Assessing stress at the workplace: an explorative study on measuring emotion using unobtrusive sensor techniques

A thesis by Loïs van de Water - 4141814
27,5 ECTS

07-07-2017

Universiteit Utrecht
First examiner: Stella Donker
Second examiner: Ignace Hooge

TNO
Anne-Marie Brouwer
Koen Hogenelst

Abstract

Stress in the workplace is a substantial problem for many employees and can, in the worst case, result in burnout. Therefore a method for measuring stress in an unobtrusive and office situation applicable manner is explored in the current study. Participants worked on their thesis in an office environment at TNO for 10 days during work hours. Heart rate, facial expression and computer usage were being gathered as parameters and related to a ground truth of subjective emotion experience sampling. Visualization of the dataset and results of correlation analysis leave promising outcomes and many possibilities for further analysis, such as classification analysis. It is important to take an individual approach in these analyses, since it concerns datasets with considerable individual differences in behavior. Privacy, sharing of data and employee needs are to bear in mind when further developing a work stress measuring system.

TNO innovation for life

Universiteit Utrecht

Introduction

Stress has become a substantial problem within the context of the workplace. In Western societies, the magnitude of sickness absence due to psychosocial health complaints represents a major concern (Duijts, Kant, Swaen, van den Brandt & Zeegers, 2007). When unhealthy stress levels are present for longer periods of time, some individuals may eventually develop serious mental health problems such as burn-out, anxiety or depression, resulting in long-term absence and disability (Koldijk, Kraai & Neerincx, 2016). Furthermore, stress in the workplace is not just directly, but also indirectly contributing to illness via maladaptive health behaviors such as smoking, poor eating habits or lack of sleep (Glanz & Schwartz, 2008).

Together, high job demands (such as workload, emotional demands and work-home interference) and limited resources (such as job control, feedback and social support) may, according to the job demands-resources model, result in work stress (Schaufeli, Bakker & Van Rhenen, 2009). Other models that strive to explain the concept of work stress are based on the balance between 1) effort and rewards (Siegrist, 2012), 2) effort and recovery (Meijman, Mulder, Drenth & Thierry, 1998), or 3) matching environment and personal characteristics (French, Rogers, Cobb, 1974). Stress in the workplace will in this study be defined as the change in one's physical or mental state in response to workplaces that pose an appraised challenge or threat to an employee (Colligan & Higgins, 2006).

The negative consequences of work-related stress on health and sustainable employment call for accurate stress monitoring to timely recognize unhealthy stress patterns and guide appropriate interventions. To assess work-related stress, several researchers use subjective measures composed of several items from common subjective stress measures (i.e., Michigan Diagnostic Survey or Stress Diagnostic Survey and Job Stress Index, Cavanaugh, Boswell, Roehling & Boudreau, 2000). However when stress in the workplace is to be assessed in a more continuous manner, repeated survey measurement asks a great amount of attention and motivation from the employee.

Modern sensor techniques, that are objective and far less intrusive, may also inform about a person's stress level. For instance, heart rate and facial expression correlate with subjectively experienced stress levels (Bakker, Holenderski, Kocielnik, Pechenizkiy & Sidorova, 2012). Moreover, objective sensor data are easily collected in a continuous way, yielding a vast amount of data points, which means natural variation can be captured more easily. Such unobtrusive monitoring techniques have previously been used in laboratory research in the TNO SWELL project (Smart Reasoning for Well-being at Home and at Work, Koldijk, Sapelli, Verberne, Neerincx & Kraaij, 2014; Koldijk, Kraaij & Neerincx, 2016). This research revealed that statistical models containing posture, facial expression, physiology and computer usage data could reach a performance of about 90% accuracy classifying whether a person was in specifically manipulated neutral or a stressful condition. However, little is known about the predictive accuracy of these measures in a real office setting, where stress can be associated with a number of different task situations, and where stress levels may vary in a more continuous, less extreme way.

In a pilot case study with 5 employees, Bakker, et al. (2012) used a prototype galvanic skin response (GSR) device to measure stress during workdays. They concluded that GSR data acquired from a wearable device yielded large methodological challenges and that stress detection using such data was difficult. Other monitoring techniques have been used by Mark, Iqbal, Czerwinski & Johns (2014). Mark et al. (2014) conducted an in-situ study in which engagement in workplace activities was assessed using software that logged computer usage. A wide range of digital activities was automatically being collected by the software. Furthermore, experience sampling was used to collect several self-report measures on subjective perception of engagement and challenge. From this study conclusions on boredom and focus during work hours were drawn. These studies show the possibilities for objectively and unobtrusively measuring stress in a workplace context. From these studies, it can be concluded that it is not only relevant but also reasonable to explore the use of monitoring objective stress measures, such as heart rate, facial expression and computer usage in a workplace context

Heart rate (in bpm) is often used as a physiological marker of emotions related to stress (Vrijkotte, Van Doornen & De Geus, 2000; Ditzen, Neumann, Bodenmann, von Dawans, Turner, Ehlert & Heinrichs, 2007; Gulati, Shaw, Thisted, Black, Merz & Arnsdorf, 2010). Stress namely affects several physiological processes in the body via the autonomic nervous system, among which heart rate (HR) and heart rate variability (HRV, Taelman, Vandeput, Spaepen & Huffel, 2009). Heart rate shows its applicability in a real life situation in that it can nowadays be measured by wearable health devices via photoplethysmography. Heart rate variability however cannot be determined. Besides cardiography, fingertip and earlobe photoplethysmography are stated to be reliable measures of HRV (Selvaraj, Jaryal, Santhosh, Deepak & Anand, 2008; Lu, Yang, Taylor & Stein, 2009), yet these are somewhat obtrusive measures to apply in an office setting. Schäfer and Vagedes (2013) however state that HRV estimation via photoplethysmography is only sufficiently accurate when subjects are at physical and mental rest. Another consideration is that although mobile HRV measuring is possible (Salahuddin, Cho, Jeong & Kim, 2007), commercially available alternatives require chest or finger placement, which would cause considerate practical challenges within an office setting.

Facial expression is a somewhat newer, but also widely used measure of emotion and has mainly been used in studies in which emotions were induced in participants (Koldijk, et al., 2016). Dinges, et al. (2005) manipulated workload and social feedback in order to create high and low stressor scenarios during long-duration spaceflights. Optical Computer Recognition algorithms were applied for detecting facial changes in mouth and eyebrow regions, and machine learning (Hidden Markov model) was used to identify the different stressor conditions. They concluded that their algorithm has potential to, in 75-80 % of the subjects, discriminate high- from low-stressor performance. Craig, D'Mello, Witherspoon & Graesser (2008) used expert ratings for investigating facial expression while students were working with an online tutoring system. Expert ratings were based on Ekman’s Facial Action Coding System. From association rule mining analysis was concluded that frustration and confusion were associated with specific facial actions. Another unobtrusive and applicable method for measuring facial expression, which is also based on the Ekman faces, is the software program FaceReader. According to a validation study by Lewinski, den Uyl and Butler (2014), FaceReader is a reliable indicator of basic emotions and shows promising results for facial expression components, called action units.

Computer usage has been used by Vizer, Zhou and Sears (2009) to detect stress after a mentally or physically stressful task. Keystroke and linguistic features of spontaneously generated text were assessed and results show that it is possible to classify cognitive and physical stress with accuracy rates compared to those obtained with affective computing methods. Furthermore in an explorative study by Khan, Brinkman and Hierons (2008) computer users’ mood was inferred from their computer behavior. Keypresses and mouse clicks were recorded from 13 programmers and 13 frequent computer users through a background software application. Mood was rated on a two-dimensional valance-arousal grid. They found significant correlations between keyboard/mouse use and valence for 31% of 26 participants and between keyboard/mouse use and arousal for 27%. Epp, Lipold and Mandryk (2011) used decision tree classification for identifying self-reported emotional states using keystrokes. They were able to recognize 15 emotional states based on keystroke dynamics. Rodrigues, Gonçalves, Carneiro, Novais & Fdez-Riverola (2013) also conclude that it is highly feasible to detect stress through mouse and keyboard activity. The software program uLog is able to log various parameters of computer usage. It can run on the background and is therefore highly unobtrusive. The program has already shown correlations between mouse behavior and HRV when participants were told to complete a search task on an internet page (van Drunen, van den Broek, Spink & Heffelaar, 2009).

The current study aims to combine the aforementioned objective measures - heart rate, facial expression and computer usage - in an attempt to explore the possibility of unobtrusively and continuously measuring emotion, which could eventually be related to stress, in a real office environment. For each of these measures a minimally obtrusive application is possible. Heart rate can nowadays be measured by wearable devices via photoplethysmography, while existing software can measure computer usage (uLog) and facial expression (FaceReader). As a ground truth, subjective emotion over the course of the day will also be measured, using experience sampling. This method has been successfully used in an alike study situation by Mark, et al. (2014) before. Csikszentmihalyi and Larson (2014) furthermore argue that experience sampling typically provides a plausible representation of reality. When collected close in time to experience, self-report using experience sampling is a reliable data source (Ericsson and Simon, 1980; Mischel, 1981). A great benefit of the method, besides the higher ecological validity and the possibility to analyze within subjects processes, is the reduction in memory bias it provides compared to retrospective methods (Napa Scollon, Kim-Prieto & Diener, 2009). Because we earlier saw that stress in a workplace context is defined in various ways, we expect this concept to be understood differently by individuals. Emotion, divided easily in valence and arousal, is therefore expected to be more suited for experience sampling. Eventually valence and arousal could be combined into the concept of stress.

Software programs FaceReader and uLog collect numerous variables, of which a selection will be analyzed. Concerning FaceReader, Craig, et al. (2008) found that frustration was associated with activity in the action units inner and outer brow raiser and dimpler, while students were working with an online tutoring system. They furthermore found that confusion was associated with action units brow lowerer, lid tightener and lip corner puller. Inner and outer brow raiser, dimpler, brow lowerer, lid tightener and lip corner puller are therefore expected to be informative when assessing emotions related to stress. Moreover, the basic emotions ‘angry’ and ‘scared’ are hypothesized to be informative for stress-like emotions, because anger (Holdeman, Good & Moore, 1976) as well as fear (Onaka & Yagi, 1990) are suggested to be related to emotional stress. Other basic emotions (neutral, happy, sad and disgusted) are also taken into analysis, as they are interesting for predicting other emotions than those related to stress and therefore can yield meaningful results in this explorative study. Concerning uLog, as demonstrated earlier, keystrokes and mouse clicks offer promising results as predictors of emotion and stress (Vizer, et al.2009; Khan, et al.,2008; Epp, et al., 2011; Rodrigues, et al., 2013). Khan, et al. (2008) in addition found correlations between valence and arousal and task switches and correction keys. These parameters are therefore also expected to be informative for emotion. Heart rate is hypothesized to be related to emotion in specifically the arousal dimension, because of the physical effects of stress-related emotions (Taelman, Vandeput, Spaepen & Huffel, 2009).

Literature points towards using an individual approach when assessing emotion with modern sensor techniques. Khan et al. (2008) found that participants showed different correlation patterns when computer behavior was related to their mood. They concluded that the findings could not be generalized in the used sample. Furthermore, Koldijk, et al. (2016) explicitly suggest further research to focus on building personalized stress models, because people differ in their work behavior. Because of the clearly implicated value of this kind of approach, in this study individual analysis will be used. This study is part of a larger TNO research project, in which also individual machine learning classification analysis will be executed. The more concrete aim of the current study is therefore to explore the dataset in a manner that can give direction to further classification analysis.

A second aim of this study is to investigate the subjective experiences with the study and the experience of privacy that participants have with the types of monitoring used in this study. Because many privacy sensitive data are collected from the participants, this presents a valuable objective. In particular, this question is important when considering the possibility of an application of the SWELL system in a real office environment. Because all participants agreed to the method used in this study when making the decision to participate, we do not expect them to have major concerns regarding their privacy in this study at the start. However, we are interested to know how their continuing observation is experienced, and how they feel at the end of the experiment regarding privacy. Another interesting concern is that of privacy when the current study situation is extended to a real office environment. A last interesting topic regarding the participants' point of view, is their overall experience with the study. Because this type of extensively monitoring people during their work is a relatively new method, all well substantiated notions are valuable in exploring future research and applications of these monitoring techniques.

Methods

Participants

In this study, 20 participants (of which 14 female) were included. One participant resigned from study participation without declaration, because of which a sample of 19 remained. The mean age was 25.84 (*SD* = 4.78). All participants were students at the time of the experiment and working on their thesis or other study projects. Participants were recruited by advertising within the University of Utrecht community, the TNO participant pool and within clients of certain thesis support companies. A monetary reward of €150 was available for each participant that completed the study. When people showed interest in participating in the study, an extensive amount of information about study procedures was sent to them. Declared reasons for deciding not to participate, after having shown initial interest, were the travel time, relatively low financial reward, long participation hours and practical issues like the request to work on a laptop that was not their own. Age and sex distributions are displayed in *Table 1.*

Table 1

Mean participant age (standard deviation) and number of male and female participants

Mean age (SD)	Males	Females
25.84 (4.78)	6	13

Procedure

Participants were asked to do their own computer work - concerning their thesis or other study projects - for 10 working days from 9.00 till 17.00 at TNO location Soesterberg. The participation days were planned as consecutively as possible and in consultation with the experimenter. The reason that each participant was asked for ten days of participation which were as consecutive as possible, is that a large amount of data - ideally of consecutive days - per participant was required for other analysis goals, regarding the larger research project that this study is part of.

For each participant, an own quiet work area was available which consisted of a desk, an adaptable chair, a Dell Windows 7 laptop with webcam and laptop stand, a mouse, a keyboard and a USB hub. Facilities such as a restroom and a drinks dispenser were freely available. Each morning participants logged in on the study software application with their participant ID and personal password. Then the heart rate wearable was placed on the participant's wrist and connection with the personal study computer and the local data server was checked. While participants were working, several unobtrusive measurements were gathered. Furthermore, they were asked to rate their emotion every 15 minutes. For details, see *Measurements*. A visual impression of the test set-up is displayed in *Figure 1.*

Short individual breaks were allowed. In addition each day at twelve o'clock all participants were guided to the canteen for a lunch break. When participants had to leave the location for an appointment, they asked permission from the experimenter. The experimenter then registered their absence and guided them to the exit. If participants were ill, a new participation date was planned.

On their first participation day, participants received a short instruction about all measurements. An informed consent form, including an attached contact form, was signed. Afterwards, two questionnaires were filled in. At the end of each participation day, a short questionnaire was taken on global activities, emotion, stress and mental effort over the whole day. These questionnaires are on purpose of gathering data for the larger research project that this study is part of. The results of the questionnaires have not been analyzed in the current study.

On their last participation day, a one-to-one interview with the participant was conducted in order to measure privacy perception and subjective study experience.

Figure 1. A visual impression of the test set-up.

