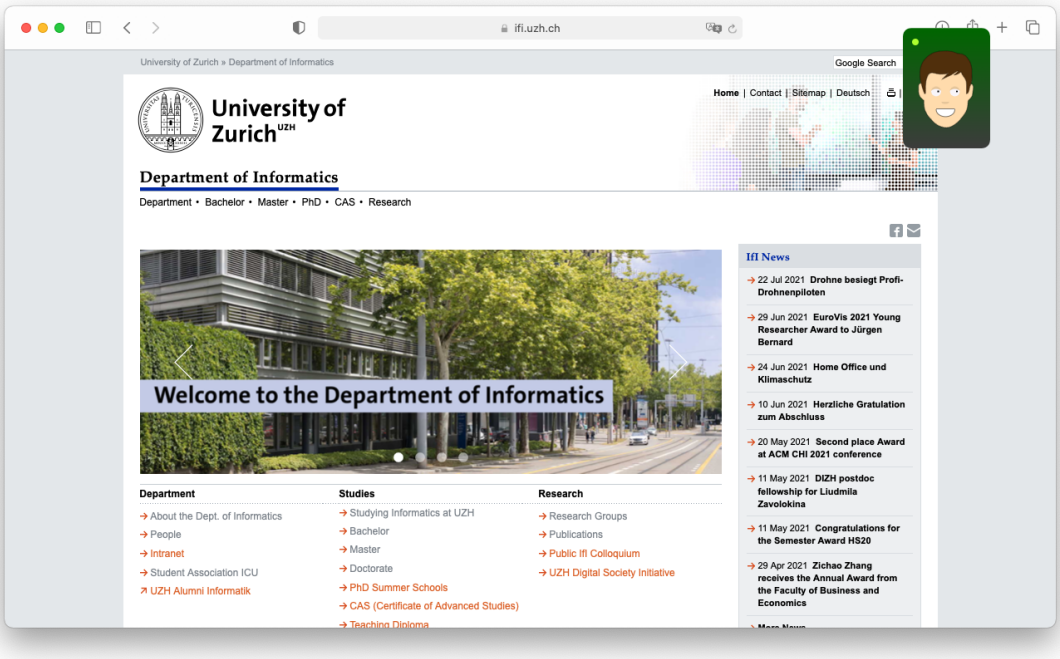

Figure 3.8: Glanceable window sitting on top of another window.

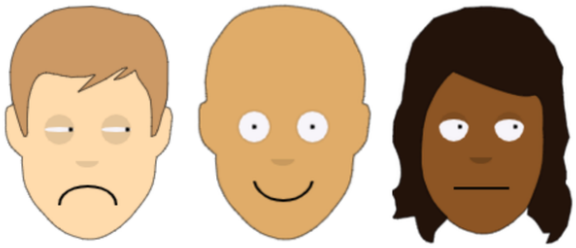

Figure 3.9: Customized Chernoff faces.

Chapter 4

Study Method

To evaluate the accuracy of the estimations regarding valence, fatigue and engagement and to get feedback about the chosen visualization approach, we performed a user study. The study was conducted with four participants affiliated with the University of Zurich and lasted for at least four days. During this period, participants were working as usual while the EmotionalAwareness-Tool was running in the background. In this chapter, details about the study procedures, participants, collected data, and analysis are provided.

4.1 Procedures

We performed an in-situ user study with four participants including data collection, a survey, and a follow-up interview. The user study contained two phases, an experience sampling phase, and an intervention phase. During the experience sampling phase, the glanceable display was disabled and self-reports were collected to evaluate the accuracy of the estimates regarding valence, fatigue and engagement. In the intervention phase, the glanceable display was visible to get feedback about the visualization approach. The study was preceded by a pilot study.

Pilot Study. To get early feedback on the usability of the EmotionalAwareness-Tool and to get first insights into the sensed data, a pilot study was performed. Four people (all male) involved in the project used the tool for several days. The pilot proved that the data collection was working reliably and revealed some possible usability improvements, like better visual cues to indicate the application's state. The feedback regarding usability was then included into the application before the next steps of the study were performed.

Recruitment. Information about the study was sent out by email to knowledge workers from the personal and professional network of the researchers. If people were interested in participating, a consent form was sent to them and a date for a kick-off meeting was scheduled.

Kick-Off Meeting. The primary user study started with an individual kick-off meeting. There, the goals and procedure of the study were explained to participants, and they had the opportunity to ask remaining questions. In case there were no more questions and the participant had sent the signed consent form to one of the researchers, participants were asked to install the EmotionalAwareness-Tool on their work computer. One of the investigators explained the main functionalities of the tool and made sure that it was working correctly.

Experience Sampling Phase. Subsequently, for 3 days, the participants were asked to work as usual while the EmotionalAwareness-Tool was running in the background. As the participants should work as usual, there were no requirements regarding the positioning of the webcam. In this phase of the study, the glanceable display was disabled to avoid biasing the participants. Once every 30 minutes or when extreme states were sensed (e.g. very happy or very engaged), a pop-up appeared on participants' computers. There, participants were asked to self-report their perceived levels of valence, fatigue and engagement. The self-reporting pop-up is shown in Figure 4.1. In a subsequent page of the pop-up (Figure 4.2), which was only visible every third time, the submitted self-report was visualized in a Chernoff face to study how participants interpret the face and how well they think the face mirrors their emotions. Participants could postpone the self-report at any point.

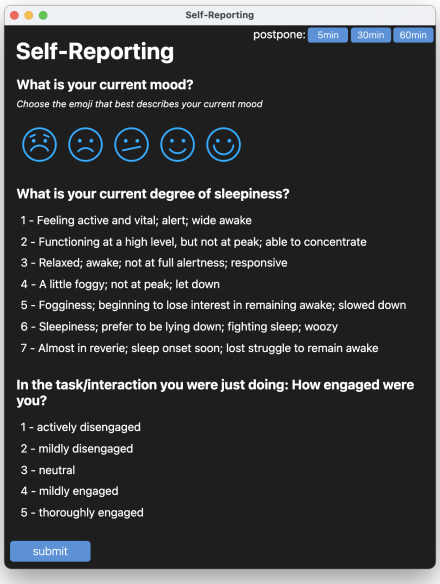

Figure 4.1: Self-reporting pop-up.

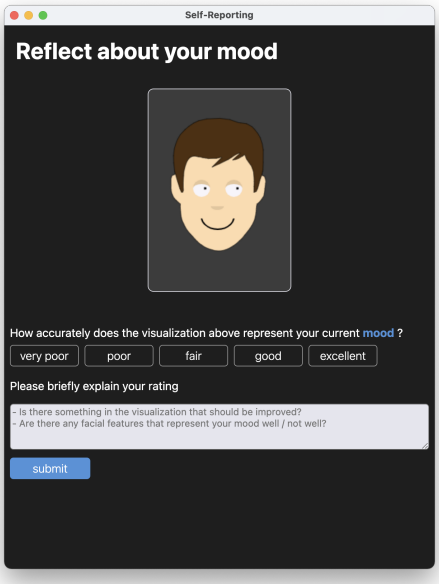

Figure 4.2: Visualization feedback page.

Intervention Phase. In the second phase of the study, the self-reporting pop-up was disabled and the glanceable display was visible. With the glanceable display turned on, the participant saw her or his sensed emotional and cognitive state in real-time. This phase lasted between 1 and 2 days, depending on the time availability of the participant.

Survey. On day 4 of the study, alongside a reminder to switch to the next study phase, a survey was sent to the participant. The survey was used to collect demographic data and to ask some closed-ended questions about the participant's desktop setup and work. The questions of the survey are listed in Appendix B.

Final Interview. We concluded the study with a final interview. Participants were asked about the interpretability of the visualization, their experience with the glanceable display and their

learnings. In addition, we asked them about their thoughts about the heuristics used for sensing valence, fatigue and engagement. In the final interview, participants were further asked to export and upload the data that the EmotionalAwareness-Tool stored on their computer.

4.2 Participants

The participants of the primary study (4 participants, 3 male, 1 female) were recruited using the researchers' personal and professional network. Three participants used the EmotionalAwareness-Tool for 4 days, one participant for 5 days. Further, all participants submitted their locally stored data for further analysis. All participants are working in the field of computer science (computer science students or graduates).

The age of participants of the primary study was between 23 and 30. The mean age was 25.75 years, the standard deviation 2.95 years. Regarding the job title, two participants answered that they are students. One of them additionally mentioned being a part-time developer. Two participants stated that they are working as research assistants. All mentioned that they are spending most of their working time in front of the computer. The reported minimum was 75%, the maximum 100%, with a mean of 92.5% and a standard deviation of 10.31%. Three participants reported that they were using external webcams for the study and used multiple computer screens. One participant used a laptop webcam and no external screens.

4.3 Data Collection

The following types of data are collected by the EmotionalAwareness-Tool.

Webcam Data. The EmotionalAwareness-Tool processes raw webcam images using convolutional neural networks (CNN). The used machine learning models work locally and compute estimates of landmark coordinates as well as the estimated facial expression. Based on landmark coordinates, aggregations are made. These include facial expressions over time, number of blinks, head orientation, head area, mouth openness, proximity to screen and the presence or absence of wrists and elbows. The sampling rate differs depending on the model (see Section 3.2). While the raw image data is discarded immediately after the inference step, the estimates are stored locally in a database.

Experience Sampling. To measure the accuracy of the EmotionalAwareness-Tool and to answer RQ1, experience sampling is used. The participant's responses to the self-reports are stored locally in the database. The self-report consists of the following three questions:

- What is your current mood?
- What is your current degree of sleepiness?
- In the task/interaction you were just doing: How engaged were you?

The first question can be answered by clicking on one of five provided emojis that depict different levels of valence. The word *current* was included in the question, to make sure the participant states her or his mood in the moment, to be able to link the answer later to data captured by the webcam. For the second question, the wording and the chosen labels of the provided 7-Point Likert scale are based on the *Stanford Sleepiness Scale* [HDZ72] as used in Shahid et al. [SWMS12]. We use the term *sleepiness* in the self-report instead of *fatigue*, as we could not find

an established scale that uses the term *fatigue* in the literature. The third question is taken from a related study by Mark et al. [MICJ14a]. The labels of the 5-Point Likert scale that is provided to answer the question are as used in Moore et al. [MSM+16].

Visualization Feedback. Every third time the user answers the self-reporting questions, a second page appears to gather feedback about the chosen visualization approach. To get more concise feedback about the three depicted dimensions, the application iterates over the following three questions:

- Does the visualization above represent your current mood?
- Does the visualization above represent your current degree of sleepiness?
- Does the visualization above represent your current engagement?

The user can answer the question by using a 5-Point Likert scale. Further, a text field is provided where the participant has to explain her or his rating in more detail. The collected feedback about the visualization is used to answer RQ2a and RQ2b.

System Load. To assess the processing power consumption of the EmotionalAwareness-Tool and its memory usage, the overall CPU usage and the free memory of the participant's system are measured every 5 minutes and stored into the database. For this purpose, the npm-package *os-utils*[1] is used.

4.4 Data Analysis

Interviews. The follow-up interviews were held online in English using *Zoom*[2]. All participants agreed that the interview was recorded and that an automated transcription service was used. To transcribe the interviews, the tool *Descript*[3] was used.

To analyze the interviews, we performed a coding approach. All answers given by the participants were tagged with one to four codes. By using those codes, it was later possible to make connections between different answers from different participants and to extract the main themes. Based on those themes, we then tried to answer our research questions. To refer to answers from one of the four participants, we use the abbreviations P1 to P4 in the analysis.

Pre-Processing of Sensed Data. The data sensed by the EmotionalAwareness-Tool was analyzed using statistical analysis by using Python and Jupiter Notebooks. To link data captured by the webcam to self-reports submitted by the participants, a time window has to be introduced. However, there are usually no guidelines on which time window has to be used for with feature [ZMMF18]. For the analysis, if not stated otherwise, we consider data collected during the last 3 minutes before a self-report was submitted. The pop-up itself could have an influence on the user's facial expression and posture. As the time window ends with the submission of the self-report, with a shorter time-window, most data captured during the specified time would be impaired by this influence. To consider a longer time frame introduces the problem that it is possible that a user filled out two self-reports shortly after each other, and thereby certain data points would be considered multiple times.

[1] `https://www.npmjs.com/package/os-utils`, verified 17.08.2021
[2] `https://zoom.us`, verified 04.08.2021
[3] `https://www.descript.com`, verified 04.08.2021

Some self-reports were submitted before the EmotionalAwareness-Tool started storing data or in the first minute of sensing. This makes those self-reports less useful for statistical analysis, as they cannot be linked to webcam data. Therefore, those self-reports were excluded from the analysis.

Chapter 5

Results

The results of the user study are presented in this chapter. Section 5.1 gives an overview of the participants' self-reports. Differences in how participants express their emotional and cognitive state could have an impact on the results of the study. Therefore, insights about interpersonal differences are provided in Section 5.2. In Section 5.3, we present a statistical analysis regarding the accuracy of the implemented algorithms that try to sense valence, fatigue and engagement. Further, ideas are presented about how the accuracy of the algorithms could be improved based on insights from the user study (Section 5.4). In Section 5.5, we present feedback about the visualization approach. Lastly, the key findings of the study are summarized in Section 5.6.

5.1 Self-Reporting

In total, the four participants answered 158 self-reports. The number of self-reports per person is between 30 and 49. As mentioned in Section 4.4, some self-reports are not useful for the analysis, as they cannot be linked to sensed data. 143 self-reports have at least one minute of data collection before them and are thus considered in the analysis.

The range of self-reported values differ between participants. As shown in the bar charts in Figure 5.1, not all participants made use of the whole spectrum of possible values during the study period. Over all participants, a valence of 4 (on a scale of 1 to 5) was reported most frequently (72 times). Low valence (state 1 or 2) was only reported 10 times. Regarding fatigue, the most reported state was 2 (on a scale of 1 to 7, 45 times). High fatigue was rarely reported: Four self-reports that include a fatigue of 6 exist, and no participant reported a fatigue of 7. The self-reported engagement of participants is generally high. Sixty-eight self-reports contained the highest possible level of engagement (state 5), while only 10 self-reports included an engagement level of 1 or 2.

The reported valence, fatigue and engagement of the participants change over a day. As an example, one arbitrarily chosen day of P2 and P3 is depicted in Figure 5.2. On the one hand, general trends are visible, like a decrease in valence and engagement over time alongside an increase in fatigue for P2. Or for P3, peaks of high valence and engagement in the morning and evening. Those continuous trends support the aspect of our implemented valence and fatigue algorithms that a new estimated state relies on the previously sensed state. On the other hand, also short-term fluctuations with lower magnitude are present. Even though only the self-reports of P2 and P3 are visualized in the figures, similar patterns occur for all participants.

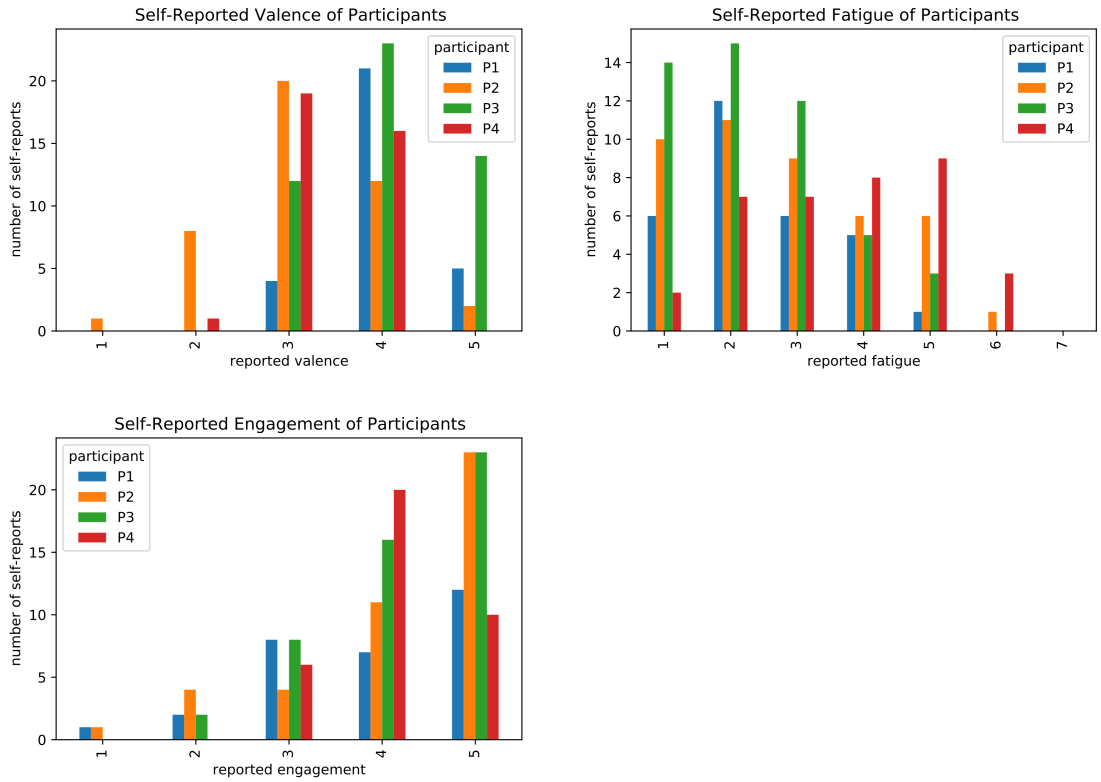

Figure 5.1: Overview of self-reported values of valence, fatigue and engagement of the four participants.

5.2 Interpersonal Differences

To answer RQ1, it is useful to get an overview of how differently users physically express their emotional and cognitive state. In the current heuristic-based algorithms, the estimations regarding valence, fatigue and engagement are based on the same assumptions for all users. This makes our approach potentially vulnerable to interpersonal differences. Therefore, we present examples of variations between users in this chapter. Data stored by the EmotionalAwareness-Tool as well as answers from the survey and the follow-up interviews are included.

Differences in Expressiveness. People differ in how their emotional and cognitive state is reflected in their facial expression. At least, this is how people self-assessed their expressiveness during the interviews. P4 stated that *"usually even if I'm in like kind of a positive mood, typically I'm not smiling"*. P3 answered *"usually I am quite an emotional person maybe, and then you can really see the emotions in a face [...] even when I'm on my own"*. However, three out of four participants noted that self-assessing their expressiveness is a hard task. Differences in the facial expressions of users are also present in the sensed data. As an example, we compare the facial expressions of different participants before a self-reported valence of 4 was submitted. On average, the face-api model returned 1.4 sad expressions per minute for P1, while it returned 25.9 for P4. Sensed happy expressions are rare for all participants. As expected, users are not necessarily smiling when their valence is high. However, differences are still present. For P1, the model returned on average 1.9 happy expressions per minute, while it returned only 0.3 for P3. The mean number of occurrences

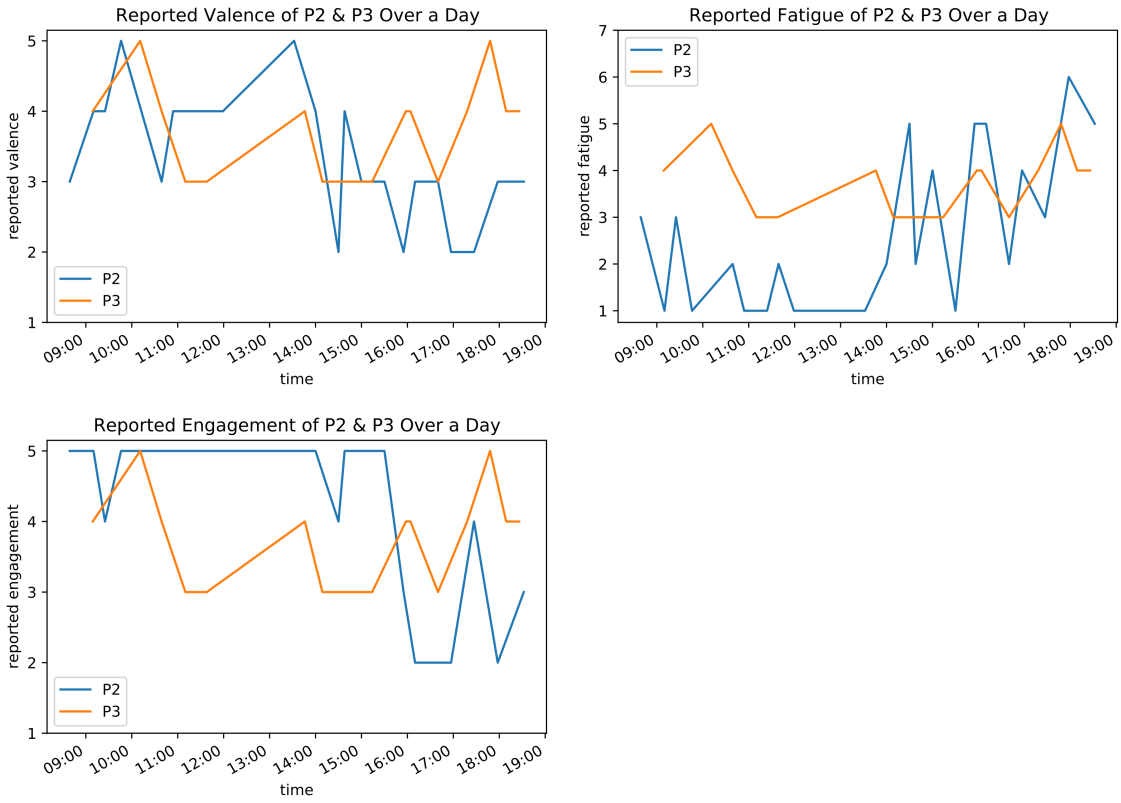

Figure 5.2: Reported valence, fatigue and engagement of P2 and P3 over a day.

of all returned expressions are shown in Table 5.1. More insights about how the facial expressions of participants changed depending on the reported level of valence are provided in Section 5.4.2.

User ID	sad	neutral	happy	angry	disgusted	fearful	surprised
P1	1.4	36.3	1.9	0.5	2.4	0.7	8.3
P2	1.2	50.5	1.3	1.9	0.0	0.6	0.5
P3	11.9	43.5	0.3	0.6	0.1	0.3	0.5
P4	25.9	30.5	1.5	0.0	0.1	0.1	0.1

Table 5.1: Mean number of sensed expressions per minute for a self-reported valence of 4.

Different Behaviors. People behave differently when being engaged or fatigued. In the follow-up interviews, participants were asked if they have certain behaviors they do when being engaged. P2 mentioned *"maybe if I'm very engaged, like in programming or something, listening to music, so you might move your head and this might be like an indicator"*. P3 reported that she or he is looking to the ceiling when being thinking and engaged. Further, the person mentioned *"when I'm engaged... I also tend to move my hands up in front of my face"*. However, participants noted that their behavior can be task-dependent. P1 reported *"So I like [...] sit back in my chair and still be engaged while other times, I try to sit up and, and yeah, come closer to the screen or whatever, while not being*

engaged". Regarding holding a hand in front of the face when being engaged, P2 noted that *"it might also depend on the task. Like if you reading, I might be like this and then just scrolling and whatever. And if I'm programming, I'm more like I need the hands"*.