Measurements

Measurements used in this study are subjective emotion measurement, heart rate - as assessed by a wearable heart rate measuring device, computer activity- as assessed by uLog - and facial expressions - as assessed by FaceReader 7.0. Also one-to-one interviews in order to measure privacy perception and subjective study experience were conducted with participants. All measurements are explained in more detail below.

Subjective emotion measurements. For measuring emotion, the Affect Grid (Russel, Weiss & Mendelsohn, 1989) was used. This grid divides emotion along two separate scales - valence and arousal. The grid showed strong evidence of convergent validity with other measures of pleasure and arousal, of discriminant validity between the pleasure and the arousal dimension and of reliable aggregated judgments (Russel, Weiss & Mendelsohn, 1989). Surrounding state labels were translated into Dutch. The used Affect Grid is shown in *Figure 2*. Each 15 minutes a pop-up screen with this Affect Grid appeared on the participant's computer screen. Participants were instructed to indicate their current emotion by clicking the appropriate location in the grid. They could also click the pop-up away, but participants were asked to do this only when they really did not feel to answer. We suggested to first determine the score on the horizontal *valence* axis and then the score on the vertical *arousal* axis. Scores could be checked by considering the surrounding emotion labels as current emotional state.

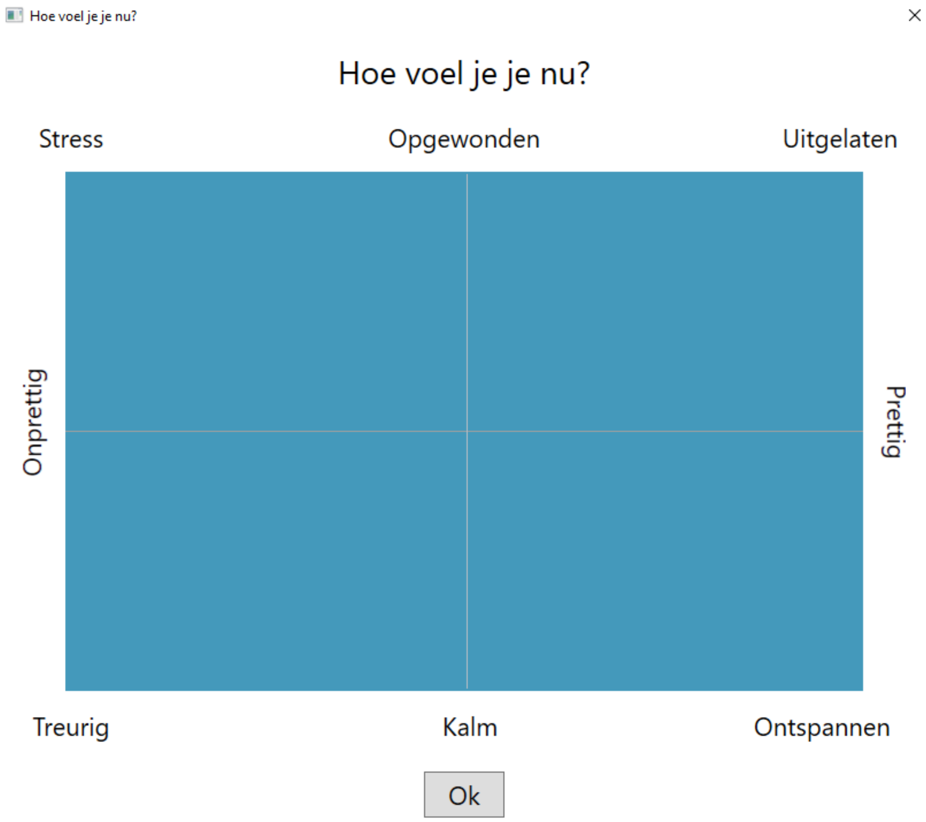


Figure 2. The Affect Grid that was used in this study. The surrounding labels were translated from the Affect Grid of Russel, et al. (1989).

Heart rate. For measuring participants heart rate, a MIO Fuse heart rate wrist wearable was used. The choice for a wrist wearable came from the aim of using unobtrusive measures that can be translated into a real office environment. Participants were instructed to wear the wearable during all working days. Results were communicated with the local data server via TNO software on the participant's computer. From the wearable, at every heartbeat a BPM value was sent to the server together with the ID number of the wearable.

Computer activity. For measuring computer activity, Noldus uLog keylogger (research Edition version 3.3) was used. This program logged various parameters of keyboard usage and mouse usage, of which the number of keypresses per minute, the number of error keypresses per minute, the error-ratio and the number of application switches per minute were assessed in this study. Typed strings of text and names or content of documents, emails and websites however were not recorded, to protect the participant's privacy.

Facial expression. For measuring the participant's facial expression, the study software application was programmed to take a webcam snapshot every minute. Afterwards we used Noldus FaceReader 7.0 to extract different basic emotions (such as happy or angry) and expression components, so-called action units. All FaceReader analysis is based on the Ekman faces (Ekman & Friesen, 1978; Ekman, Friesen & Hager, 2002). In this study, the complete emotional expressions neutral, happy, sad, angry, scared and disgusted as well as the action units inner and outer brow raiser, dimpler, brow lowerer, lid tightener and lip corner puller, were examined.

Interviews. For investigating participants' overall experience and privacy experience with the study, the second aim of this study, semi-structured one-to-one interviews were conducted. Open (qualitative) as well as - 1 to 10 scaled - Likert scale questions (quantitative) were used. An example of a qualitative question is: "How did you experience this study related to your privacy?". An example of a quantitative question is: "Do you think your privacy is violated with participation in this study on a scale from 1 to 10, when 1 = not at all and 10 = very much?". For all interview questions in the original language (Dutch), see *Appendix 1*. With all quantitative questions, an explanation of the interviewee's statement was gathered. When answers, on both qualitative or quantitative questions, gave cause to asking for clarification, further questions were asked. The audio was recorded and textual notes were made by the experimenter.

Design

In this study, a form of exploratory individual analysis was used. No interventions or conditions were applied. The subjective variables are reported arousal and valence and the objective variables are heart rate and parameters of computer activity and facial expression. Interview topics of interest are subjective experiences with the study and privacy perception of gathered data and possible future applications of these data.

Analysis

After pre-analysis of the dataset, an exploratory analysis, comprised of data visualization and correlation analyses, was performed when assessing monitoring data. Specifically, within this analysis an individual approach was used because this is expected to be more suitable and valuable than a group approach. Furthermore, some suggestions for correlation enhancement were analyzed. Interview data were analyzed separately. All analysis components are explained in more detail below.

Pre-analysis. Because of the magnitude of the data, pre-analysis was necessary. All parameters were averaged per subjective emotion data point. The averaging interval was set on 15 minutes previous to emotion judgment, because of the sampling rate - every 15 minutes - for this measure.

Dataset. Based on the number of participation days, 4 participants were excluded from further analysis. Second, visualization of the data was realized to give direction to further analysis and to ensure to not miss out on important confounds. Microsoft Excel was used for creating accessible graphs of all parameters of interest. To assess the statistic relationships between the different objective variables - heart rate, computer activity and facial expression - and subjective emotion parameters - valence and arousal - multiple bivariate correlation analyses were carried out in IBM SPSS.

Suggestions for correlation enhancement. First, a method of enhancing heart rate data was carried out. Indicated by visualization, data points for which heart rate was above a specific value were excluded from analysis, in an attempt to filter out heart rate peaks that were mainly caused by physical movement. The applied cut-off was based on visual inspection of heart rate data and differed per participant. Correlation analyses were then performed on the filtered dataset.

Second, a method of filtering out data points at which the participant was not present at his workstation, was applied. Heart rate and FaceReader data could be noisy when a considerable amount of data in the 15 minute time window was missing. Especially for computer usage data this could lead to bias, because for these data there was no label that indicated missing data. Therefore computer usage parameters could be artificially low.

In order to filter out these data points, an exclusion criterion based on the occurrence of FaceReader not finding a face to analyze was applied. In these cases, FaceReader gives 'Find_Failed' as output. The applied cut-off was a Find_Failed output at 10 of 15 original data points per data point used for correlation analysis, because 10 out of 15 minutes was judged to be a fair amount of time to conclude the participant was generally not present. Correlation analyses were then performed on the filtered dataset.

Third, data were split following the used software application. Different kinds of working behavior might be dependent on the sort of application that was used. Participant 1 was used as an example because this participant had a high mean usage percentage of one software application, that is Google Chrome (*M* = 46,56%). A data split was applied using a criterion of 85% Chrome usage averaged over previous 15 minutes, in order to have approximately even data samples. Correlation analyses were performed for both sets separately.

Interviews. For analysis of quantitative data from semi-structured interviews, mean scores and standard deviations were computed. For analysis of qualitative data from semi-structured interviews, evaluation in the form of structured summarization of answers and remarks per question, over all participants was used.

Results

Dataset

In this section, an overview of the monitoring data set is displayed. Analysis consisted of personal data visualization and correlation analyses. Furthermore, some suggestions of data enhancement are done at the end of this section. In the section *Interviews*, data from one-to-one semi-structured interviews will be described.

Table 2 presents general descriptives of the gathered data set. When participants had not succeeded to participate 10 participation days, they were excluded from further analysis. This was the case for participant 6, 10, 14 and 16. Declared reasons for shortened participation, were different kinds of personal issues, such as sickness. Reasons for elongated participation were serious absentness from workstation due to appointments (participant 1), or the will to participate more days because of the pleasant and quiet working area (participant 18).

Table 2
Descriptive quantities on emotion ratings, heart rate entries, total participation days and participation days at which heart rate data was successfully gathered

Participant number	Total number of emotion ratings	Number of heart rate entries	Total number of participation days	Number of participation days with heart rate data
1	238	84	11	6
2	288	287	10	10
3	292	288	10	10
4	290	278	10	10
5	298	283	10	10
6	128	27	4	1
7	261	254	10	10
8	255	198	10	9
9	268	261	10	10
10	0	0	0	0
11	249	240	10	10
12	290	273	10	10
13	277	265	10	10
14	30	12	1	1
15	290	228	10	9
16	150	146	5	5
17	234	210	10	8
18	463	407	16	14
19	297	290	10	10
20	289	251	10	9

Note: Heart rate data were missing due to unforeseen technical problems in the first weeks of testing.

Visualization

Subjective emotion. The spread of subjective emotion measurements is displayed in *Figure 3*. The amount of spread was different for each participant; some participants did not report to vary much in their emotional state (such as participant 5), while others did (such as participant 9). But in each participant some amount of spread was found. Another remarkable aspect of the figures is the vertical 'gap' that is found on the valence axis with some participants (such as participant 2 and 18).

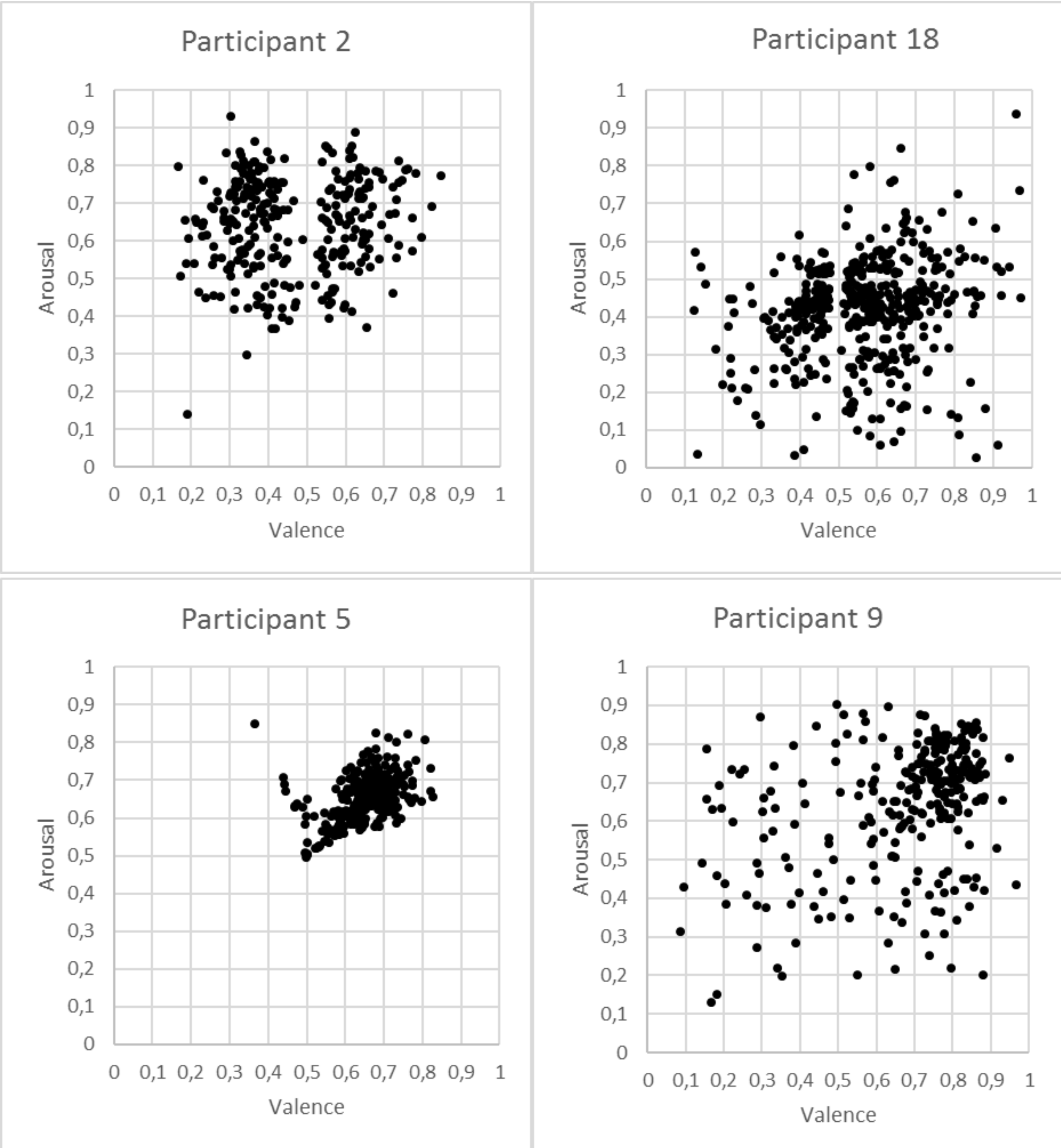


Figure 3. Emotion ratings over all participation days, displayed for four participants with a typical spread. Reported arousal is displayed on the y-axis and reported valence is displayed on the x-axis. Visual representation of the emotions ratings, resembles the Affect Grid which was used as method for gathering the emotion ratings. For emotion rating figures of all 16 participants, see *Appendix 2*.

Figure 4 presents the spread of subjective emotion ratings over the total participation period. Some participants did not report to have varied much in their emotion between the participation days and on different parts of the day (such as participant 4 and 8), while others did report to be in very different states at different days and parts of the day (such as participant 15 and 17).

Figure 4. Arousal and valence ratings over total participation period, for four participants with a typical pattern. Reported arousal is displayed in yellow. Reported valence is displayed in blue. Participation day number divided by ten is displayed in gray. Measurement number is displayed on the x-axis, subjective score on the y-axis. For the arousal and valence ratings over total participation period for all 16 participants, see *Appendix 3*.

Heart rate. In *Figure 5*, the typical heart rate pattern that was found in participants is displayed. Remarkable is the peak in heart rate around lunchtime.

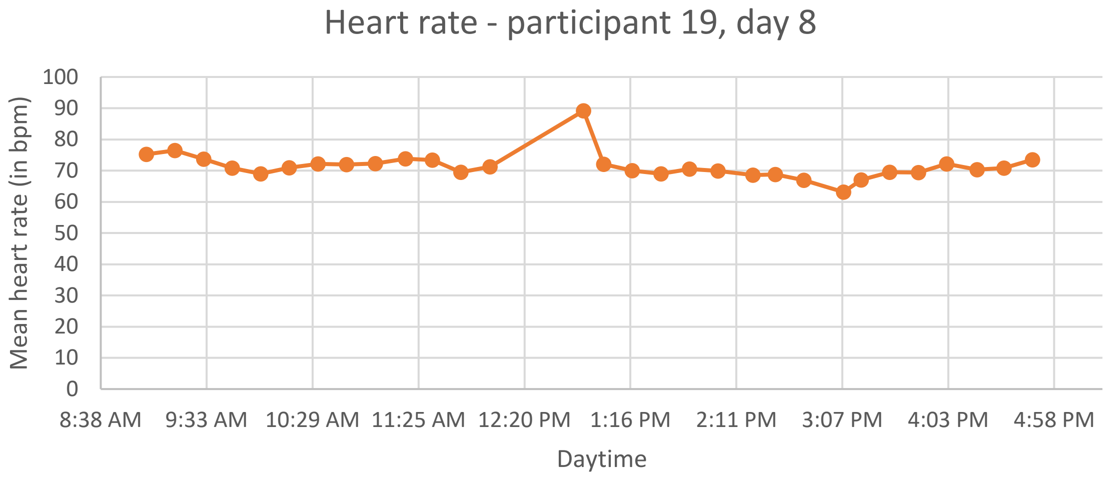


Figure 5. The typical heart rate pattern over daytime (participant 19 on participation day 8). Mean heart rate over previous 15 minutes (in beats per minute) is displayed on the y-axis. Daytime is displayed on the x-axis. For heart rate patterns on all participation days of participant 19, see *Appendix 4*.

Facial expression & computer usage. Visualizations on facial expression and computer usage data are less important for further analyses and are therefore displayed in *Appendix 5*, *6* and *7*.

Correlation analysis

In *Figure 6*, the number of participants for whom we found significant correlation coefficients is displayed. Because the assumption of normality was not met, based on visual inspection of histograms, Spearman correlation coefficients were displayed. Computer usage yielded most significant correlations, heart rate and facial expression yielded less. Found significant correlation coefficients were often not all in the same direction, for each parameter. Therefore the directions of the coefficients are distinctively displayed.

Figure 6. Number of participants (out of 17) for whom Spearman correlation coefficients were significant, separately for each analyzed correlation between monitoring parameters and arousal and valence. The number of significant negative correlation coefficients is displayed in orange, with a negative number, the number of significant positive correlation coefficients is displayed in blue, with a positive number. Parameters are displayed on the y-axis and frequency is displayed on the x-axis.

Suggestions for correlation enhancement

The number of correlations, i.e. the strength of the association between rated emotion and the examined variables, may be further increased in several ways. Some of these are explored for a limited amount of data.

First, a method of enhancing heart rate data based on filtering out data points influenced by physical movement, was carried out. Two example cases are displayed in *Table 3* and *Table 4*. In both cases, a cut-off of 80 bpm was used based on visual inspection of heart rate diagrams of each participation day (see *Figure 5* and *Appendix 4*). For participant 19 (*Table 3*) correlations were not enhanced when the dataset was filtered on high heart rates, while for participant 8 (*Table 4)* they were.

Table 3

Spearman correlation coefficients for correlations between heart rate (HR) and reported valence and arousal of participant 19 in filtered and unfiltered datasets

	Unfiltered HR	*Filtered* HR
Valence	*.124*	*.088*
Arousal	*.423*	*.399*

Table 4

Spearman correlation coefficients for correlations between heart rate (HR) and reported valence and arousal of participant 8 in filtered and unfiltered datasets

	Unfiltered HR	*Filtered* HR
Valence	*.033*	*.060*
Arousal	*.025*	*.039*

Furthermore, a method of filtering out data points at which the participant was not present on his workstation, was applied. The most interesting parameters for this correction are computer measures, because of the sensitiveness for absentness. Therefore results of computer measures analyses are shown in *Table 5*. With this suggestion, no clear differences in correlations were found.

Table 5

Sample size and computer measures correlation analysis results when a Find_Failed criterion was applied, compared to when not. Cursive numbers are Spearman correlation coefficients, non-cursive numbers are frequencies

Dataset	Find_Failed<10	Complete
Number of valid emotion rating data points	442	463
Total number of significant Spearman correlation coefficients	11	11
Keystrokes-Arousal	*.046*	*.022*
Keystrokes-Valence	*-.187**	*-.190**
Errorkeys-Arousal	*.065*	*.047*
Errorkeys-Valence	*-.196**	*-.194**
Mouseclicks-Arousal	*.064*	*.037*

Mouseclicks-Valence	*-.181**	*-.188**
Error-ratio-Arousal	*.055*	*.039*
Error-ratio-Valence	*-.171**	*-.171**
Task switches-Arousal	*.041*	*.052*
Task switches-Valence	*-.215**	*-.221**

*Significant Spearman's ρ when α = .05

For the last suggestion, data were split following the used software application. Correlation analyses were performed for both sets separately. Results from correlation analysis are displayed in *Table 6*. When Google Chrome was used for less than 85% of the time, mainly facial expression correlations seemed to remain. When Google Chrome was used for 85% of the time or more, only computer usage correlations seemed to remain.

Table 6

Sample size and correlation analysis results for both samples when a criterion of 85% was used on chrome usage averaged over previous 15 minutes and when no data split was applied. Cursive numbers are Spearman correlation coefficients, of which only significant correlation coefficients were displayed. Non-cursive numbers are frequencies

Dataset	Chrome usage >=85%	Chrome usage <85%	Complete
Number of valid emotion rating data points	104	107	238
Total number of significant Spearman correlation coefficients	10	7	18
HR-Valence			*.249*
Neutral-Arousal		*-.200*	*-.212*
Happy-Arousal		*.400*	*.246*
Sad-Arousal		*-.229*	*-.169*
Sad-Valence		*-.200*	*-.164*
Angry-Arousal			*-.220*
Angry-Valence		*-.218*	*-.220*
Lip Corner Puller-Arousal		*.218*	
Keystrokes-Arousal	*-.434*		*-.407*
Keystrokes-Valence	*-.326*		*-.407*
Error keys-Arousal	*-.402*		*-.398*
Error keys-Valence	*-.333*		*-.423*
Mouse clicks-Arousal	*-.260*		*-.423*
Mouse clicks-Valence	*-.312*		*-.437*
Error-ratio-Arousal	*-.364*		*-.376*
Error-ratio-Valence	*-.340*		*-.455*
Task switches-Arousal	*-.305*		*-.432*
Task switches-Valence	*-.420*	*-.236*	*-.500*

Interviews

Study experiences. Participants' general experiences with the study were positive. Participants liked the study with a mean score of 7,32 (*SD* = 1,07) on a 1 to 10 scale. When asked to rate the possibility that they would participate in a comparable study again, the mean score was 7,47 (*SD* = 1,18). They stated to have had a pleasant, quiet work area and that they were able to work in a concentrated and productive manner. Mentioned disadvantages were the long working days, non-flexible working hours and the lack of freedom assigned to participants. Most participants report to not have been bothered much by the measures.

The subjective emotion pop-up screen, in specific, was not experienced as interfering or bothering, except in a few cases when participants were very concentrated and wishing to finish a certain task. They seldom closed the pop-up without giving an answer. Participants often remarked that it was challenging to get wind of their true emotional state at certain times. They often felt to have clicked mostly in the same area in these and other cases. The affect grid itself was judged as pleasant, easy and quick to use.

The other - meant to be unobtrusive - measures were indeed experienced as convenient and not bothersome. Most participants reported to not have adjusted their behavior according to the measures, because of decreasing awareness of these measures. Aspects that sometimes caused awareness of the measures were the occasional buzzing of the wearable heart rate device and the flickering of the webcam when it turned on. The heart rate device was in some cases named as the less pleasant measure, because of the buzzing when other consumer functions were active. When in certain cases awareness of the monitoring was raised, the feeling of 'being watched' could sometimes arise as reported by some cases.

When participants were asked about a system that could be developed based on the gathered data, for instance a mobile or desktop application that can provide insight in personal stress levels and generate advice based on this, a mean score of 5,00 (*SD* = 2,42) on a 1 to 10 scale was given for wanting to use this kind of system themselves in a real work situation. The main potential advantages that were mentioned, were the value of getting insight in their own stress and working pattern which could be helpful for structuring their work activities and the possibility of generating personal health advices and making management decisions. While some participants reported finding it useful to have this insight, others did not see the use of a system that creates insight in something they 'are already aware of'. Another mentioned disadvantage was the fear of experiencing stress caused by using the system itself. Furthermore, not everyone believed that it would be possible for such a system to monitor stress accurately enough in order to be useful. When later asked about the trust that they had in the measures predicting their current level of stress, they gave a mean score of 6,35 (*SD* = 1,36) on a 1 to 10 scale. Another reported disadvantage was the privacy issue and the feeling to be watched.

Privacy. Addressing the participants privacy in the current study, the majority of the participants reported being fine with the manner of gathering and analyzing data that was used in the study. When asked to what extent their privacy was harmed in their opinion, a mean score of 2,59 (*SD* = 1,70) on a 1 to 10 scale was given. Participants had trust in the discretion of the researchers, although some did point out that they were aware that the researchers did in principle have access to all personal data. Others felt that the recorded data was not really personal to the extent that they were worried about it being shared or accessed. The overall conclusion was that participants reported to not have strong concerns about privacy issues, since they had beforehand agreed to the procedure of the study when deciding to participate in the study.

When their willingness was questioned to, in a real office setting, share the same kind of data with their manager, 8 of 17 interviewees reported willingness to share at least some of the data on a personal level with their manager. Proponents motivate their answer mostly with the use they saw for the manager to make changes in the employees working situation. Opponents mostly are afraid that wrong conclusions will be taken by the manager or that they would be judged or compared with other colleagues. Many opponents however stated that they would share their data with their manager if the data were anonymized or aggregated between more colleagues.

When the question was if they would be willing to share the data with their colleagues, only 5 of 17 said to be willing to share some data on an individual level. Their main motivation was that it would be useful for colleagues to compare their data. Others thought it would lead to competition, lack of trust, a bad atmosphere or wrong conclusions. Some subjects also did not see the use in this exchange of data. As with the previous case, many participants stated to be willing to share the data on aggregated or anonymous level.

When presented with the final case of sharing data with the company doctor, 15 of 17 reported being willing to share some data with their company doctor. They would rely on the professional confidentiality of the doctor and the justified interest of a company doctor. However many participants mentioned that they would only share data if they have a problem or if it has a clear purpose.

In *Figure 7* and *Figure 8*, quantitative data from the interviews are displayed.

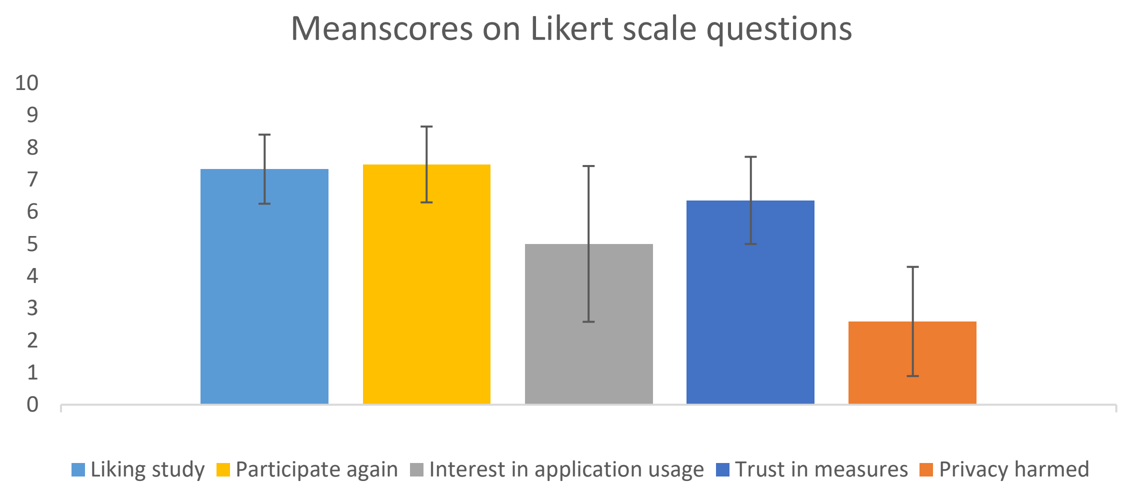


Figure 7. Mean scores on quantitative Likert scale (1 - very little to 10 - very much) questions. The subject of the question is displayed below. Error bars show standard deviations from the mean.

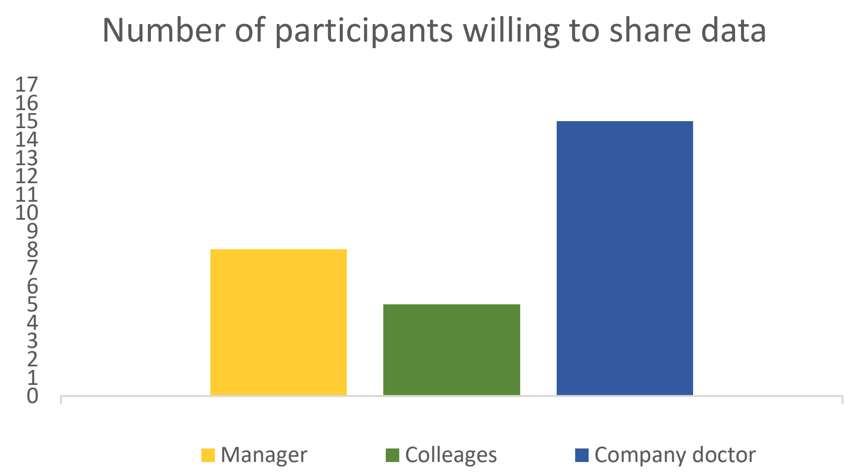


Figure 8. Number of interviewed participants that would be willing to share at least some of the alike data on a personal level with their manager, their colleagues or their company doctor in a real office setting, from a total of 17 interviewed participants

Discussion

The first aim of this study was to explore the possibility of unobtrusively and continuously measuring emotions in real life working conditions, which could eventually be related to stress, by combining different objective monitoring measures. The second aim was to investigate subjective experiences with overall study procedures, especially those concerning privacy.