As discussed in Section 3.2.2, we make use of the blink frequency to sense fatigue. Based on literature, large inter-individual differences in the mean blink count per minute are expected [SGBG08]. This effect can also be seen in the collected data. While the mean blink count of P4 was 25.6 (standard deviation 7.5), the mean of P3 was 15.5 (standard deviation 6.7). High standard deviations can also be expected based on the literature [SGBG08]. The blink counts per minute of P1 and P2 were below 10. Those blink frequencies were potentially affected by sensing inaccuracies (see next paragraph). Inter-individual differences with reasonable blink frequencies were also present in the data from the pilot study. Possible improvements regarding the eye blink detection are discussed in Section 5.4.3.

More information about differences in the sensed facial and posture related features are provided when presenting their correlations to valence, fatigue and engagement in Section 5.4.

Differences in Data Collection. Not only differences in the facial expressions of users or different behaviors can have an impact on the sensing approach, but also how their workplace is set up. P1 used two computer screens side by side, with the camera standing on the desk in the middle of the two monitors. P2 and P3 used a laptop on the desk and a second monitor above with an external webcam on it. P4 used a laptop and the built-in camera. As we did not specify a required camera position, the users are recorded from different angles. Further, camera quality and lighting differed between participants and some participants were wearing glasses. Those aspects could lead to differences in the accuracy of the sensed data.

5.3 Accuracy of the Current Algorithms

In this section, we compare the estimates regarding valence, fatigue and engagement returned by the implemented prototype algorithms with the self-reports submitted by the participants. Further, fluctuations in the estimated values are considered and we present, how the participants perceived the sensing accuracy during the second phase of the study. By doing so, we try to answer RQ1.

5.3.1 Overall Accuracy

To evaluate the overall accuracy of the algorithms, difference-measures are used. In the following, the mean estimated valence, fatigue and engagement of the 3 minutes before a self-report are compared to the corresponding value reported by the user. To do so, the absolute difference of the two values is calculated. Then, the distribution of the differences is analyzed. It has to be noted though that the scales the participants used to report fatigue and engagement do not match the number of states the corresponding algorithms can return. The scale for reporting fatigue goes from 1 to 7, the scale for engagement from 1 to 5. Both corresponding algorithms can return only 3 different states. Therefore, a rescaling is needed. The rescaling was performed according to the scheme shown in Table 5.2. Valence is reported on a scale of 1 to 5 and also the corresponding algorithm can return 5 different states.

Valence (1-5): The mean absolute difference between estimated and self-reported valence considering 143 self-reports is 0.99 scale items with a standard deviation of 0.85. In the pilot study, the accuracy of the valence estimation was better. There, the mean difference was 0.61 with a standard deviation of 0.62 considering 148 self-reports. This difference in the results could hint

Fatigue		Engagement	
Reported	Rescaled	Reported	Rescaled
1	1	1	1
2	1	2	1
3	2	3	2
4	2	4	3
5	2	5	3
6	3		
7	3		

Table 5.2: Rescaling applied to reported fatigue and engagement to match range of returned values by algorithms. Reported valence does not require a rescaling.

towards a limited ability of the algorithms in coping with interpersonal differences, as discussed in Section 5.2.

Fatigue (1-3): For fatigue, the mean difference is 0.74 scale items, the standard deviation is 0.61. This result is comparable to the results from the pilot (mean 0.74, standard deviation 0.57).

Engagement (1-3): The mean difference between estimated and self-reported engagement is 0.84 scale items with a standard deviation of 0.76. This result is also similar to the difference-measures in the pilot study (mean 0.80, standard deviation 0.65).

When analyzing those results, it has to be noted that the range of possible values is from 1 to 5 for valence and from 1 to 3 for fatigue and engagement. A boxplot of the difference-measures of the primary study is shown in Figure 5.3.

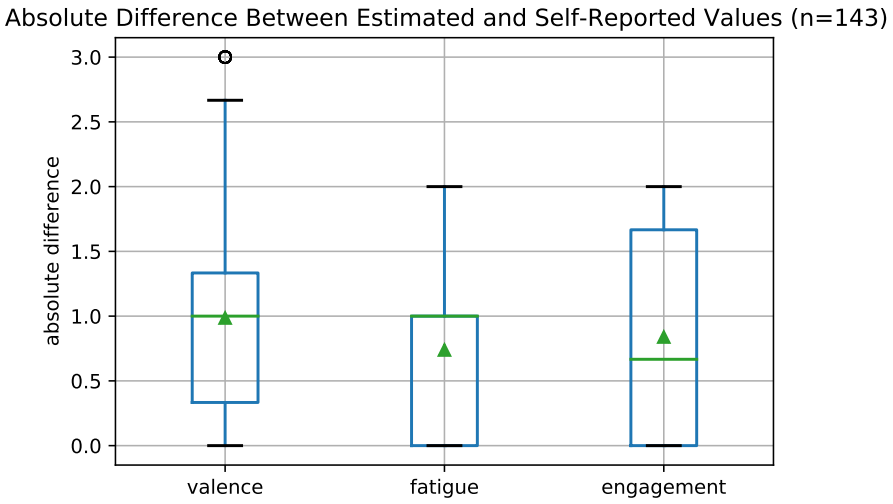

Figure 5.3: Absolute difference between estimated and self-reported values. Note: Valence is measured on a scale of 1 to 5, fatigue and engagement on a scale of 1 to 3.

To evaluate the accuracy of the algorithms given specific perceived emotional and cognitive states of users, the confusion matrices shown in Figure 5.4 are analyzed. As we compare the mean of estimated values over three minutes with the self-reported values, decimal numbers exist. To be able to compute the confusion matrices, the mean of the estimates is rounded to the next integer.

The confusion matrices reveal that the accuracy of the algorithms as they are currently implemented is very limited. In terms of valence, for a self-reported valance of 4, the most frequent estimated value is 4, which is correct. However, for this level of valence, the algorithm misjudged the user's valence more often than it was correct (36 incorrect estimates, 29 correct estimates). Further, when the user's reported valence was 3, the algorithm returned 4 once more than it returned 3. In addition, when the perceived valence of the user was in its highest state (5), the algorithm returned a valence of 2 most of the time. For fatigue, the algorithm estimated a level of 2 most of the time, independent of the reported fatigue. However, only three self-reports exist that include a fatigue of 3, which makes the evaluation for high levels of fatigue difficult. Lastly, when the self-reported engagement was high (3), the algorithm returned an engagement of 3 most frequently. However, the algorithm also returned a level of 1 or 2 many times. To evaluate the algorithm for lower levels of engagement is difficult, as much less self-reports for low engagement were submitted.

Based on these results, we can say that our current heuristic-based algorithms are not good enough to accurately sense the emotional and cognitive state of users. However, undesirable states, like low valence, high fatigue or low engagement were only sparsely reported, making an evaluation for those states difficult. This finding answers RQ1.

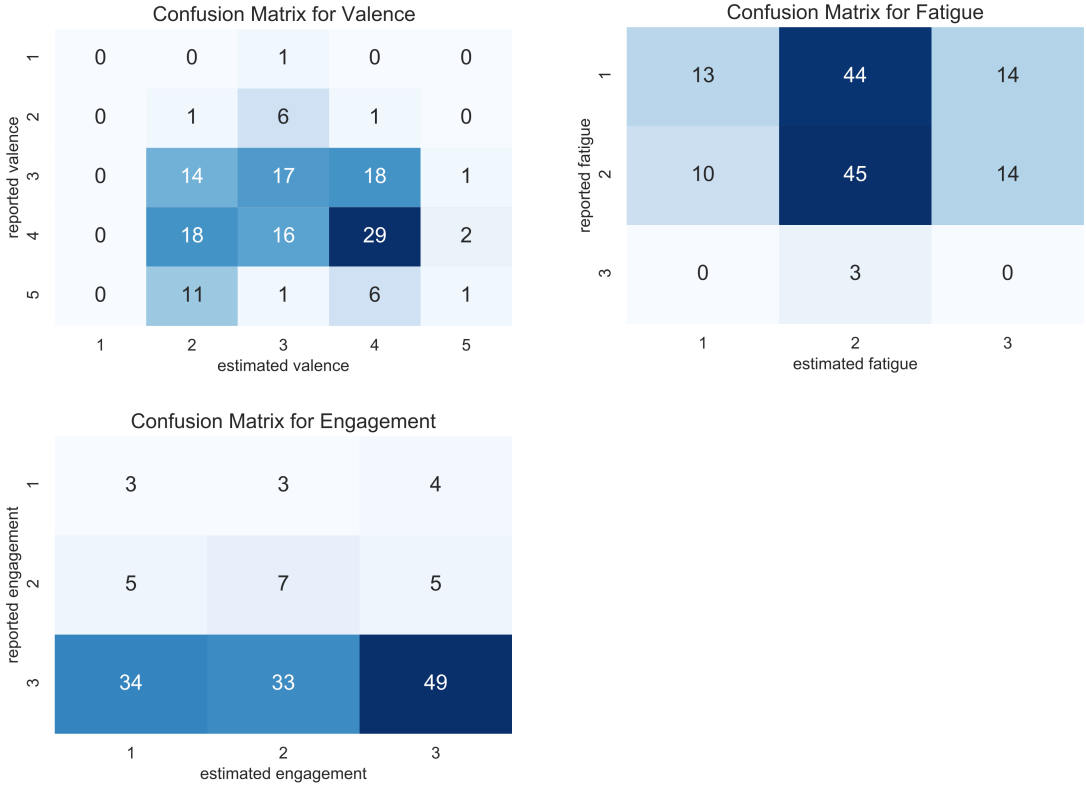

Figure 5.4: Confusion matrices comparing the estimated dimensions with the self-reports.

5.3.2 Fluctuations of Estimates

To get further insights into the behavior of the implemented algorithms, this section looks at the fluctuation of estimated values. While the algorithms that estimate valence and fatigue change their returned value infrequently, the estimates of the user's engagement are subject to high fluctuations. As it is unlikely that the engagement of the user changes that frequently, high fluctuations are undesirable and show potential limitations of the heuristic-based algorithm. Figure 5.5 shows the returned values of the three algorithms. The same user and the same date is chosen as in Figure 5.2, but to not clutter the visualizations, only the morning is depicted (4 hours). Even though only data of P2 is visualized in the figure, similar patterns are present in the data of all participants.