Overall we can conclude that the design worked out well. Generally participants were motivated to show up and not resign from participation during the study. Such high motivation even occurred that participants voluntarily joined the study for a considerately longer period than 10 days. Furthermore, this study's specific monitoring methods were experienced as easy to use and not bothersome or raising awareness. This suggests an unobtrusive measuring method, as was intended.

In data visualization we looked extensively at ground truth subjective emotion measurements. From these visualizations can be concluded that spread in subjective emotion data is existent, though differing per participant. According to subjective emotion data, participant's emotion varied during the day, but also between different days. This is a valuable conclusion when considering further correlation analysis, because this spread is important in revealing relations between the subjective and objective measures. Furthermore, an interesting vertical 'gap' on the valence axis is found in some cases. This gap indicates that some participants were not likely to report feeling completely 'neutral', but rather slightly happy or slightly unhappy. Another conclusion to be taken from data visualization is that in the heart rate data over the day, clear moments of increased heart rate are visible at certain times of the day. For instance after lunch time, when participants had got physically active by walking from the canteen to their workstation, an increase in heart rate is often visible. This indicates that the visible peaks are likely caused by physical movement.

In further analysis, several correlations were found, although varying in direction. We can conclude from this that the dataset does yield some relations between subjective emotion and objective monitoring parameters. Like in the comparable study by Khan, et al (2008) the direction and the amount of these relations depend on the individual. Regarding this and further classification analysis, this result suggests an individual approach is indeed necessary.

In this study heart rate yields surprising results when related to valence and arousal. Especially because the results are near to evenly distributed between positive and negative relations for both valence and arousal, although in the case of arousal a clear direction (more arousal, higher heart rate) is commonly expected. However, previous TNO lab research shows that mental arousal can be positively related to heart rate (e.g. in social stress, Brouwer & Hogervorst, 2014) or negatively (e.g. when reading highly arousing sections in a novel, Brouwer et al., 2015). An explanation for this could be that there are more than one arousal systems that affect behavior in different ways (Graham & Jackson, 1970). According to the theory of Sokolov (1963), the defense reflex and the orienting reflex are distinguished as different arousal systems. In the defense system, response is energized and receptive and consolidating processes are inhibited, while in the orienting system, these processes are facilitated. The defense system is thus characterized as the system that stimulates reaction to a stimulus (breaking away from or limiting the activity of the stimulus), and the orienting system as the system that enhances sensitivity, sensory processing, memory and learning (Turpin, 1968; Graham & Jackson, 1970). With the defense system heart rate accelerations were found, while with the orienting system decelerations in heart rate were found (Graham & Clifton, 1966). When considering the orienting system, it could thus be possible to reach low heart rates while arousal is actually high: for instance when someone is in a very concentrated state. This suggests that in finding both negative and positive relations between heart rate and arousal, these separate arousal systems could have been involved.

Another important aspect to consider is the reactivity of heart rate with physical activity. When high heart rates that were probably caused mainly by physical activity of walking from or to the workstation are filtered out, this yields contradictive results in participants. As one shows slight correlation improvement, the other does not. It might be possible that some participants judged their emotional state to be high in arousal when recovering from physical activity, while others did not. Anyhow these results show that physical activity might be a factor to take into account in further analysis.

When looking at the amount of participants for whom significant correlations were found for each parameter, computer behavior seems to be considerately informative for predicting emotion. A possible cause for the variation in correlation direction can be differences in work behavior. One for instance might get a good mood when increasing typing behavior because of the feeling of making progress, while the other might feel time pressure, frustration or stress when he or she is accelerating in typing behavior.

Computer behavior and emotion correlations seem not to clearly improve when absence at the workstation, according to FaceReader, is taken into account. When a participant is absent from his or her computer, this could lower the computer behavior measures in the 15 minute time window. However, results show improvements as well as aggravations in correlations between computer measures and valence and arousal. It might however be possible to create another type of indicator of absence by looking at sustained low key presses or mouse clicks. Also a more strict Find_Failed criterion might yield different results.

When considering the computer usage of students working on their thesis, it could be that the analyzed parameters are differing per subtask that is performed. For instance when searching for or reading through literature or analyzing quantitative data, mouse clicking might be prominent. While when typing a method section, or doing qualitative analysis, keyboard presses might be prominent. These differences in computer behavior will be hard to detect in the current dataset, but could be assessed by taking the parameter application usage into further analysis.

When the dataset of one participant was split on Google Chrome usage, primarily FaceReader correlations remained when on average less than 85% of the time Chrome was used. However when on average 85% or more of the time Chrome was used, mainly computer measure correlations remained. This means that software application usage could indeed be an interesting parameter to bear in mind for further analysis, because the findings suggest that computer behavior parameters are more informative for predicting emotion of this participant when spending much time on the internet and facial expression is more informative when not much time on the internet was spent.

Considering facial expression and emotion, somewhat less clear results were found in correlation analysis. FaceReader's expression components, so-called action units, yield less promising outcomes when related to valence and arousal than when these components are combined into complete expressions, so-called basic emotions. A possible explanation for the lack of found correlations, could be that FaceReader is not sensitive enough to pick up the subtle expression differences that are present on the face of someone that is working on a thesis, compared to a design in which certain emotions are provoked. Because the basic emotions are composed of several action units, the latter might even have less potential picking up emotional expressions in real working environments. Supporting this explanation is a study by Koldijk, Bernard, Ruppert, Kohlhammer, Neerincx and Kraaij (2015) in which FaceReader data from a knowledge work dataset are visualized. The data were gathered on 25 participants performing knowledge work for 3 hours in different conditions. In the neutral condition participants were instructed to work as they usually do and in the two stressor conditions either email interruptions or time pressure were added. When FaceReader action units were visualized, no great activity was found on the several action units, although differing per individual. In the current study, no manipulations were made, which suggests even less action unit activity to be found.

When looking at the directions of the correlations, we can conclude that FaceReader's labels do not quite seem to fit the current dataset. For example also positive correlations were found when relating anger to valence, when specifically negative correlations would be expected. Yet also negative correlations were found when relating anger to arousal, in which case positive correlations would be expected. Again this could have to do with the study setting; facial expressions might be not really clear when performing computer work in an office setting.

When looking at the second aim of the current study, we can conclude that the used measures were all judged as accessible and easy to use. Participants seemed to have least pleasant experiences with the wearable heart rate device. The device had certain functions meant for fitness use, such as a buzz when a high or low heart rate was suddenly reached or when a stopwatch started. These functions made the device somewhat less suitable for an office application. Furthermore some technical problems arose with the communication between the heart rate devices and the logging software. Consequences were missing data for several hours or sometimes days. Suggested for future research is therefore is to develop software that is capable of logging many devices at the same time, and of showing battery duration information.

Regarding privacy, we can conclude that the study did not cause considerate issues. However, it must be noted that all participants joined the study after being informed about the study procedures and agreeing with them by signing informed consent, as participants themselves also bring forward. Only when future applications and willingness to share alike data with other people in a company were addressed, issues arose in some participants. When in the future a software application would be developed that helps employees creating insight in work stress patterns, the admission to the data is something to bear in mind. Also the needs of employees should be taken into consideration, because the participants' interests in such an application were not very convincing.

Although participants reported decreasing awareness for the measures, a current issue in this study might be the fact that participants are theoretically aware of being monitored. Therefore their behavior could have been different from their behavior in a real office. Furthermore, the created office situation was slightly different from a real office situation in a few aspects. For instance, in a company office probably more tight and short-term deadlines and more social pressure from colleagues or managers would be expected.

Lastly, the frequent demand of judging current emotional state, could have influenced participants' emotional state. The fact that participants reported to not have been bothered much by the pop-ups, however does not support this issue. A considerable issue that does match participants' considerations, is finding it difficult to get wind of their true emotional states. This implies that measured emotional states might be differing from real emotional states. Yet with the present state of monitoring emotion held in mind, a subjective method was a necessary choice for adding a ground truth measurement to the current study. Also with future applications, some form of subjective measurements will probably be needed for model calibration. When talking about an application for employees - based on machine learning models - that could give insight in personal stress levels, ideally only in the calibration phase subjective measurements will be needed. Further in the process, probably only occasionally the measure should return in order to keep the model accurate. However future research is to reveal more about the possibilities with these kinds of models.

As the current study is only a first attempt at applying modern sensor techniques in office situations, in the future also other measures are expected to develop non-obtrusive alternatives that can monitor working behavior. For instance modern eye tracking devices suggest promising applications for measuring pupil sizes and blink frequencies in order to assess cognitive load in employees.

In summary, this study yields promising results in the field of monitoring real life behavior. We show that some of the essential requirements to be able to proceed in this field are met: emotions under real life office working circumstances vary enough, and can be reported well enough, in order to uncover relations with several unobtrusively measured variables on the level of an individual person. Suggested is furthermore that differences in heart rate, facial expression and computer usage, although subtle, could be detected and in some cases related to emotion. Individual analysis has shown to be a valuable and furthermore necessary approach to this kind of data. The currently gathered dataset leaves many possibilities of further analysis, of which individual machine learning classification analysis will certainly be of considerate value.

References

Bakker, J., Holenderski, L., Kocielnik, R., Pechenizkiy, M., & Sidorova, N. (2012, January). Stess@ work: From measuring stress to its understanding, prediction and handling with personalized coaching. In *Proceedings of the 2nd ACM SIGHIT International health informatics symposium* (pp. 673-678). ACM.

Brouwer, A. M., & Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in neuroscience*, *8*.

Brouwer A-M, Hogervorst, MA, Reuderink, B, van der Werf, Y, van Erp, JBF. (2015) Physiological signals distinguish between reading emotional and non-emotional sections in a novel. *Brain-Computer Interfaces*, 1-14.

Colligan, T. W., & Higgins, E. M. (2006). Workplace stress: Etiology and consequences. *Journal of workplace behavioral health*, *21*(2), 89-97.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive–affective states during learning. *Cognition and Emotion*, *22*(5), 777-788.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35-54). Springer Netherlands.

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z. & Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, *76*(6), B172-B182.

Ditzen, B., Neumann, I. D., Bodenmann, G., von Dawans, B., Turner, R. A., Ehlert, U., & Heinrichs, M. (2007). Effects of different kinds of couple interaction on cortisol and heart rate responses to stress in women. P*sychoneuroendocrinology*, *32*(5), 565-574.

van Drunen, A., van den Broek, E. L., Spink, A. J., & Heffelaar, T. (2009). Exploring workload and attention measurements with uLog mouse data. *Behavior research methods*, *41*(3), 868-875.

Ekman, P., Friesen, W. V., & Hager, J. (1978). The Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto. *CA: Consulting Psychologists Press, Inc. Ekman, P. Levenson. RW, & Friesen WV (1983). Autonomic nervous system activity distinguishes among emotions. Science*, *221*, 1208-12.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial action coding system: The manual on CD-ROM. Instructor’s Guide. *Network Information Research Co, Salt Lake City*.

Epp, C., Lippold, M., & Mandryk, R. L. (2011, May). Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 715-724). ACM.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87,* 216–251.

French, J. R., Rodgers, W., & Cobb, S. (1974). Adjustment as person-environment fit. *Coping and adaptation*, 316-333.

Glanz, K., & Schwartz, M. D. (2008). Stress, coping, and health behavior. *Health behavior and health education: Theory, research, and practice*, *4*, 211-236.

Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological bulletin*, *65*(5), 305.

Graham, F. K., & Jackson, J. C. (1970). Arousal systems and infant heart rate responses. *Advances in child development and behavior*, *5*, 59-117.

Holdeman, L. V., Good, I. J., & Moore, W. E. (1976). Human fecal flora: variation in bacterial composition within individuals and a possible effect of emotional stress. *Applied and Environmental Microbiology*, *31*(3), 359-375.

Koldijk, S., Kraaij, W., & Neerincx, M. A. (2016). Deriving requirements for pervasive well-being technology from work stress and intervention theory: framework and case study. *JMIR mHealth and uHealth*, *4*(3).

Koldijk, S., Neerincx, M. A., & Kraaij, W. (2016). Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on Affective Computing*.

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M. A., & Kraaij, W. (2014, November). The SWELL knowledge work dataset for stress and user modeling research. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 291-298). ACM.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227.

Mark, G., Iqbal, S. T., Czerwinski, M., & Johns, P. (2014, April). Bored mondays and focused afternoons: the rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3025-3034). ACM.

Meijman, T. F., Mulder, G., Drenth, P. & Thierry, H. (1998). Psychological aspects of workload. *Handbook of Work and Organizational Psychology*, *2*.

Mischel, W. (1981). A cognitive-social learning approach to assessment. In *T. Merluzzi, C. Glass, M. Genest* (Eds.), *Cognitive assessment.* New York: Guilford.

Napa Scollon, C., Prieto, C. K., & Diener, E. (2009). Experience sampling: promises and pitfalls, strength and weaknesses. *Assessing well-being*, 157-180.

Onaka, T., & Yagi, K. (1990). Differential effects of naloxone on neuroendocrine responses to fear-related emotional stress. *Experimental brain research*, *81*(1), 53-58.

Rodrigues, M., Gonçalves, S., Carneiro, D., Novais, P., & Fdez-Riverola, F. (2013). Keystrokes and clicks: Measuring stress on e-learning students. In *Management Intelligent Systems* (pp. 119- 126). Springer International Publishing.

Russel, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493-502.

Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007, August). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (pp. 4656-4659). IEEE.

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of medical engineering & technology*, *32*(6), 479-484.

Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *International journal of cardiology, 166*(1), 15-29.

Siegrist, J. (2012). Effort-reward imbalance at work: Theory, measurement and evidence. *Department of Medical Sociology, University Düsseldorf, Düsseldorf*.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual review of physiology*, *25*(1), 545-580.

Taelman, J., Vandeput, S., Spaepen, A., & Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering* (pp. 1366-1369). Springer Berlin Heidelberg.

Turpin, G. (1986). Effects of stimulus intensity on autonomic responding: The problem of differentiating orienting and defense reflexes. *Psychophysiology*, *23*(1), 1-14.

Vrijkotte, T. G., Van Doornen, L. J., & De Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*(4), 880-886.