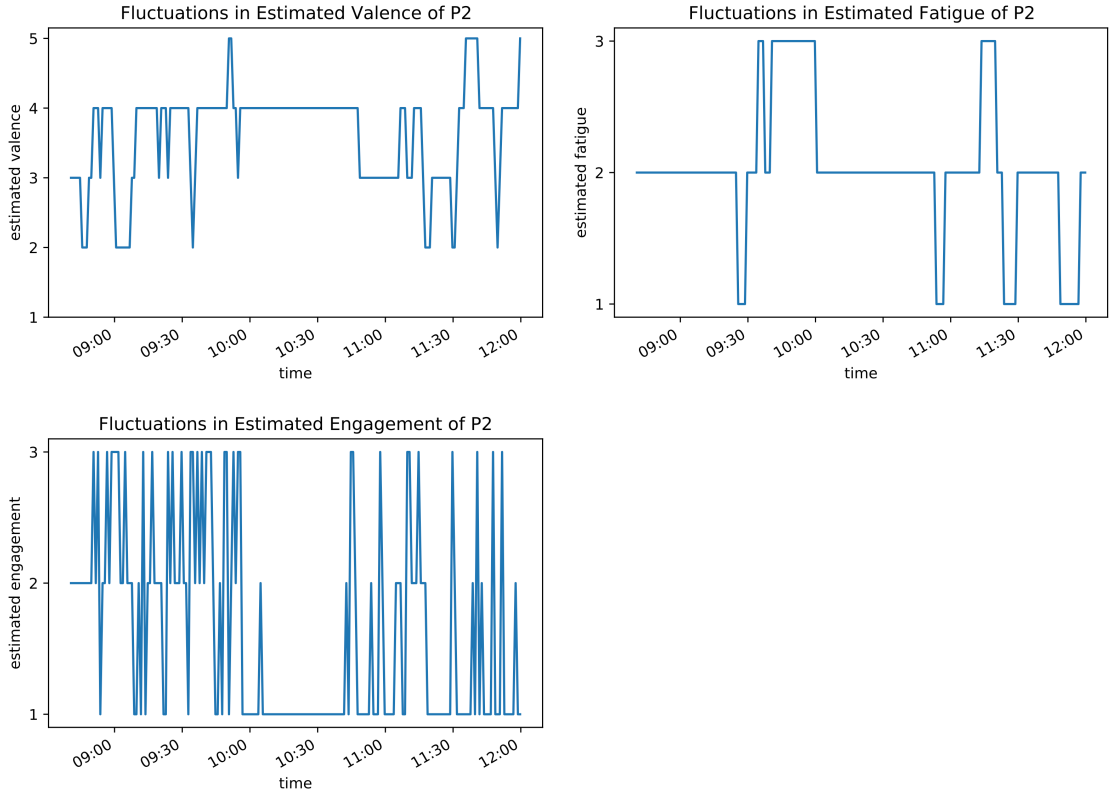

Figure 5.5: Fluctuations in estimated values of P2 over 4 hours.

5.3.3 Perceived Accuracy

Even though previous sections have shown that the accuracy of the EmotionalAwareness-Tool is very limited, participants reported that they think that the accuracy of the tool was reasonably good. This assessment of participants is based on the estimates they saw in the glanceable display during the second phase of the study. P1 mentioned *"I think it's, it's quite accurate. I mean, in sometimes, yeah, probably when, when having my hands up and it covers my face, there were some, [...] funny results that it wasn't sure what, [...] the emotion was. But, but generally I would say it was quite*

nice". According to P2, *"if I'm always neutral in front of the computer or you don't see it really, I think it's, it's hard to, to misjudge it. But I think it worked quite well"*. In addition, P2 mentioned that the fatigue estimation was the most accurate dimension and engagement could be improved the most. P3 reported *"I was quite surprised because I thought it was very accurate, quite accurate"*. P4 was most skeptical, but estimated that the accuracy of the algorithms is around 60 to 70%.

This result could be explained by the fact, that the glanceable display is biasing the participants. P2 explicitly noted that *"You might be focused on your task and then you look at this, this like, oh yeah. There's like the eyes are looking at bit sleepy. Yeah, actually, now that I see it, I feel it also"*. P3 reported *"I was really, let's say also impressed, but I was also really biased by it"* and P1 added *"If the tool says I am tired, I may feel more tired than I actually am"*. This biasing effect is an influence of the glanceable display and therefore contributes to the answer of RQ2c.

Participants also reported that a perfectly accurate algorithm do not have to be the main goal of the approach. According to P1 *"If I deliberately try to, to show some, [...] emotion and in that case, the visualization also shows it, then I'm probably happy [...] even, yeah, if, if in some cases it's not that accurate"*. P3 thinks another aspect could be more valuable than a perfectly accurate algorithm. The person asked *"how can we tell a developer or the user, when is a good time to take a break and try to work as efficient as possible? Because the accuracy does not always help to, to focus on the task"*. However, according to P4, *"accuracy could be improved before it's... Before it's useful"*.

5.4 Possible Improvements of Algorithms

To understand how the implemented algorithms could be improved in future iterations, we look at correlations between the different self-reported dimensions (Section 5.4.1). Further, we discuss how the individual algorithms could be improved based on correlations between the sensed facial and posture related features and the self-reported values (Section 5.4.2 – Section 5.4.4). Lastly, we look at the current system load of the algorithms and how it could be improved in Section 5.4.5.

5.4.1 Correlations Between Dimensions

Table 5.3 gives insights in the relationship between valence, fatigue and engagement. As we have seen in Section 5.2, interpersonal differences can be large. Therefore, we focus on correlations in the data of individual users. Generally, high valence occurs during phases of high engagement or low fatigue. However, interpersonal differences are again present. The connection between valence and engagement is much more pronounced for P3 as it is for P4. P2 mentioned such a connection during the follow-up interview as well: *"when I'm engaged in a task, I think I don't feel sleepy. And if it's like a boring task, I, it might be like a higher risk of being sleepy"*. This finding could be included in a future iteration of the algorithms. Estimations for one dimension could be made dependent on estimates of other dimensions to increase accuracy.

	Correlation			
Metric	P1	P2	P3	P4
Self-Reported Valence <-> Self-Reported Fatigue	-0.676	-0.633	-0.831	-0.485
Self-Reported Valence <-> Self-Reported Engagement	0.578	0.332	0.721	0.272
Self-Reported Fatigue <-> Self-Reported Engagement	-0.711	-0.579	-0.751	-0.613

Table 5.3: Correlations between different self-reported dimensions.

5.4.2 Valence Algorithm

The number of sensed happy expressions per minute seems to be a promising indicator for valence. However, sad expressions, which are currently considered in the algorithm, don't show a correlation. In the following, some facial and posture related features and their correlation with valence are presented in more detail. An overview of all features and their correlation with valence is provided in Table 5.4.

	Correlation to Valence			
Feature	P1	P2	P3	P4
Happy Exp.	0.292	0.218	-0.196	0.141
Neutral Exp.	0.063	-0.022	0.073	-0.03
Sad Exp.	-0.014	0.075	-0.036	0.021
Surprised Exp.	-0.039	-0.014	-0.049	-0.034
Blink Count	0.18	0.079	0.254	-0.375
Shoulder Ear Distance	0.225	0.053	-0.276	0.139
Pitch	0.02	0.086	-0.018	0.035
Roll	-0.12	0.13	-0.133	0.038
Yaw	-0.078	0.061	0.226	-0.118
Head Area	0.097	0.143	0.345	-0.083
Mouth Openness	0.175	0.032	0.013	-0.071
Wrist Is Present	-0.011	-0.105	-0.221	0
Elbow Is Present	0.135	-0.005	-0.035	0.162

Table 5.4: Correlations of facial and posture related features to valence. For each participant, the two features with the highest absolute correlation are highlighted.

Happy Expressions. For P1 and P2, the highest correlation is calculated between valence and the number of happy expressions per minute. For P4, the correlation is smaller but still positive, whereas for P3, a negative correlation is calculated. To better understand how happy expressions are linked to valence, we visualized each participant's mean number of happy expressions per minute for different levels of valence in Figure 5.6. The number of sensed happy expressions is continuously increasing with increasing valence for P1 and P4. The same holds for P2, except that a smaller number of happy expressions was sensed for a valence of 5. However, only one self-report contribute to this measurement. Therefore, this could be an outlier. The number of happy expressions of P3 is continuously decreasing with increasing valence. As the number of happy expressions per minute generally correlates with the perceived valence of the user, this measurement is promising for sensing valence and should be kept in future iterations of the algorithm. However, the way how this facial feature is integrated into the algorithm could be improved by tailoring the algorithm better to individual behavior. This could be done by calculating baseline values for the sensed facial expressions to take interpersonal differences into consideration.

Other Features. Even though the number of sensed happy expressions per minute correlates with valence for most users, the same does not hold for the number of neutral or sad expressions (see Table 5.4 for more details). As sad expressions contribute currently to the estimate made by the algorithm, the not existing correlation affects the algorithm's accuracy. Therefore, this facial expression should be removed in a future iteration. Further, head area is among the two strongest correlations for P2 and P3. This can potentially be explained by the fact that valence

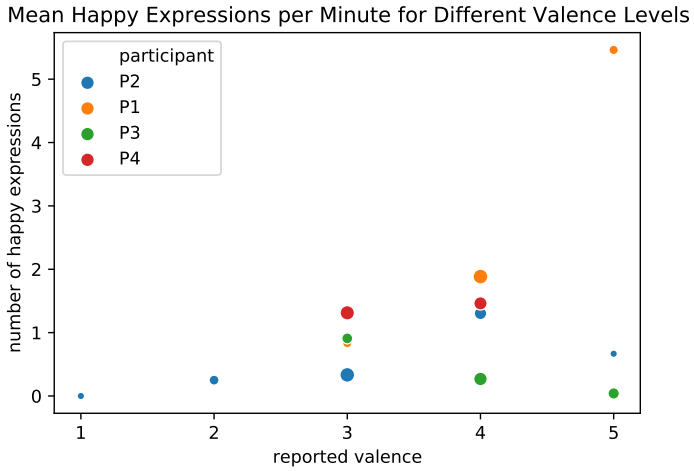

Figure 5.6: Mean happy expressions per minute for different valence levels. The size of the dots represents the number of considered self-reports.

is positively correlated to engagement (see Table 5.3) and head-area is a promising indicator for this dimension (Section 5.4.4). Apart from that, no other facial or pose related feature correlates notably with valence for multiple participants. Therefore, in order to improve the algorithm, raw facial landmark data could be studied to find more indicators for valence.

5.4.3 Fatigue Algorithm

In contrast to the results of the pilot study, no increase in the blink frequency of participants is visible with increasing reported fatigue, potentially due to sensing inaccuracies. Moreover, no other facial or posture related feature shows promising results to detect fatigue. In the following, those findings and possible improvements are presented in more detail. Table 5.5 provides an overview of all features and their correlation with fatigue.

Blink Frequency. According to the literature, an increased blink frequency indicates high fatigue [SGBG08]. In contrast to our pilot study, this effect is not visible in the data of our primary study. Considering different self-reported levels of fatigue, the participants' number of eye blinks per minute remains approximately constant. The mean blink counts per minute for different levels of fatigue are visualized in Figure 5.7. The camera position could affect the sensing accuracy of the EmotionalAwareness-Tool. P1, P2 and P3 used multiple computer screens during the study. If the camera is capturing the user's face from the side, blinks could be missed. In the pilot study, where the results were more similar to findings from the literature, camera positions were more comparable and the users were looking straight towards the camera. To overcome this possible limitation, a future iteration of the algorithm could only sense blinks when the user is looking towards the webcam. Further, it could be required to mount the camera on top of the main screen to get more comparable results.

Yawning. As described in Section 3.2.2, we tried to sense yawning by looking at surprised expressions returned by the face-api model. According to the calculated correlations in Table 5.5,

	Correlation to Fatigue			
Feature	P1	P2	P3	P4
Happy Exp.	-0.031	-0.124	0.105	0.119
Neutral Exp.	0.123	0.045	-0.087	0.306
Sad Exp.	-0.083	-0.09	-0.02	-0.31
Surprised Exp.	0.026	0.134	0.083	-0.082
Blink Count	0.017	-0.061	-0.103	0.068
Shoulder Ear Distance	-0.001	-0.045	0.166	0.001
Pitch	-0.025	0.051	0.052	-0.034
Roll	0.171	0.093	-0.021	0.091
Yaw	0.116	-0.1	-0.011	-0.129
Head Area	0.188	-0.177	-0.379	0.04
Mouth Openness	-0.172	0.022	-0.014	0.04
Wrist Is Present	0.002	0.056	0.157	-0.14
Elbow Is Present	-0.011	0.056	0.068	0.029

Table 5.5: Correlations of facial and posture related features to fatigue. The chosen time window is 15min. For each participant, the two features with the highest absolute correlation are highlighted.