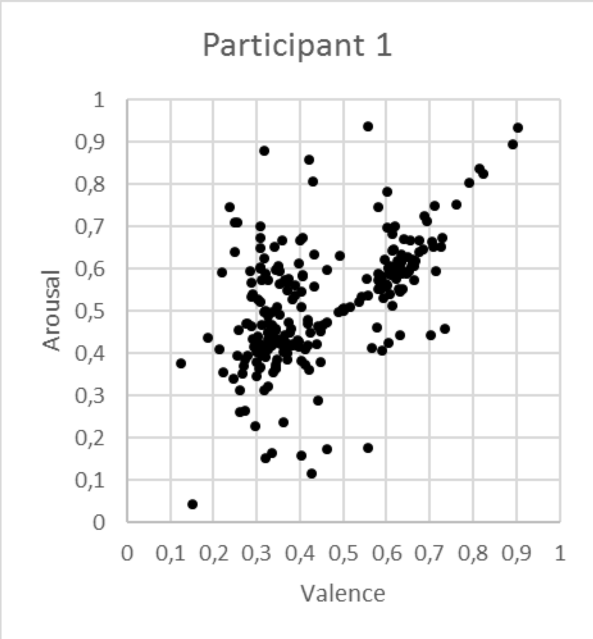
Appendix 1: Interview questions

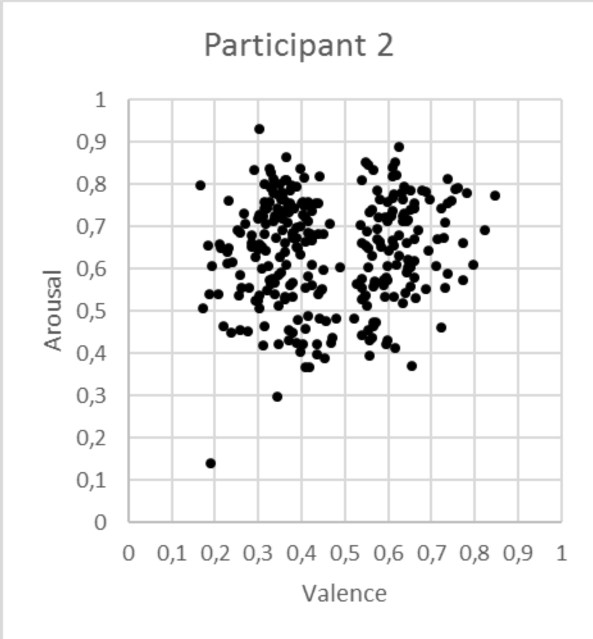
1. Hoe heeft u de studie over het algemeen ervaren?
2. Hoe vond u het om mee te doen aan deze studie op een schaal van 1 tot 10 (waarbij 1 = helemaal niet leuk en 10 = heel leuk)?
3. Hoe heeft u het gebruik van het hartslaghorloge ervaren?
4. Hoe heeft u de elk kwartier terugkerende emotievragen ervaren?
5. Hoe heeft u ervaren dat er regelmatig webcamfoto's van u werden gemaakt om uw gezichtsexpressie te analyseren?
6. Hoe heeft u ervaren dat uw computergedrag werd opgeslagen en geanalyseerd?
7. Hoe ervaarde u deze studie met betrekking tot uw privacy?
8. Vindt u dat uw privacy wordt geschonden bij participatie aan dit onderzoek op een schaal van 1 tot 10 (waarbij 1 = helemaal niet en 10 = ja, heel erg)?
9. Wanneer u werknemer van een bedrijf was, zou u dan toestemming geven de in dit onderzoek verzamelde data te delen met uw manager? Zo ja, welke data wel en welke data niet?
10. Wanneer u werknemer van een bedrijf was, zou u dan toestemming geven de in dit onderzoek verzamelde data te delen met uw collega's? Zo ja, welke data wel en welke data niet?
11. Wanneer u werknemer van een bedrijf was, zou u dan toestemming geven de in dit onderzoek verzamelde data te delen met uw bedrijfsarts? Zo ja, welke data wel en welke data niet?
12. Op basis van deze data zou een systeem ontwikkeld kunnen worden dat werknemers kan helpen hun stress inzichtelijk te maken en eventueel adviezen hierover kan geven. Zou u dit systeem willen gebruiken, wanneer u werknemer van een bedrijf was, om uw stress inzichtelijk te maken, op schaal van 1 tot 10 (waarbij 1 = helemaal niet graag en 10 = heel graag)?
13. Wat ziet u bij dit systeem als de grootste voor- en nadelen?
14. Hoe waarschijnlijk is het dat u nog een keer mee zou doen aan een dergelijk onderzoek op een schaal van 1 tot 10 (waarbij 1 = helemaal niet waarschijnlijk en 10 = heel waarschijnlijk)?

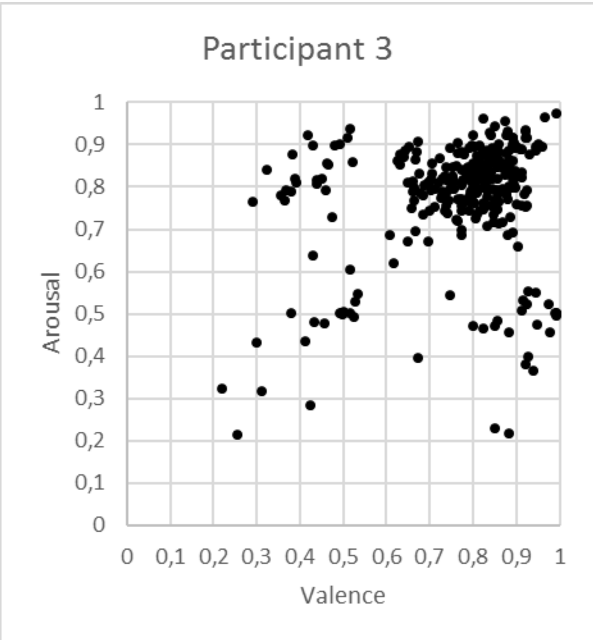
Following questions were not analyzed in this study, but were included for the interest of the bigger project that this study is part of.

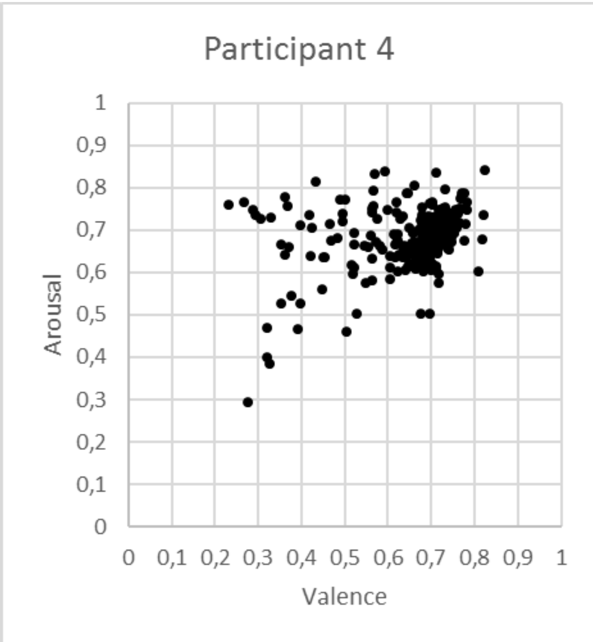
15. Wat denkt u zelf dat er voor uw data uit dit onderzoek komt?
16. Hoeveel vertrouwen heeft u erin dat de maten in dit onderzoek uw stress kunnen voorspellen op een schaal van 1 tot 10 (waarbij 1 = heel weinig vertrouwen en 10 = heel veel vertrouwen)?
17. Hoe goed denkt u dat uw hartslag uw stress kan duiden op een schaal van 1 tot 10 (waarbij 1 = helemaal niet goed en 10 = heel goed)?
18. Hoe goed denkt u dat uw gezichtsexpressie uw stress kan duiden op een schaal van 1 tot 10 (waarbij 1 = helemaal niet goed en 10 = heel goed)?
19. Hoe goed denkt u dat uw computergedrag uw stress kan duiden op een schaal van 1 tot 10 (waarbij 1 = helemaal niet goed en 10 = heel goed)?

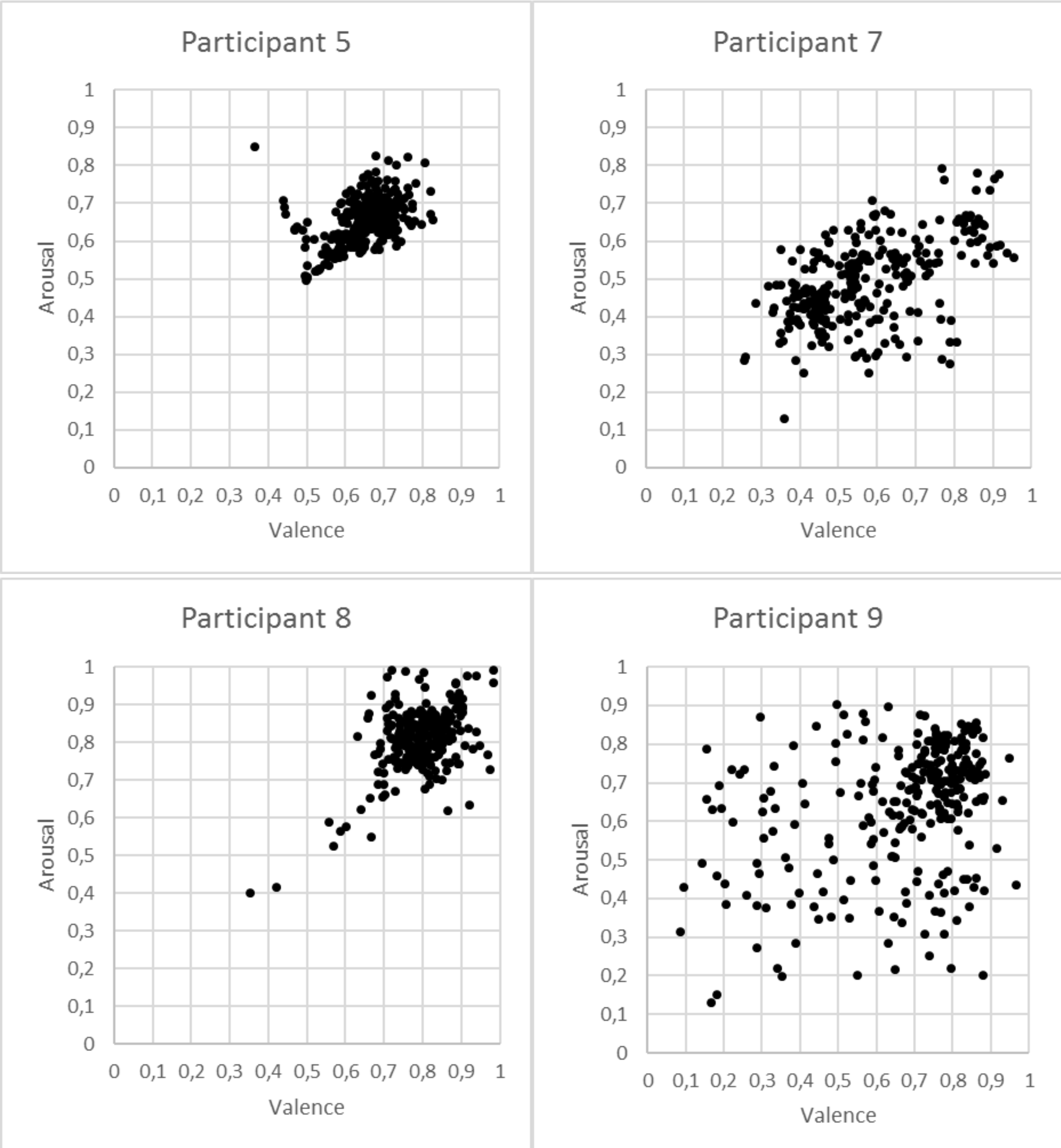
Appendix 2: Subjective emotion spread

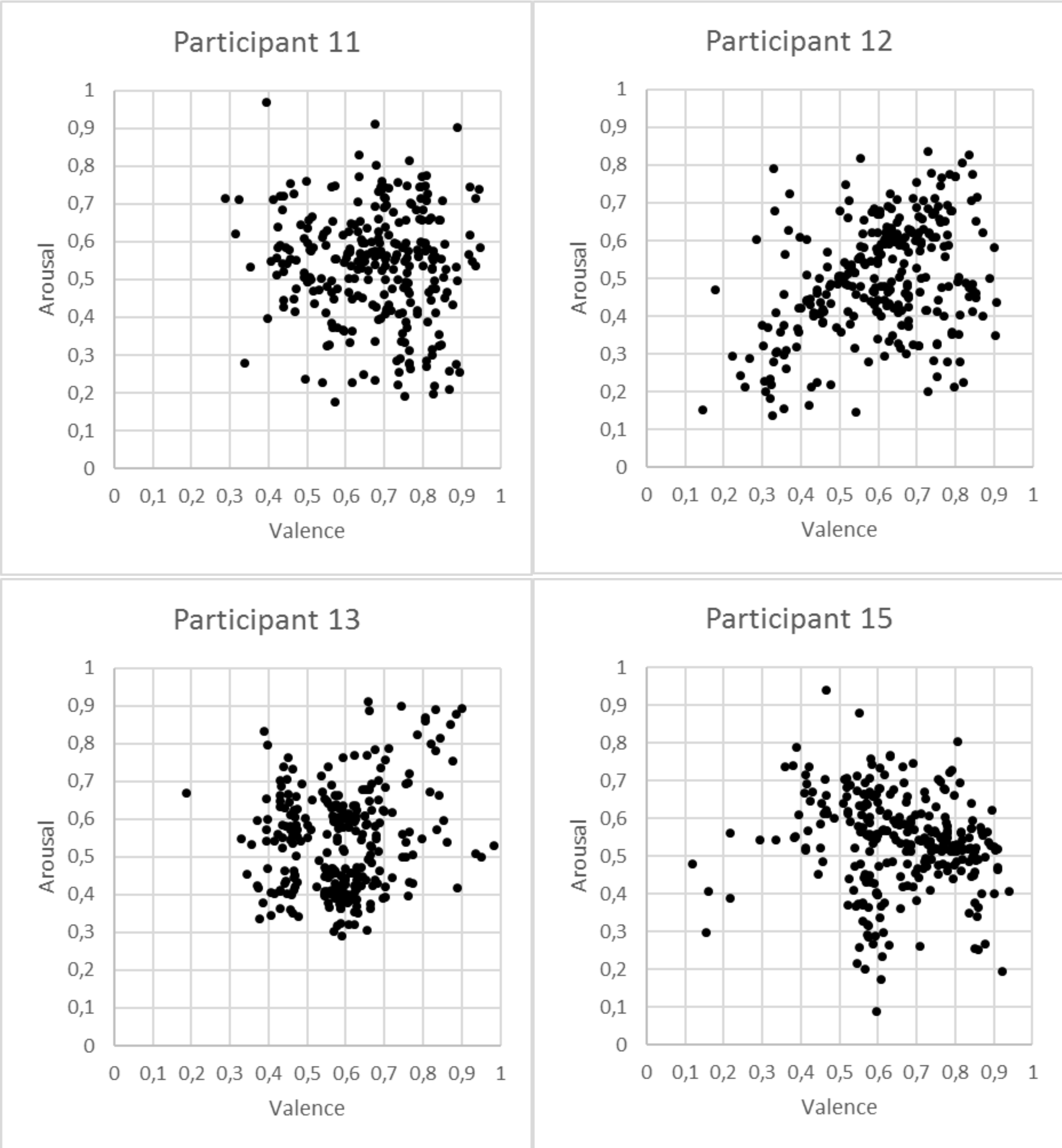









Participant 5
Arousal
Valence
Participant 7
Arousal
Valence
Participant 8
Arousal
Valence
Participant 9
Arousal
Valence
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1