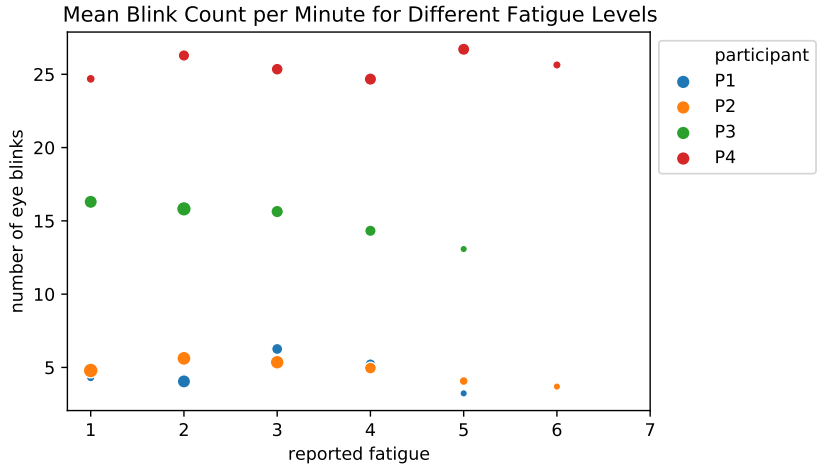

Figure 5.7: Mean blink count per minute for different fatigue levels. The size of the dots represents the number of considered self-reports. The chosen time window is 15min.

there is no notable correlation for multiple participants between the number of surprised expressions and fatigue. In the pilot study, no correlation was present either. Therefore, the assumption that the model returns *surprised* when the user is yawning, does not hold reliably. Therefore, in a future iteration of the algorithm, raw data have to be used to create a more reliable model for yawning detection.

Other Features. By looking at the correlations, no facial or pose related feature seems to be promising to estimate fatigue. Head area is among the two strongest correlations for three participants, but not with the same sign. Further, fatigue is negatively correlated with engagement (Table 5.3), for which head-area is a promising indicator (Section 5.4.4). To make the estimation of fatigue more accurate in a future iteration of the algorithm, it could be useful to add more eye-

related features, as presented in Section 2.2. However, as discussed before, the webcam-based sensing approach could be affected by the camera's viewing angle. Therefore, more robust algorithms should be implemented.

5.4.4 Engagement Algorithm

The user's proximity to the screen, which is estimated by calculating the user's head area, is a promising feature to estimate engagement. Other features that are included in the current algorithm do not show a correlation with engagement for multiple users. More details and possible improvements are provided in the following paragraphs. An overview of the correlations is presented in Table 5.6.

	Correlation to Engagement			
Feature	P1	P2	P3	P4
Happy Exp.	0.086	0.119	-0.18	-0.08
Neutral Exp.	0.344	-0.03	0.105	-0.014
Sad Exp.	0.01	0.058	-0.104	0.021
Surprised Exp.	0.166	-0.055	-0.142	-0.205
Blink Count	0.02	-0.084	0.273	-0.044
Shoulder Ear Distance	0.289	0.09	-0.227	0.238
Pitch	-0.095	0.074	-0.037	-0.001
Roll	-0.12	0.025	-0.022	-0.205
Yaw	-0.049	0.028	0.096	-0.027
Head Area	-0.022	0.193	0.389	0.277
Mouth Openness	0.341	-0.001	0.036	-0.079
Wrist Is Present	-0.014	-0.028	-0.261	-0.032
Elbow Is Present	0.019	0.168	-0.041	0.055

Table 5.6: Correlations of facial and posture related features to engagement. For each participant, the two features with the highest absolute correlation are highlighted.

Head Area. To approximate the distance of the user to the computer screen, respectively to the webcam, we calculated the user's head area. For P2, P3 and P4, the calculated head area has the most positive correlation with engagement among the considered features. Therefore, head area is a promising feature for engagement estimation. To get more insights about this relationship, we visualized the mean sensed head area for different reported levels of engagement in Figure 5.8. As the calculated area for P1 was much higher than the head area of other participants, we applied a multiplication factor of 0.4 to all values of P1. The sensed head area is continuously increasing with increasing engagement for P3 and P4. For P2, a positive trend is also visible, but the mean head area for a self-reported value of 4 is lower than for a self-reported engagement of 3. For P1, no clear trend is present. Since the head area is increasing with increasing engagement for multiple users, this feature should be kept in future iterations of the algorithm.

Wrist Presence. In our approach, we made the assumption that engaged users tend to hold their hand in front of their face. Based on the correlations visible in Table 5.6, this assumption does not hold. The correlations are even slightly negative for all participants. P2 pointed out that *"if I'm programming, I'm more like I need the hands"*. Therefore, our assumption might only hold

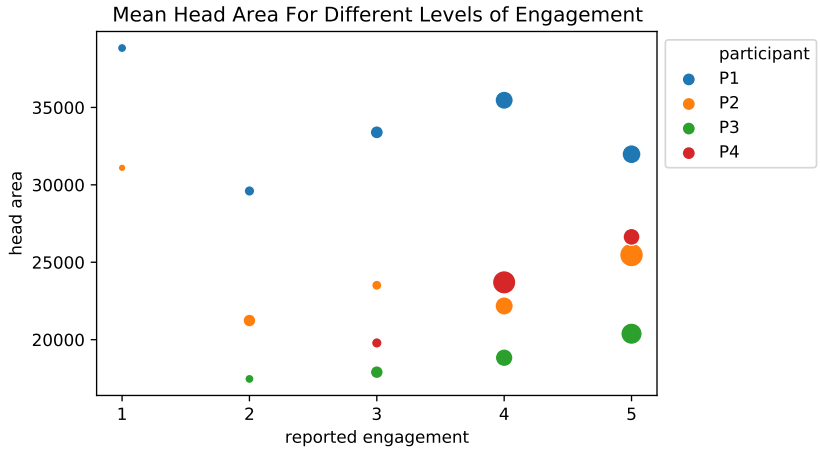

Figure 5.8: Mean head area for different engagement levels. The size of the dots represents the number of considered self-reports.

for certain tasks or is only applicable for a minority of people. Therefore, this feature could be omitted in a future iteration. Another approach would be to include a more sophisticated model that is able to differentiate between different hand poses.

Other Features. To get some insights into the direction of lean of the user, the projected distance between shoulders and ears is calculated by the EmotionalAwareness-Tool. This feature shows positive correlations for P1 and P4, but a negative correlation for P3. It's noteworthy that the correlations are therefore different from the correlations of the head area, which tries to approximate the user's distance to the screen. Probably, the viewing angle of the camera affects both measurements in different ways. Other features used for engagement estimation (head orientation, mouth openness, elbow presence) do not show notable correlations to engagement for multiple participants. Three out of four participants used multiple screens during the study. This could affect the results regarding head orientation. Even if the assumption that disengaged people tend not to focus on the screen held, it would be difficult to see this in the stored data, since users are looking at multiple screens while working. P1 noticed this behavior in the interview. *"Whenever I'm looking to the left or the right of my monitors it probably sometimes thinks that I am looking out of the window"*. In a future iteration of the algorithm, this effect could be reduced by requiring a standardized camera position on top of the mainly used screen. Further, to find more features that indicate engagement, the recorded raw data could be analyzed.

5.4.5 System Load of Algorithms

The current prototype of the EmotionalAwareness-Tool requires an amount of computing power that is noticeable by the user. Especially, running the model for blink detection at 20 Hz requires a lot of processing power. Two participants mentioned the high system load in the follow-up interviews. P2 reported *"So this was maybe the one negative thing that my computer was getting hotter than normal. And also like the fan was running louder than normally"*. The participant used macOS, an 8-Core CPU and 32 GB of RAM. The mean overall CPU usage of the participants' machines while running the tool was between 15% and 25%. All participants used computers with at least 8-Core CPUs and 16 GB of RAM. As the processing power consumption of the EmotionalAwareness-

Tool is recognizable, the system load of the algorithms should be reduced in future iterations. This could be achieved by reducing the frequency the models are evaluated. However, this is not possible for the model that is used to detect eye blinking, which is also used to calculate the user's head area and mouth openness. As this model cannot be evaluated less frequently while remaining accurate, another approach could be to run the model only selectively at the frequency of 20 Hz and slower otherwise. For example, after 3 minutes of running the tool at 20 Hz and sensing eye blinking, the tool could evaluate the model only every second and could not sense blinks for 3 minutes. After this, the frequency could be increased again. In the current algorithm, data from the preceding 15 minutes is considered for fatigue estimation. Therefore, this proposed approach should not have a major effect on the real-time aspect of the visualization. In addition, since features that rely on the posenet model do not show promising results, this model could be removed to reduce system load.

5.5 Visualization Approach

To answer RQ2, we present feedback from the participants about the visualization approach in this section. Data from the follow-up interviews as well as feedback collected on the second page of the self-reporting pop-up is considered.

Value of the Approach. RQ2d focuses on the value users see in the chosen approach. In the performed follow-up interviews, two participants mentioned that they think they are generally aware of their emotional and cognitive state. The two other participants noted that now, after doing self-reporting for several days, they think that their awareness before the study was not as good as they thought it would be. P2 pointed out that *"I would say I'm aware, but I think I'm not reflecting about it enough"*. However, all participants mentioned that self-reporting valence, fatigue and engagement increased their awareness during the period of the study and that they learned thereby something about themselves.

All participants reported that they think that being aware of their emotional and cognitive state is valuable. P3 mentioned that *"If you know, when you're wide awake, then you could push the tasks that require, let's say more, cognitive load or higher demanding tasks into these sections. And then maybe when you realize in the afternoon, for example, you're not as awake, then you could usually do like smaller tasks"*. According to P4, *"If I find myself distracted, then I maybe see it as an opportunity to take a break and get some more tea or a snack or walk around a bit and then come back to the screen"*. P2 explained that an increased awareness could be useful because *"normally, I think I notice only when I'm almost falling asleep and then it's like, the last hour was not productive"*.

In this regard, all participants mentioned that the glanceable display made them more aware of their current emotional and cognitive state. *"This would really help me to monitor myself better"* (P3). But the EmotionalAwareness-Tool did not only increase awareness by sensing the participants' current state, but also as P4 described *"I was oftentimes think like, kind of looking into myself and thinking like, okay, does this match or not?"*.

However, participants mentioned that a tool making them aware of their current state is not in all cases useful or beneficial. P1 mentioned that *"sometimes maybe when I want to be productive while being a little bit tired, it doesn't help if, if the face tells me very clear, 'Hey, you're tired. You should probably take a break' when I don't want to"*. P2 noted that *"if I would have multiple tasks, I could switch in-between and say like, now I'm too sleepy. I should do something else now. [...] But for my study time, it was not possible to switch to something else"* Further, three participants reported, that the tool is not useful during phases of high cognitive load or when thoroughly engaged, because then the user is very focused and could easily be distracted.

Influence. RQ2c is interested in the influence and learnings from the glanceable display. Besides increasing the participant's awareness, the tool led to tangible actions for P3. The person reported *"When I saw that the person was not happy. I tried to actively change my mood or actively try to be in a happier state to be more efficient"* and regarding the background color *"I tried to get out of my red state as fast as possible"*. In addition, P4 tried to outsmart the tool when the estimated valence was off. *"I would like try to prove to it that I'm not sad"* P4 reported. As mentioned earlier, the glanceable display was biasing participants in how they perceived their valence, fatigue and engagement. In addition, P3 added *"When it tells me that maybe I'm not as motivated, then I'm more biased towards taking a break and when it shows me a happy face, then I'm also more biased to continue working"*.