Participant 11
Participant 12
Participant 13
Participant 15
Arousal
Valence
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1

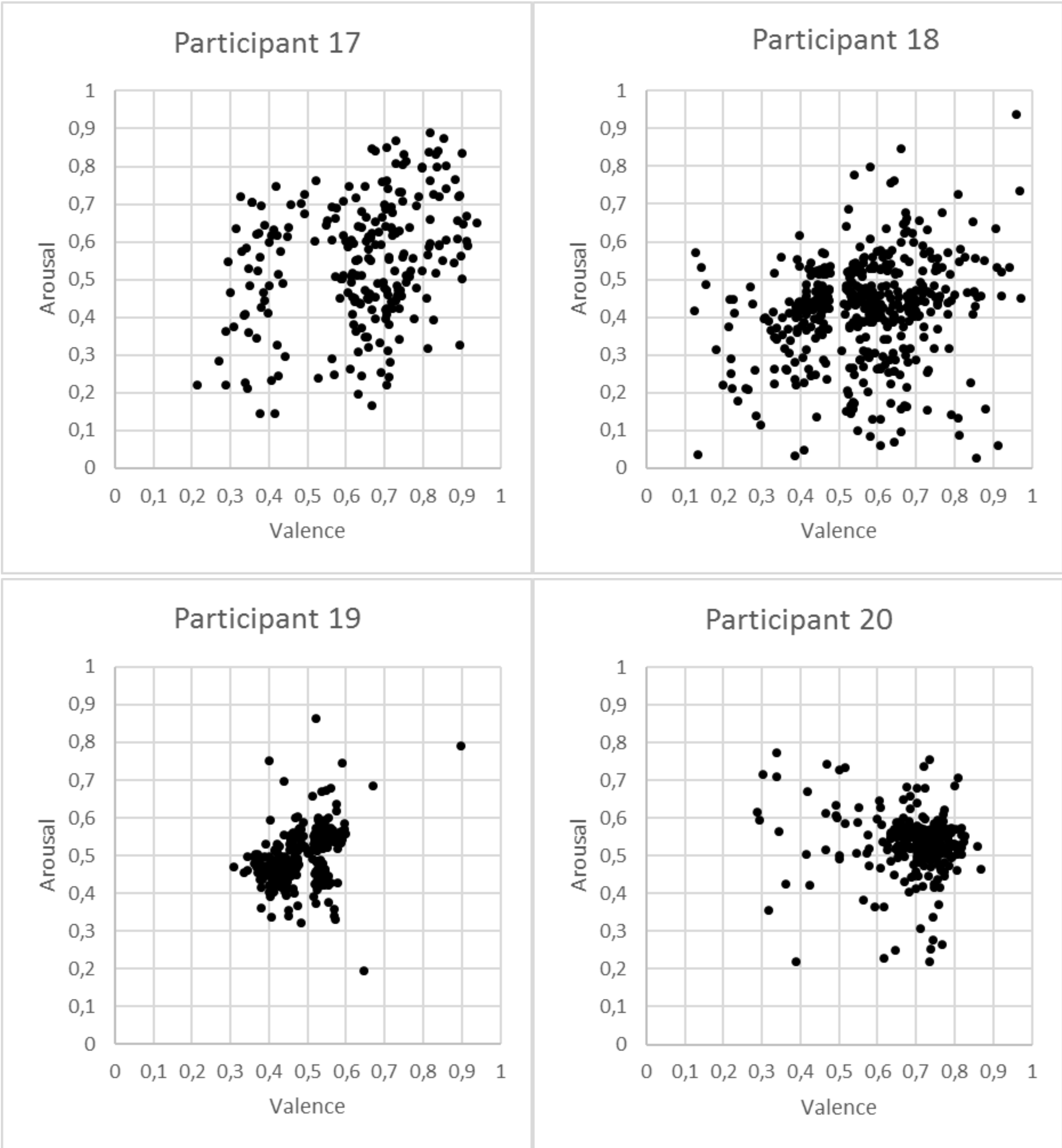
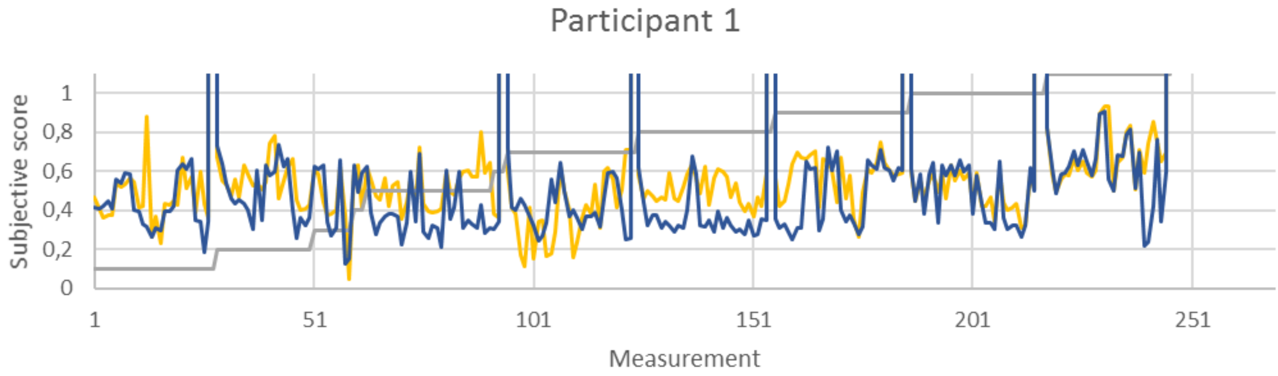
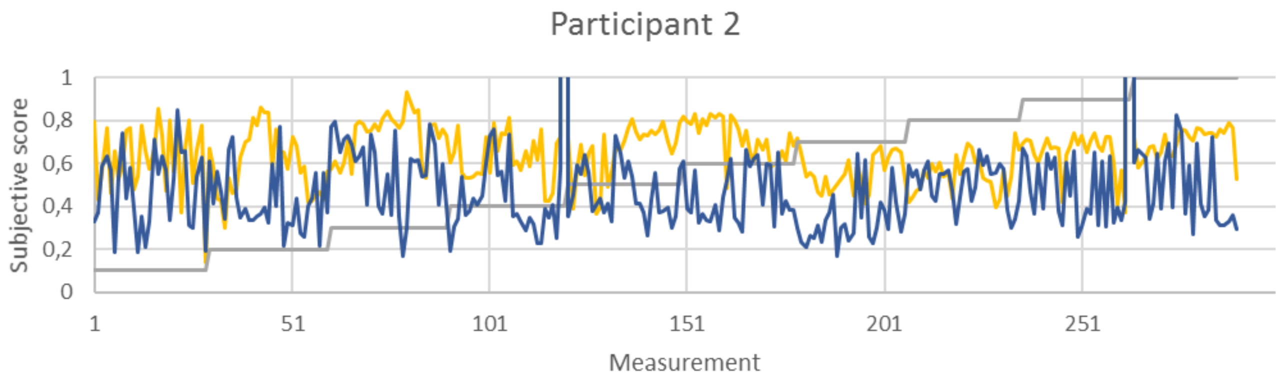
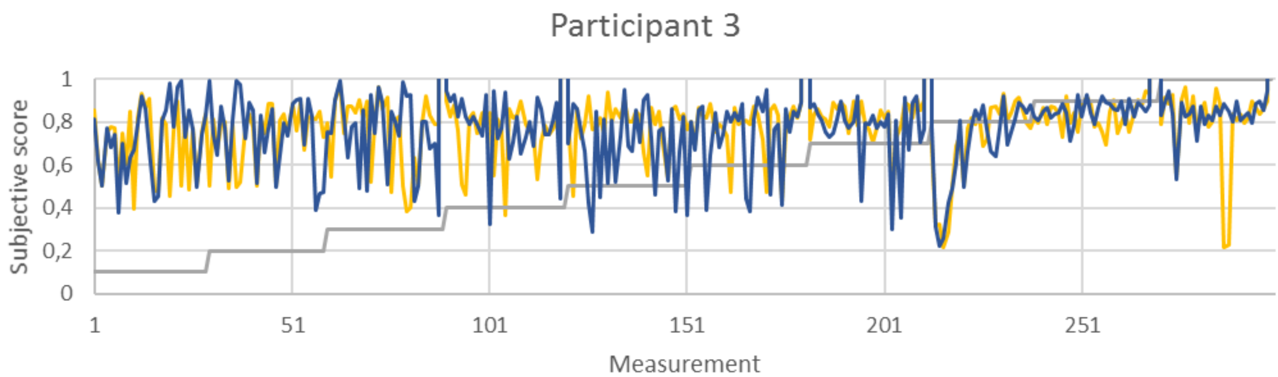


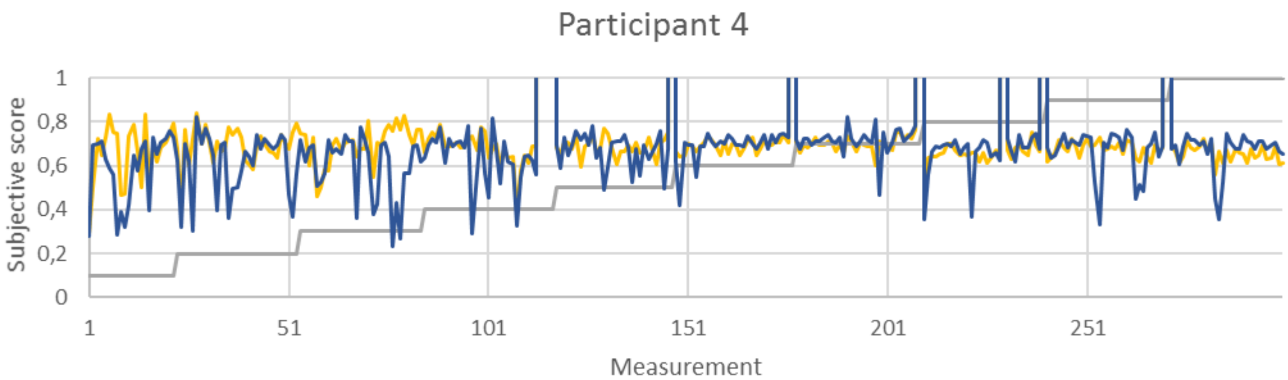
Figure 9. Emotion ratings over all participation days, displayed per participant. Reported arousal is displayed on the y-axis and reported valence is displayed on the x-axis. Visual representation of the emotions ratings, resembles the Affect Grid which was used as method for gathering the emotion ratings.

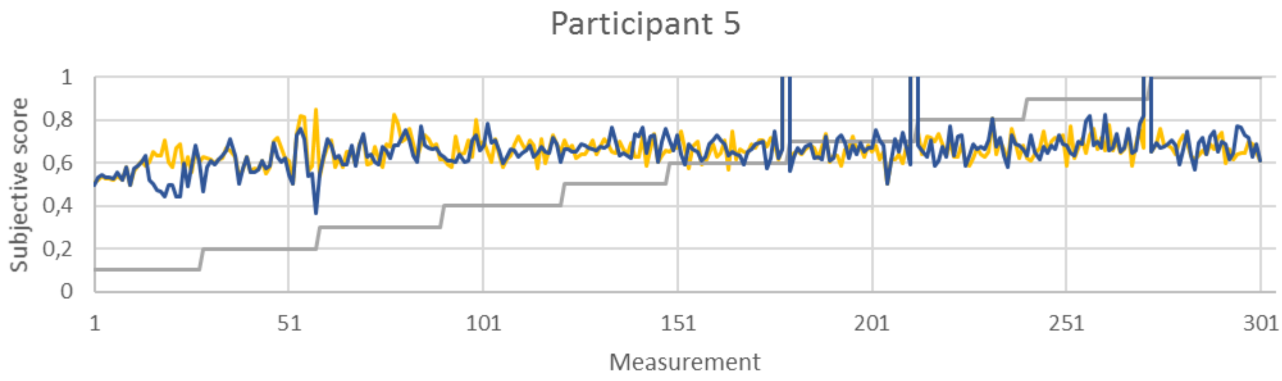
Appendix 3: Subjective emotion over time

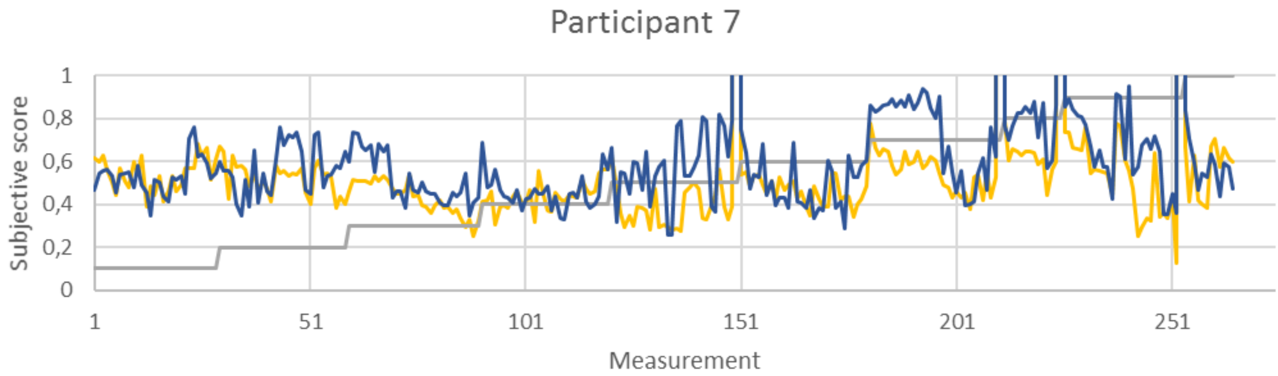


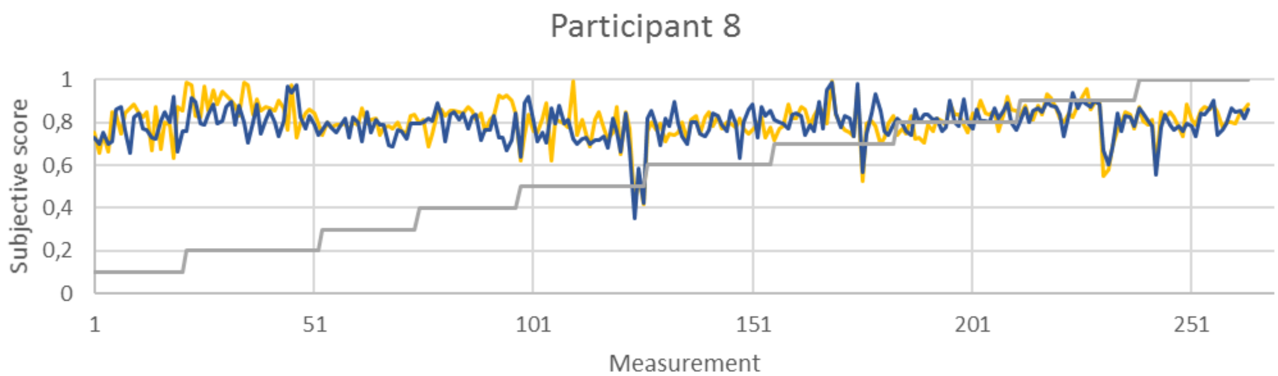


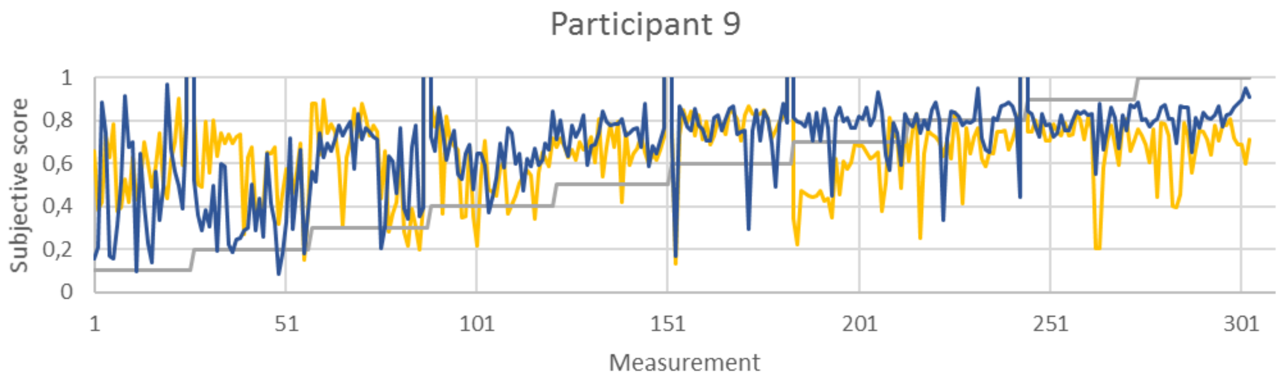


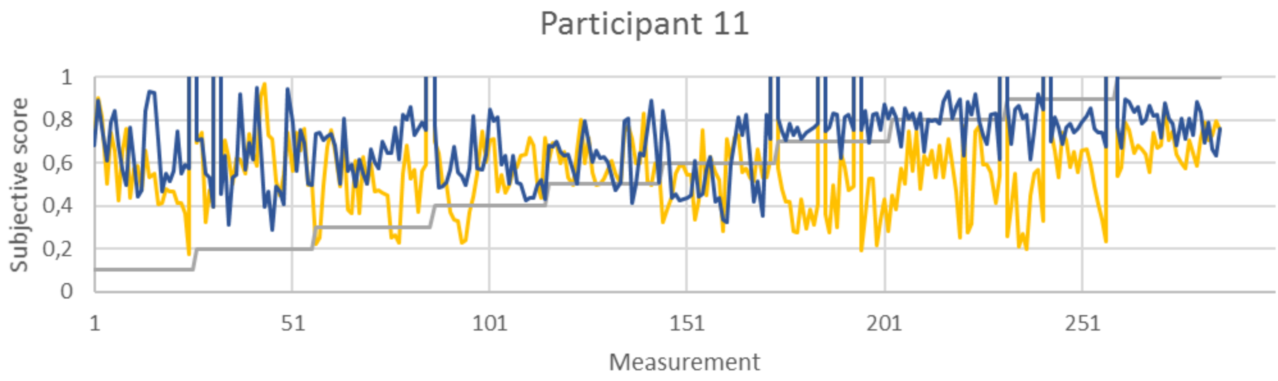


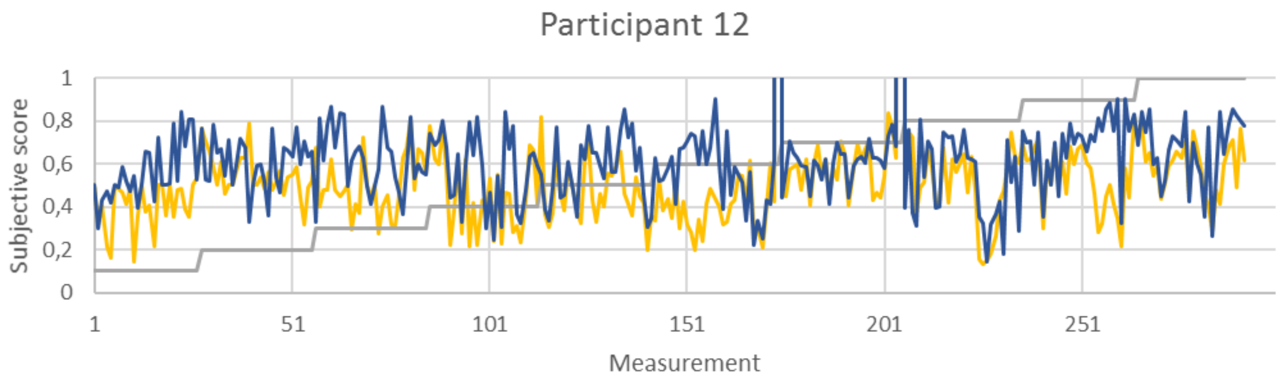

Participant 5
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
301
Measurement

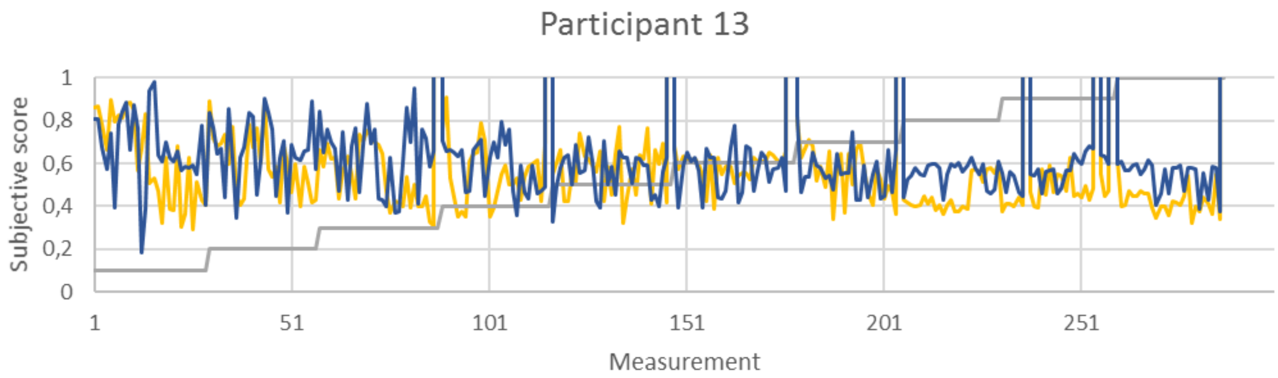

Participant 7
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement

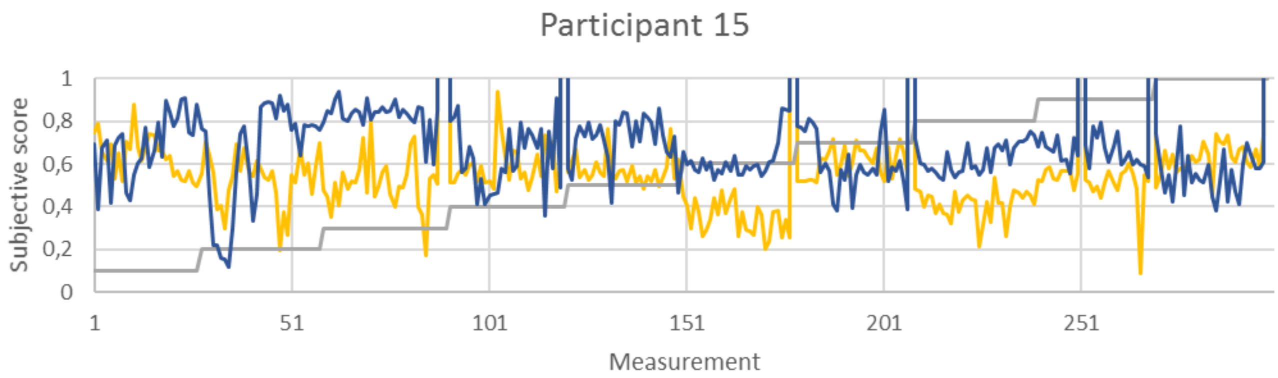

Participant 8
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement


Participant 9
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
301
Measurement


Participant 11
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement


Participant 12
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement


Participant 13
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement


Participant 15
Subjective score
1
0,8
0,6
0,4
0,2
0
1
51
101
151
201
251
Measurement

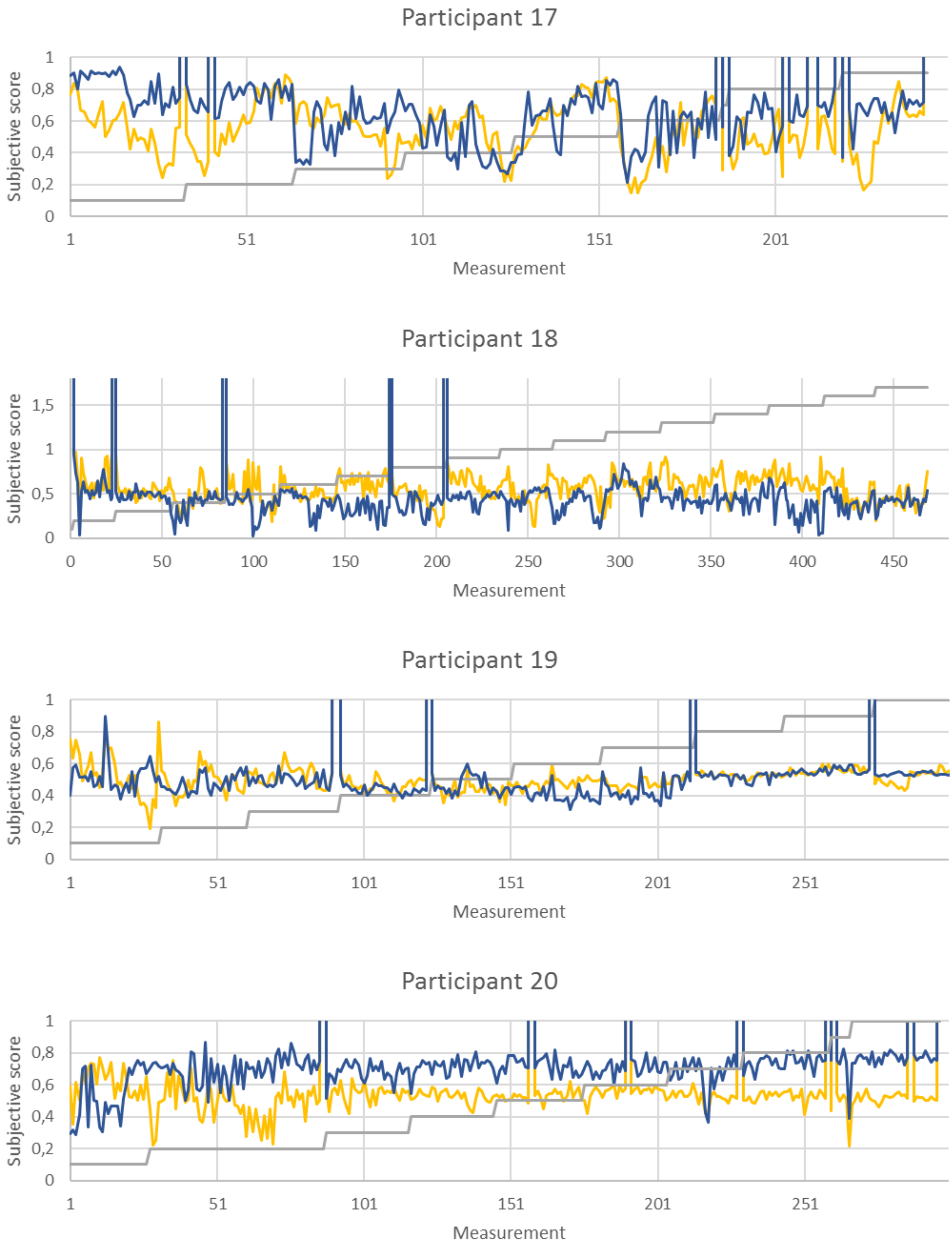
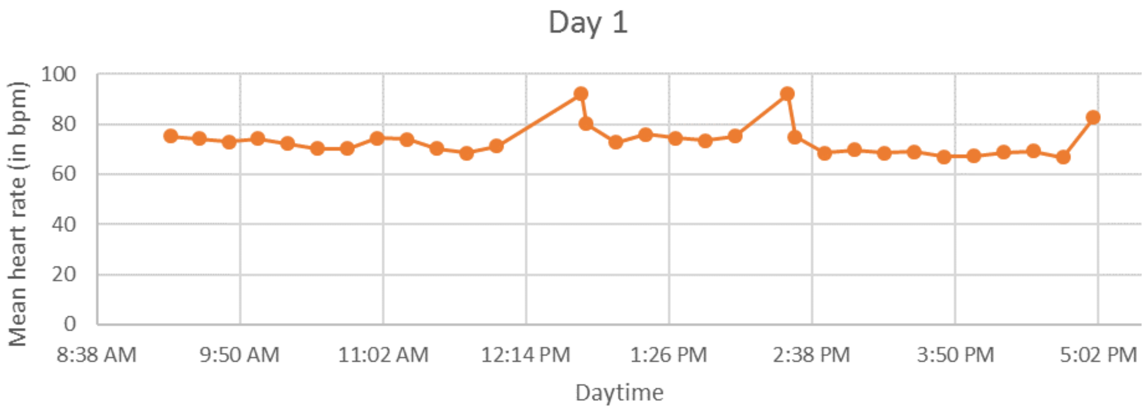
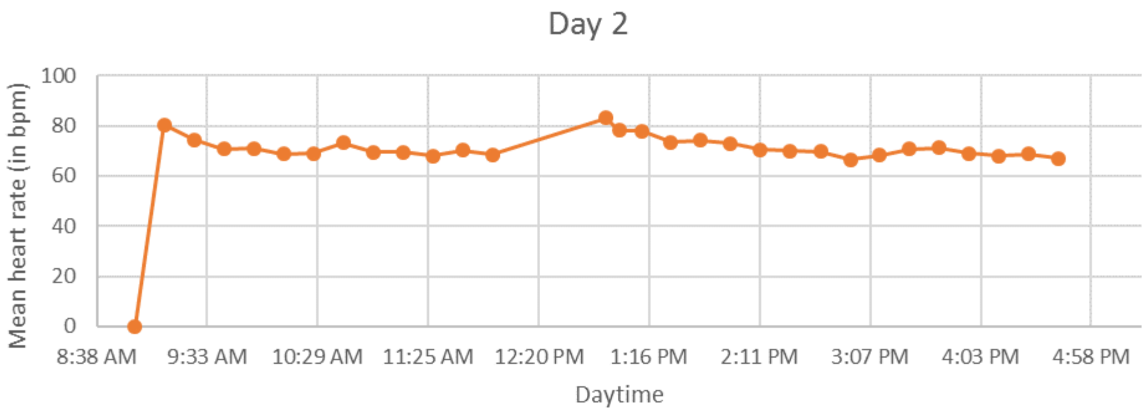
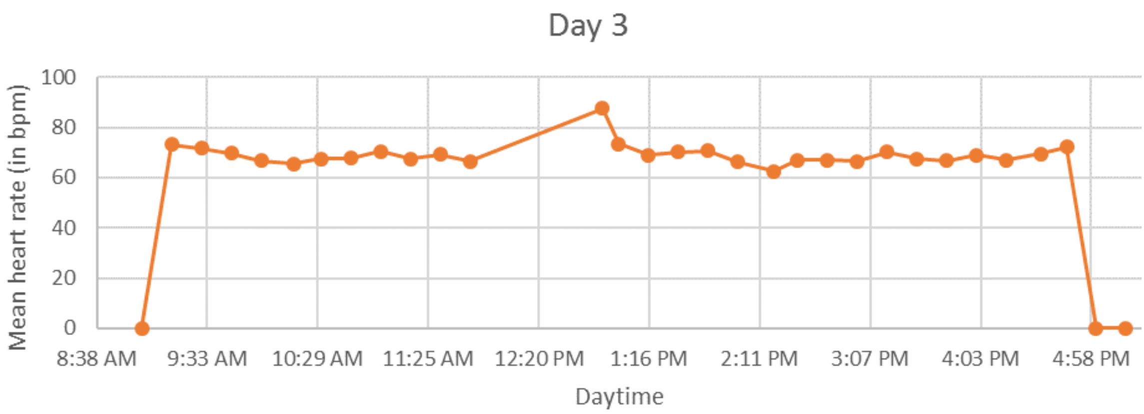