Intrusiveness. As related approaches can be perceived as intrusive, we asked participants if the glanceable display was distracting them. The reactions were mixed. P3 and P4 mentioned that it was sometimes distracting. *"I did find myself looking at it quite often to see [...] what is it revealing about my inner self. So it was actually like a little bit distracting perhaps in that way"* (P4). *"And then I just moved away from my work and I was more focused on the visualization and how it represented my state"* (P3). P1 answered *"Even if, if it's very small and, and somewhere on the edge of the screen, you're pretty much constantly looking at it and... Or, or you're, you're sensing whenever it changes"*. However, P1 added that this effect decreased with time. P2 on the other hand didn't think that the glanceable display was distracting because the participant moved the window to a second screen. P3 moved the window to another screen as well from time to time to get less distracted. P4 pointed out that sensing inaccuracies have an effect on the tool's intrusiveness. *"Other times it was like show that I was distracted, even though I thought I was being engaged. So, so that difference, I guess, kind of made it a bit also distracting"*.

Chosen Visualization. On the second page of the self-reporting pop-up, participants rated the accuracy of the visualization in terms of valence and engagement as *good*, the depiction of fatigue was assessed as slightly below *good*. In more detail, participants were asked to rate the accuracy of the visualization, which was based on their previously submitted self-report, on a scale of 1 (labeled as *very poor*) to 5 (labeled as *excellent*). On average, the reaction of participants towards valence, which was called *mood* in the asked question, was 4.03, standard deviation 0.85. For fatigue, which was called *sleepiness*, the mean rating was 3.75, standard deviation 1.12. Lastly, regarding engagement, the mean reported accuracy was 4.17, standard deviation 1.03. The mean ratings of all participants and dimensions is shown in Figure 5.9. Remarkably low is the rating for fatigue of P1. The person noted that *"The eyes are too wide open"* and *"I don't have my mouth wide open when I am awake"*. The wide open mouth was caused by a self-reported valence of 5. Based on those results, we can say that it's possible to accurately visualize the emotional and cognitive state of a person by the chosen visualization. By creating an understandable visualization approach, RQ2a is answered. The ratings submitted by participants answer RQ2b.

P3 explicitly noted that he likes the idea that a face is used to visualize data. *"I think it's a good idea to use a face because usually you can relate to a face and think about how you would look like from an outside perspective and if you would make this face currently"*. However, two participants noted that there is a mismatch between how the emotional and cognitive state is depicted in the glanceable display and how they actually look like at this moment. P1 noted that *"the smile [...] is always very extreme because I mean, nobody can even smile that way"*. P4 mentioned *"I could tell the logic, like, okay. I'm more tired. So then the eyes are more droopy or something, but I don't think it reflected, what I actually looked like in that moment"*. Further, all four participants were unable to make sense out of the changing background color, as this visual cue was described nowhere. In addition, the only female participant (P4) reported, that the customization options of the glanceable display should be extended to better mirror herself.

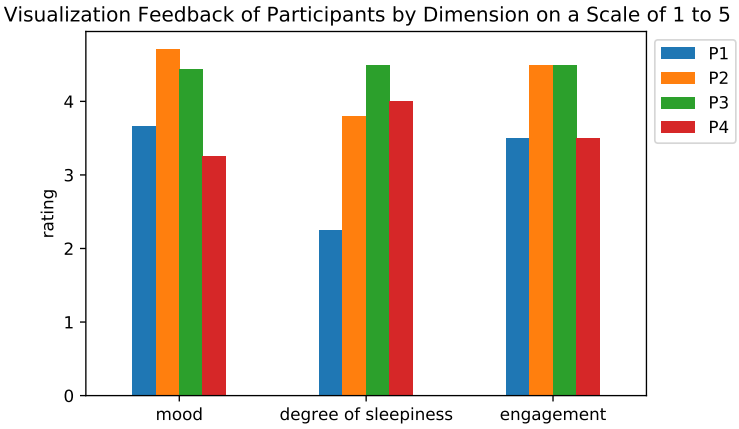

Figure 5.9: Visualization feedback of participants by dimension on a scale of 1 to 5.

5.6 Summary of the Results

The analysis has shown some interesting insights in the self-awareness of users, their emotional and cognitive state and the sensing and visualization approach of the EmotionalAwareness-Tool. The key findings of the analysis are summarized in Table 5.7.

#	Finding	Section
F1	People differ in how they physically express their emotional and cognitive state.	5.2
F2	The accuracy of the heuristic based algorithms that sense valence, fatigue and engagement is very limited.	5.3.1
F3	Participants perceived the accuracy of the tool as reasonably good, possibly because the glanceable display was biasing them.	5.3.3
F4	Only a few of the implemented facial and posture related features have a correlation to valence, fatigue or engagement for multiple participants.	5.4
F5	All participants think that being aware of their emotional and cognitive state is valuable.	5.5
F6	All participants mentioned that the glanceable display made them more aware of their emotional and cognitive state.	5.5
F7	Even though we tried to design our tool as unobtrusive as possible, participants still got distracted by it.	5.5

Table 5.7: Summary of key findings of the study.

Chapter 6

Threats and Limitations

The primary threats and limitations of this work are the short intervention phase of the study (only 1 to 2 days), the not homogeneous set of participants and the fact that participants could behave differently when being recorded by a webcam.

Short Intervention Phase. The intervention phase of the performed user study lasted only for 1 or 2 days, depending on the time availability of the participant. In the follow-up interviews, we asked participants about their learnings and how the glanceable display influenced them. We assume that the answers we received for those questions would have been different and more insightful if the second phase of the study had been longer. Further, a participant noted that she or he perceived the glanceable display as less distracting over time. Therefore, our findings towards the intrusiveness could not be applicable for a long-term usage of the EmotionalAwareness-Tool. Lastly, we assume that by performing a longer intervention phase, participants would have noticed the limited accuracy of the estimated valence, fatigue and engagement.

Homogeneous Set of Participants. All participants of the user study were recruited from the personal and professional network of the researchers. By doing so, all participants already were aware of the general topic of this work, before their recruitment. This could have affected their responses during the interviews. Further, the set of participants was homogeneous in terms of demographics. All participants are between 23 and 30 years old and are working in the field of computer science. This is only a very limited sample for the population of knowledge workers. As we found out that interpersonal differences in how participants express their emotional and cognitive state can be of importance, our set of participants lacks cultural and ethnic diversity. More diverse participants as well as a larger user group would have made the results more insightful.

Hawthorne Effect. A threat to the approach could be that users behave differently when they know that they are captured by a camera. This so called Hawthorne Effect was explicitly pointed out by P3 and is also discussed in related work that uses webcams [BSN+20]. Not only how people behave in front of the computer could be affected by the study, but also how they self-report. As discussed by Soto et al. [SSF+21], participants could be afraid to report undesirable states, like very low engagement or high fatigue. This would affect the results of the study.

Chapter 7

Discussion

Based on the results and limitations presented in the preceding chapters, this chapter discusses the main findings and how the approach could be improved. Further, future work is presented.

7.1 Discussion

Value of Awareness. One of the main assumptions of our approach is that self-awareness is useful and desirable. This assumption was confirmed by participants of the study. Further, we have seen that a digital aid, like the EmotionalAwareness-Tool, can help users to be more aware of their emotional and cognitive state. In our approach, this was achieved by showing them their sensed state in real-time, but also because users were questioning the sensed estimates. To understand the value of awareness better, it is interesting to examine the submitted self-reports of the participants. By looking at the self-reported values of valence, fatigue and engagement, it becomes apparent that the perceived emotional and cognitive state of users changes over a day. As discussed earlier in this work, to make users more aware of those short and long-term states could be supportive. For example, participants mentioned that they can take countermeasures when they are aware of an undesirable state. However, the causes of those changes were not investigated in this work. We assume that short-term fluctuations can be task-dependent and like this, the EmotionalAwareness-Tool could encourage users to switch the task when being in an undesirable state. In this regard, it could be useful to investigate the main root causes of changes in the emotional and cognitive state of users in a future study (see Section 7.2).

Intrusiveness & Fluctuations. Many related approaches can be perceived as intrusive, since they require body-worn sensors and make use of interventions that can affect the user's concentration [CAJ+16]. This is why we tried to design our tool as unobtrusive as possible, by relying on webcams and making use of a glanceable display. However, participants still got distracted by the always on top visualization. It has to be noted though that the participants used the glanceable display only for one or two days and one participant even reported, that the display's intrusiveness decreased over time. Therefore, only a longer user study could give more insights in the intrusiveness of the approach.

There is a connection between the perceived intrusiveness of the glanceable display and fluctuations of the estimates. As presented in Section 5.3.2, the estimated valence and fatigue change their value less frequently than the estimation of engagement. It is not only unlikely that the high fluctuations of the engagement estimate correctly represent the user's cognitive state, the high fluctuations were also distracting participants. P1 looked at the glanceable display whenever it

was changing. Consequently, high fluctuations are undesirable, as they increase the tool's intrusiveness. The algorithms for valence and fatigue are implemented as state machines. Therefore, a new estimate is always depending on the last estimate. The algorithm for engagement is not implemented this way, which could explain the high fluctuations. Hence, a state machine based sensing approach seems to be more useful, at least when real-time visualization is provided. A smoothing of the estimates could also reduce fluctuations. However, this could impair the real-time aspect of the visualization approach.

How to Inform Users About Their State. Since participants got distracted by the glanceable display when it was changing, different approaches to inform the user about their emotional and cognitive state should be considered to extend or replace the always on top display. P4 mentioned *"I can imagine it would be useful if, if it does pop up at moments when like, could use an intervention or a break or something"*. Also, P3 asked *"how can we tell a developer or the user, when is a good time to take a break and try to work as efficient as possible?"*. This process of reminding the user of taking countermeasures against an undesirable state can be described as *nudging* [TS08]. The approach of this work is based on the idea that users gain useful insights by examining visualized data. However, as mentioned in Section 2.4, related work has pointed out that those insights not necessarily lead to further actions by people. A nudging based approach could therefore not only make users aware of undesirable states, but also actively help them to leave those states. It has to be noted though that in such an approach, the right moment to show a nudge is crucial. Nudging at an intrusive time can be counter-productive [BK06] and users might stop using the tool or not trust in the information anymore. Three participants of the study explicitly mentioned that they do not want to be interrupted by a digital aid during phases of high engagement. How an approach that makes use of nudging could look like is described in Section 7.2.

Perceived Accuracy. As the accuracy of the sensing approach is very limited, the high perceived accuracy by the users is surprising. This leak of skepticism towards the accuracy of a digital aid was also reported by Snyder et al. [SMC+15] in *MoodLight*. In our study, the high perceived accuracy could be caused by at least two effects. On the one hand, self-reports were not equally distributed over all possible values. The algorithms returned a valence of 4, a fatigue of 2 and an engagement of 3 most of the time. Therefore, as long as the user is in a non-extreme state, the inaccuracy might not be apparent. On the other hand, as reported by participants, the glanceable display could be biasing. If this is the case, future studies have to be careful to not influence the user negatively by providing incorrect estimates. Further, ethical concerns have to be taken into consideration as the user's emotions could get manipulated.

Improvements of Sensing Approach. A key finding of the study is that people differ in how they physically express their emotional and cognitive state (F1). This is apparent in the answers of the four participants during the follow-up interviews, as well as in the correlations between the reported dimensions and the sensed facial and posture related features. Another key finding is that the accuracy of the heuristic-based algorithms is very limited (F2). To some extent, F2 could be caused by F1. The heuristic-based algorithms to estimate valence, fatigue and engagement were designed and tested based on assumptions that might only hold for the authors of this work. This could explain the bad accuracy of the estimates. However, a simple-heuristic based approach might also be unsuitable if it was better tailored to individuals because the interviews revealed that the behavior of users can be task-dependent. Therefore, a more sophisticated estimation approach could be needed to increase accuracy. A machine learning based approach could help to identify behaviors that are signs of valence, fatigue and engagement of individual users. By performing self-reporting while webcam data is collected, an individual model could be created

during a first phase of using the EmotionalAwareness-Tool. This model could then be used to predict valence, fatigue and engagement in a following phase.

If a heuristic-based approach should remain in place, the accuracy could potentially be improved by a better pre-processing of the sensed data. As no required camera position was specified, the participants were recorded from different angles in our study. It is likely that this impaired the accuracy of the approach. However, the goal was to develop a tool that is applicable in real-world scenarios, where a standardized camera position is not desirable. A better approach would be to improve automatic pre-processing. For example, blink counts could only be taken into consideration if the user is looking towards the camera for a predefined time. Or the tool could ask users in the beginning to look at all their computer screens one by one and save the corresponding head orientations. Then, by using this information, the EmotionalAwareness-Tool could differentiate better between viewing towards a second monitor or towards the window or smartphone.

7.2 Future Work

As mentioned before, the sensing approach could be improved in future work. For example, by making use of machine learning or better pre-processing. To further improve the accuracy of the estimates, alongside webcam data, keyboard and mouse input as well as context information such as window titles could be considered. Those input features do not increase the intrusiveness of the approach and could give insights in a potential task dependency of the emotional and cognitive state of users.

Our approach tries to visualize the current emotional and cognitive state of users. In a future approach, it could be valuable to focus more on changing states rather than reached states. For example, the user could be notified if the sensed fatigue was continuously increasing over the last 30 minutes. We asked participants in the interviews to share their thoughts about this idea, and they mentioned that changes could be even more insightful. P2 argued that *"when you're looking very happy in the middle of like a boring task, the tool might actually notice that you're now procrastinating"*. Further, notifying users about changes could prevent them from entering undesirable states. P1 reported in this regard *"if I'm already tired, then I probably know it [...] but if I was not tired, but I start to get tired, it's probably more helpful to be notified of this"*.

As discussed earlier, an approach that makes use of nudging could be the subject of future work. When an undesirable emotional or cognitive state is sensed or the user is heading towards such a state, a notification could be sent. This nudge could contain explicit ideas of what users can do to improve their state, depending on the sensed data. However, those ideas should go beyond a prompt to take a break, as P3 and P4 noted that not all types of breaks have the same effect on them. For example, a short walk could be more helpful than watching a video. Ideas of countermeasures mentioned by the participants include getting a coffee or switching to a less demanding task when being fatigued. Further, one participant proposed doing a short meditation session or watching some funny videos when being in a bad mood. In addition, participants mentioned standing up, looking out the window, eating something, taking a nap, taking a few steps and going outside as countermeasures to improve their state. Those ideas could be included in the notifications to nudge users.

A future approach could also keep the glanceable window and try to make it less intrusive. As the intrusiveness of the glanceable display can to some extent be related to the frequently changing visualization, a future approach could improve the way data is visualized in real-time. One idea would be to make transitions between visualized states more dynamic, to decrease the intrusiveness of a change. Another idea would be to visualize high or low estimates only when extreme states are sensed. For example, the visualization could not change if the user is

slightly disengaged. Only when the user is slowed-down and should consider taking a break, the visualization changes.

Chapter 8

Conclusion

Emotional awareness can help people to recognize their emotions, to reason about them, and to (pro-)actively regulate them when needed. In this work, we presented the EmotionalAwareness-Tool, which tries to increase the users' awareness by visualizing their emotional and cognitive state in real-time on a glanceable, always on top display. To this end, we tried to sense the user's valence, fatigue and engagement using a regular webcam.

A user study revealed that an improved self-awareness is indeed useful and desirable. Even though the sensing accuracy of the EmotionalAwareness-Tool is limited, it helped users to be more aware of their emotional and cognitive state. The tool's accuracy was perceived by participants as reasonably good, possibly because the glanceable display was biasing them. Many related approaches can be perceived as intrusive. Therefore, we tried to design our tool as unobtrusive as possible. However, participants still got distracted by the glanceable window. Further, the study has shown that people express their emotional and cognitive state very differently, which makes sensing their state difficult.

As interpersonal differences have limited the accuracy of our implemented sensing algorithms, future work could improve the approach by adapting the algorithms better to individuals. A machine learning based approach could help to identify behaviors that are signs of valence, fatigue and engagement of individual users. Further, the glanceable display could be expanded by providing nudging when certain conditions are met, to actively help the user leaving undesirable states.

Bibliography

[ABD17] Muhammad Awais, Nasreen Badruddin, and Micheal Drieberg. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors*, 17:1991, 08 2017.

[BK06] Brian P. Bailey and Joseph A. Konstan. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22(4):685–708, 2006. Attention aware systems.

[BL94] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[BR03] Kirk Brown and Richard Ryan. The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of personality and social psychology*, 84:822–48, 05 2003.

[BSN+20] Ebrahim Babaei, Namrata Srivastava, Joshua Newn, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. Faces of focus: A study on the facial cues of attentional states. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[BWvdB09] Leticia S. S. Bialoskorski, Joyce H. D. M. Westerink, and Egon L. van den Broek. Mood swings: An affective interactive art system. In Anton Nijholt, Dennis Reidsma, and Hendri Hondorp, editors, *Intelligent Technologies for Interactive Entertainment, Third International Conference, INTETAIN 2009, Amsterdam, The Netherlands, June 22-24, 2009. Proceedings*, volume 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 181–186. Springer, 2009.

[CAJ+16] Jean Costa, Alexander T. Adams, Malte F. Jung, François Guimbretière, and Tanzeem Choudhury. Emotioncheck: leveraging bodily signals and false feedback to regulate our emotions. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016.

[Che73] Herman Chernoff. The use of faces to represent points in k- dimensional space graphically. *Journal of the American Statistical Association*, 68:361–368, 1973.

[CMT+08] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby,

Ian Smith, and James A. Landay. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1797–1806, New York, NY, USA, 2008. Association for Computing Machinery.

[DB09] Matjaž Divjak and Horst Bischof. Eye blink based fatigue detection for prevention of computer vision syndrome. *Proceedings of the 11th IAPR Conference on Machine Vision Applications, MVA 2009*, pages 350–353, 01 2009.

[DD18] Kevin Doherty and Gavin Doherty. Engagement in hci: Conception, theory and measurement. *ACM Computing Surveys*, 51:1–39, 11 2018.

[DDB01] P.D.M.I.S.M.T.H. Davenport, T.H. Davenport, and J.C. Beck. *The Attention Economy: Understanding the New Currency of Business*. The Attention Economy: Understanding the New Currency of Business. Harvard Business School Press, 2001.

[DSA+11] Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. Understanding the impact of video quality on user engagement. *SIGCOMM Comput. Commun. Rev.*, 41(4):362–373, August 2011.

[DTT19] Ed Diener, Stuti Thapa, and Louis Tay. Positive emotions at work. *Annual Review of Organizational Psychology and Organizational Behavior*, 7, 09 2019.

[DZP+20] Na Du, Feng Zhou, Elizabeth M. Pulver, Dawn M. Tilbury, Lionel P. Robert, Anuj K. Pradhan, and X. Jessie Yang. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transportation Research Part C: Emerging Technologies*, 112:78–87, 2020.

[FDK+09] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A. Landay. *UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits*, page 1043–1052. Association for Computing Machinery, New York, NY, USA, 2009.

[Fis00] Cynthia D. Fisher. Mood and emotions while working: missing pieces of job satisfaction? *Journal of Organizational Behavior*, 21(2):185–202, 2000.

[FYS07] Xiao Fan, Bao-Cai Yin, and Yan-Feng Sun. Yawning detection for monitoring driver fatigue. In *2007 International Conference on Machine Learning and Cybernetics*, volume 2, pages 664–668, 2007.

[GKMN07] Michael Grimm, Kristian Kroschel, Emily Mower, and Shrikanth Narayanan. Primitives-based evaluation and estimation of emotions in speech. *Speech Communication*, 49(10):787–800, 2007. Intrinsic Speech Variations.

[GLNS21] Daniela Girardi, Filippo Lanubile, Nicole Novielli, and Alexander Serebrenik. Emotions and perceived productivity of software developers at the workplace. *IEEE Transactions on Software Engineering*, June 2021.

[GQEME15] Christian Grillon, David Quispe-Escudero, Ambika Mathur, and Monique Ernst. Mental fatigue impairs emotion regulation. *Emotion (Washington, D.C.)*, 15, 02 2015.

[GWA15] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. Do feelings matter? on the correlation of affects and the self-assessed productivity in software engineering. *Journal of Software: Evolution and Process*, 27(7):467–487, 2015.

[HDZ72] Eric Hoddes, William Dement, and Vincent Zarcone. The development and use of the stanford sleepiness scale (sss). *Psychophysiology*, 10:431–436, 01 1972.

[HMO+21] Javier Hernandez, Daniel McDuff, Ognjen, Rudovic, Alberto Fung, and Mary Czerwinski. Deepfn: Towards generalizable facial action unit recognition with deep face normalization, 2021.

[KBH10] Iftikhar A. Khan, Willem-Paul Brinkman, and Robert M. Hierons. Do moods affect programmers' debug performance? *Cognition, Technology & Work*, 13:245–258, 2010.

[KM19] Nataliya Kosmyna and Pattie Maes. Attentivu: a biofeedback device to monitor and improve engagement in the workplace. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1702–1708, 2019.

[KvDWBI16] Elisabeth Kersten van Dijk, Joyce Westerink, Femke Beute, and Wijnand Ijsselsteijn. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human-Computer Interaction*, 32, 01 2016.

[LdDOS17] André Teixeira Lopes, Edilson de Aguiar, Alberto F. De Souza, and Thiago Oliveira-Santos. Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. *Pattern Recognition*, 61:610–628, 2017.

[LS13] Yisi Liu and Olga Sourina. Real-time fractal-based valence level recognition from eeg. In Marina L. Gavrilova, C. J. Kenneth Tan, and Arjan Kuijper, editors, *Transactions on Computational Science XVIII*, pages 101–120, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[MICJ14a] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. *Conference on Human Factors in Computing Systems - Proceedings*, 04 2014.

[MICJ14b] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. Capturing the mood: Facebook and fast-to-face encounters in the workplace. In *Proceedings of CSCW 2014*, February 2014.

[MKK+12] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. Affectaura: An intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 849–858, New York, NY, USA, 2012. Association for Computing Machinery.

[MR11] Marwa Mahmoud and Peter Robinson. Interpreting hand-over-face gestures. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, pages 248–255, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[MSM+16] Andrew Moore, Jeff Savinda, Elizabeth Monaco, Jamie Moyes, Denise Rousseau, Samuel Perl, Jennifer Cowley, Matthew Collins, Tracy Cassidy, Nathan VanHoudnos, Palma Buttles-Valdez, Daniel Bauer, and Allison Parshall. The critical role of positive incentives for reducing insider threats. Technical Report CMU/SEI-2016-TR-014, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2016.

[ORM+08] Christoph Obermair, Wolfgang Reitberger, Alexander Meschtscherjakov, Michael Lankes, and Manfred Tscheligi. Perframes: Persuasive picture frames for proper posture. In *Proceedings of the 3rd International Conference on Persuasive Technology*, PERSUASIVE '08, page 128–139, Berlin, Heidelberg, 2008. Springer-Verlag.

[PCB+07] Christos Papadelis, Zhe Chen, Panagiotis Bamidis, Ioanna Chouvarda, Evangelos Bekiaris, and N. Maglaveras. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118:1906–22, 10 2007.

[PZC+18] George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, and Kevin Murphy. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. *CoRR*, abs/1803.08225, 2018.

[QZL04] Qiang Ji, Zhiwei Zhu, and P. Lan. Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53(4):1052–1068, 2004.

[Rus80] James Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980.

[SC16] Tereza Soukupová and Jan Cech. Eye-blink detection using facial landmarks. In *21st Computer Vision Winter Workshop*, 2016.

[SGBG08] Robert Schleicher, Niels Galley, Susanne Briest, and Lars Galley. Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, 51:982–1010, 08 2008.

[SHZ+18] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018.

[SMC+15] Jaime Snyder, Mark Matthews, Jacqueline T. Chien, Pamara Chang, Emily Sun, Saeed Abdullah, and Geri Gay. Moodlight: Exploring personal and social implications of ambient display of biosensor data. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2015.

[SSF+21] Mauricio Soto, Chris Satterfield, Thomas Fritz, Gail C. Murphy, David C. Shepherd, and Nicholas Kraft. Observing and predicting knowledge worker stress, focus and awakeness in the wild. *International Journal of Human-Computer Studies*, 146:102560, 2021.

[SSGR02] Wilmar Schaufeli, Marisa Salanova, and Vicente González-Romá. The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3:71–92, 02 2002.

[SWMS12] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M. Shapiro. *Stanford Sleepiness Scale (SSS)*, pages 369–370. Springer New York, New York, NY, 2012.

[TS08] R.H. Thaler and C.R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, 2008.

[ULS18] Muhammad Umair, Muhammad Hamza Latif, and Corina Sas. Dynamic displays at wrist for real time visualization of affective data. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, DIS '18 Companion, page 201–205, New York, NY, USA, 2018. Association for Computing Machinery.

[VCl+14] Gaetano Valenza, Luca Citi, Antonio lanatà, Enzo Scilingo, and Riccardo Barbieri. Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Scientific reports*, 4:4998, 05 2014.

[ZCM+17] Manuela Züger, Christopher Corley, André N. Meyer, Boyang Li, Thomas Fritz, David Shepherd, Vinay Augustine, Patrick Francis, Nicholas Kraft, and Will Snipes. *Reducing Interruptions at Work: A Large-Scale Field Study of FlowLight*, page 61–72. Association for Computing Machinery, New York, NY, USA, 2017.

[ZMMF18] Manuela Züger, Sebastian C. Müller, André N. Meyer, and Thomas Fritz. *Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors*, page 1–14. Association for Computing Machinery, New York, NY, USA, 2018.

Appendices

A Component Diagram of Tracker

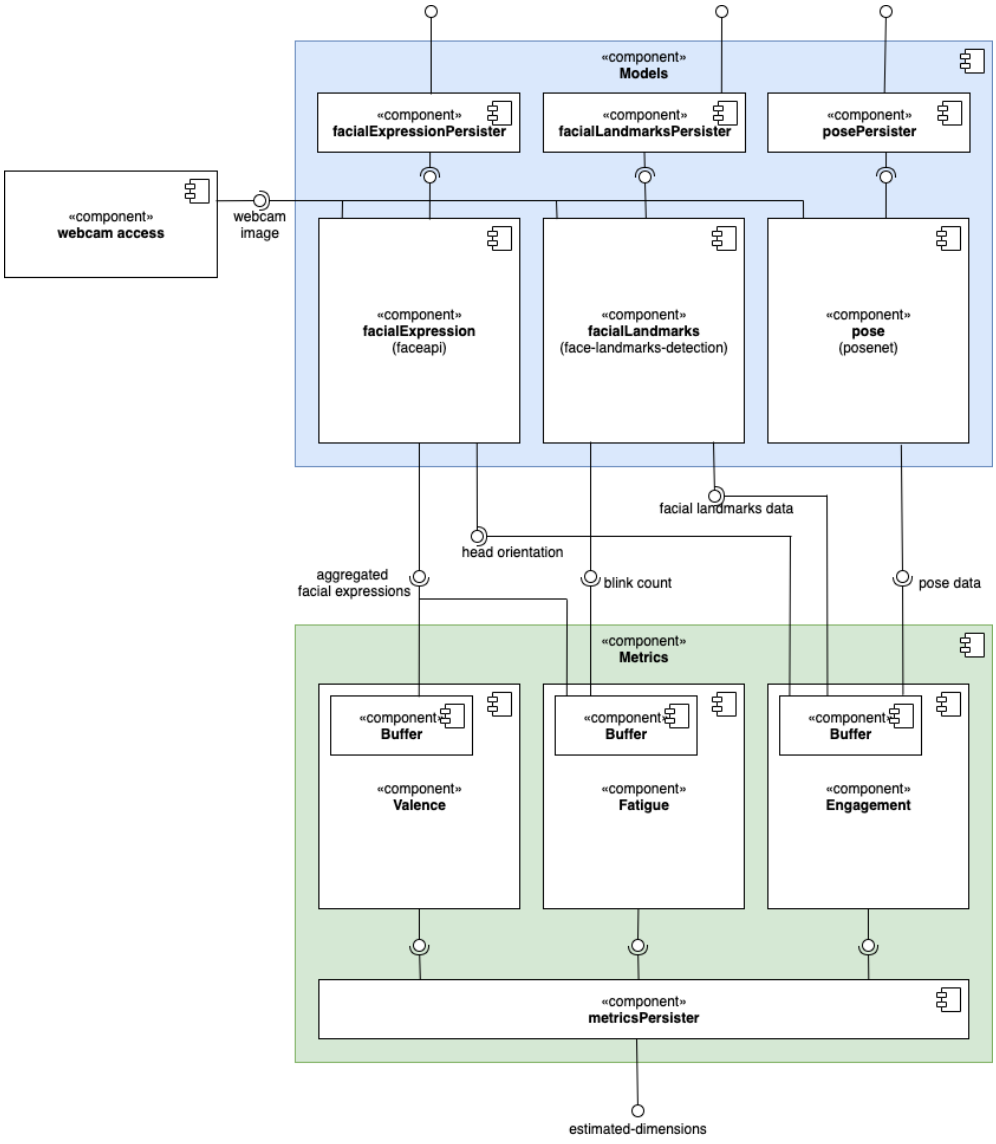

Figure 1: Component diagram of tracker.

B Survey Questions

- **Generally, how much of your work time are you spending in front of the computer? (enter screen time in %)**
- **What kind of webcam did you use during the study?**
  - laptop webcam
  - external webcam
  - professional camera
  - other
- **Did you use multiple computer screens?**
  - yes
  - no
- **What is your job title?**
- **What is your job role?**
  - individual contributor
  - leader
  - other
- **What gender do you identify as?**
  - female
  - male
  - other
  - prefer not to say
- **What is your age in years?**

C Guiding Interview Questions

Self-Reporting and Awareness on Mood/Fatigue/Engagement

1. [RQ-2c] Did you learn anything about yourself from self-reporting your mood, fatigue & engagement and if so, what?
2. [RQ-2d] Are you generally aware or do you generally reflect on your mood/fatigue/engagement during your workday?

happiness/mood

3. [RQ-2d] Do you usually notice immediately when you are happy or unhappy? If so, what kind of impact does it have on you, and how do you react to it?
4. [RQ-2d] Would it be helpful to know immediately when you are happy or unhappy during your workday and if so, why/how?
    - [If they don't come up with any] Assuming you would find out that you are always unhappy before lunch. Would you schedule meetings differently having these new insights?

fatigue

5. [RQ-2d] Do you usually notice when you are fatigued or sleepy? If so, what kind of impact does it have on you, and how do you react to it?
6. [RQ-2d] Would it be helpful to know when you are sleepy or wide awake during your workday and if so, why/how?
7. [RQ-2d] Imagine you are very sleepy and working on a task. Would it be helpful to get a reminder to take a break, go to bed or switch tasks when you are fatigued? Are there any other feedback features you can imagine?

engagement

8. [RQ-2d] Do you usually notice when you are more or less engaged in your work? If so, what kind of impact does it have on you, and how do you react to it?
9. [RQ-2d] Would it be helpful to know when you are more or less engaged in your work and if so, why/how?
    - [If they don't come up with any] For example, would you reschedule tasks (if possible) when you are not engaged in your current task for a while or would you stick to your initial schedule?

general

10. [RQ-2d] Which of these three dimensions/aspects, i.e. mood, fatigue or engagement, is most important to you and why? (what's second…)
11. [RQ-2d] How do you deal with negative emotions, high sleepiness, and low engagement at your workplace?
    - What kind of habits, practices, or tools do you already use when you are fatigued or not engaged in your work?
    - Do you regularly take breaks?

Feedback on the assumptions/heuristics for **sensing** and awareness

"The application sensed your emotional and cognitive state using a few assumptions and heuristics that we've defined. We would like to verify some assumptions with your input and experience"

12. [RQ1] Do you think that your emotions and cognitive state are reflected in your facial expressions or body language?

13. [RQ1] What kind of behaviors do you notice in yourself when you are engaged in a task or interaction?

    [If they don't come up with any] In related studies, some participants moved closer to the screen when they were engaged. Do you notice similar behaviors?

14. [RQ1] What kind of behaviors do you notice in yourself when you are tired?

Questions regarding the **visualization** of the 3 states (in the pop-up)

The next few questions focus on the visualization of the mood, sleepiness and engagement as visualized in the "emoji face" **in the self-reporting pop-up.**

15. [RQ2-b]: Overall, how well does the visualization represent your mood, fatigue/sleepiness and engagement?
    a. Were there particular visual elements of the face that reflected your emotional and cognitive state well or not well? Do you remember any examples from your time in the study?
    b. [if not answered already] What do you think of visualizing mood in the mouth going up or down in the emoji?
    c. [if not answered already] What do you think about the closing eyes to represent fatigue?
    d. [if not answered already] What do you think about eye gaze as a representation for engagement?
    e. [if not answered already] Do other facial features come to mind visualizing these 3 dimensions?

Questions on the **glanceable display**

The next few questions are about the glanceable, always on-top display that you've seen on the last day of the study.

16. [RQ2-a] On your last day of the study, the "emoji face" was displayed at all times, visualizing your mood, sleepiness and engagement as it was sensed by the approach. What was your experience with the approach?
17. [RQ2-c] Did the approach help you to be more aware of your mood (happiness), fatigue and engagement?
18. [RQ2-c] Did the approach have any impact on you (positive, negative, motivating, demotivating)?

19. [RQ2-b] How accurate was the approach for you?
    - Does accuracy have a (big) impact on the approach for you?
    - [If the participant states that valence was too high] Do you think this is a bad thing or could this even be a valuable approach (e.g. positive framing).

20. [RQ-2d] What do you think of this **always on top** visualization?
    a. How frequent did you look at it and consider it? Every 30min? Or did you ignore it after a while?
    b. [if not answered so far] Was the always on top visualization sometimes distracting?
    c. [if not answered so far] What do you generally think of the idea that an emoji face is "mirroring" (and thus, visualizing) your mood, sleepiness and engagement?

21. [RQ2-a] Regarding the glanceable display, what do you think about how we notified you about extreme values (we changed the background color from green to red)? Did you notice the visual cue?
    - At the moment, we change the background if extreme states are estimated. Would you generally be more interested in easily detecting when your emotional or cognitive states change, or when they are extreme values? E.g. either "when you switch from awake to tired" versus "when you are very tired".
    - How might you want to be notified of such changes or extreme values? Is there a difference in how you want to be notified for each?

Wrap-up

22. Do you have additional comments regarding the EmotionalAwareness-Tool?
23. Do you have any other feedback or questions regarding the study?