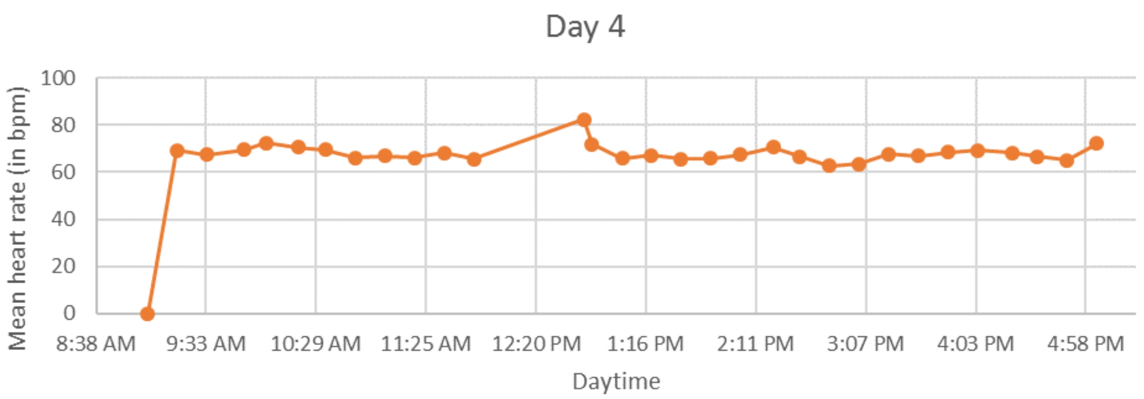
Figure 10. Arousal and valence ratings over total participation period, for each participant separately. Reported arousal is displayed in yellow. Reported valence is displayed in blue. Participation day number divided by ten is displayed in grey. Rating number is displayed on the y-axis.

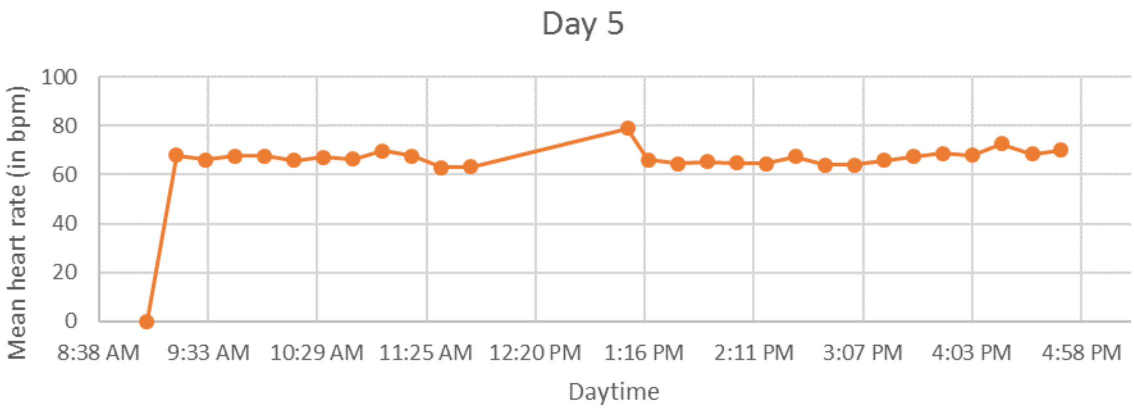
Appendix 4: Heart rate over time

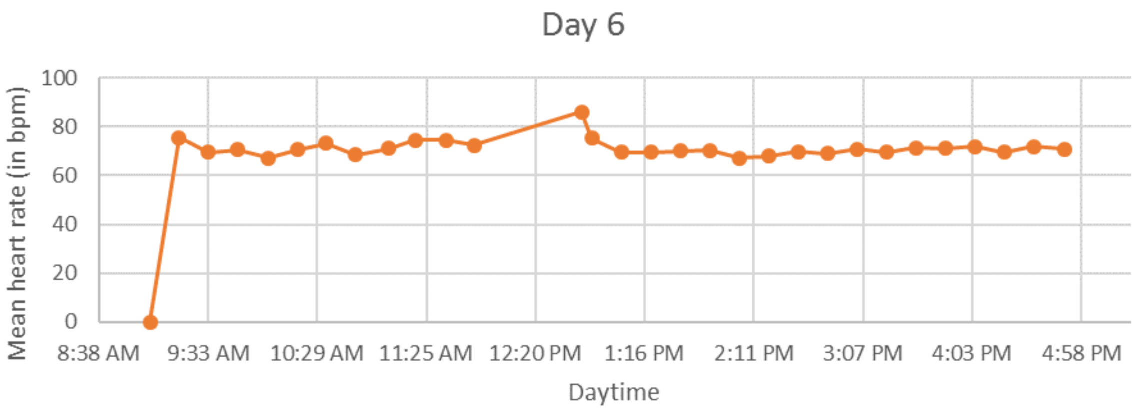


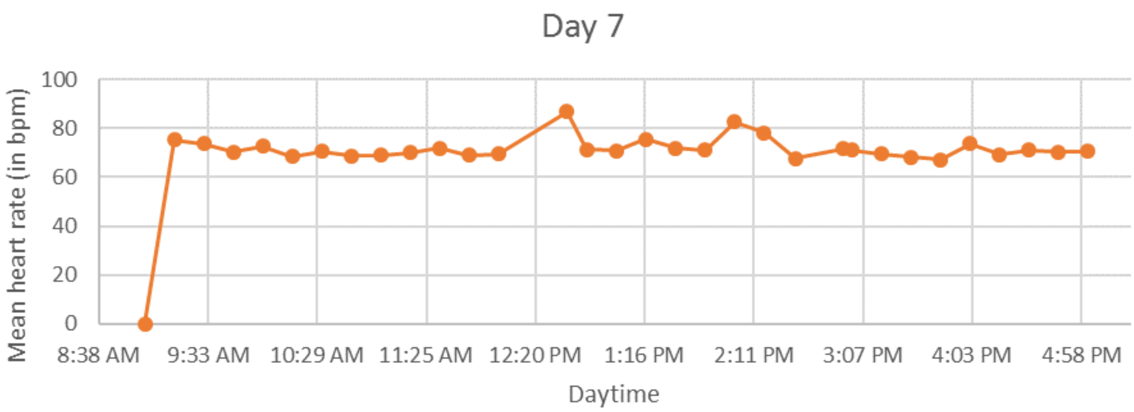


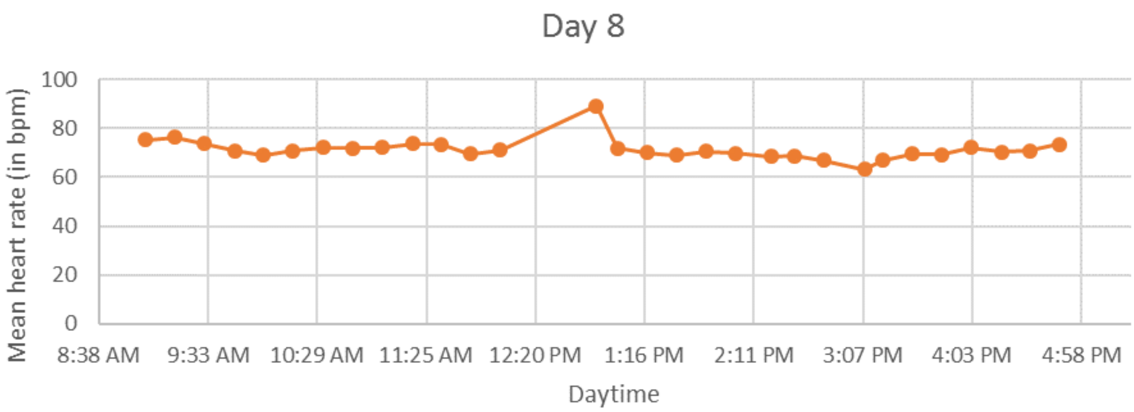





Day 5
Mean heart rate (in bpm)
100
80
60
40
20
0
8:38 AM
9:33 AM
10:29 AM
11:25 AM
12:20 PM
1:16 PM
2:11 PM
3:07 PM
4:03 PM
4:58 PM
Daytime


Day 6
Mean heart rate (in bpm)
100
80
60
40
20
0
8:38 AM
9:33 AM
10:29 AM
11:25 AM
12:20 PM
1:16 PM
2:11 PM
3:07 PM
4:03 PM
4:58 PM
Daytime


Day 7
Mean heart rate (in bpm)
100
80
60
40
20
0
8:38 AM
9:33 AM
10:29 AM
11:25 AM
12:20 PM
1:16 PM
2:11 PM
3:07 PM
4:03 PM
4:58 PM
Daytime


Day 8
Mean heart rate (in bpm)
100
80
60
40
20
0
8:38 AM
9:33 AM
10:29 AM
11:25 AM
12:20 PM
1:16 PM
2:11 PM
3:07 PM
4:03 PM
4:58 PM
Daytime

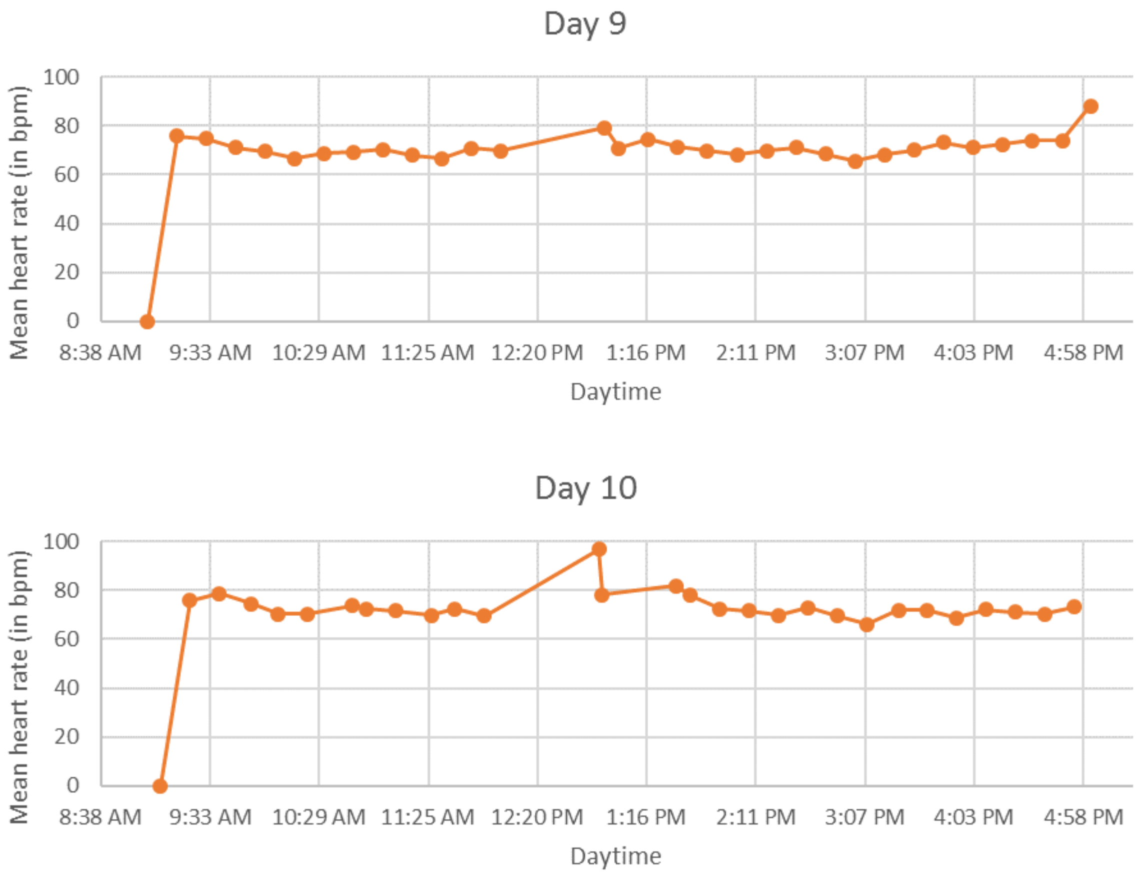


Figure 11. Heart rate of participant 19 over daytime, per participation day. Mean heart rate over previous 15 minutes (in beats per minute) is displayed on the y-axis. Daytime is displayed on the x-axis.

Appendix 5: FaceReader action units over time

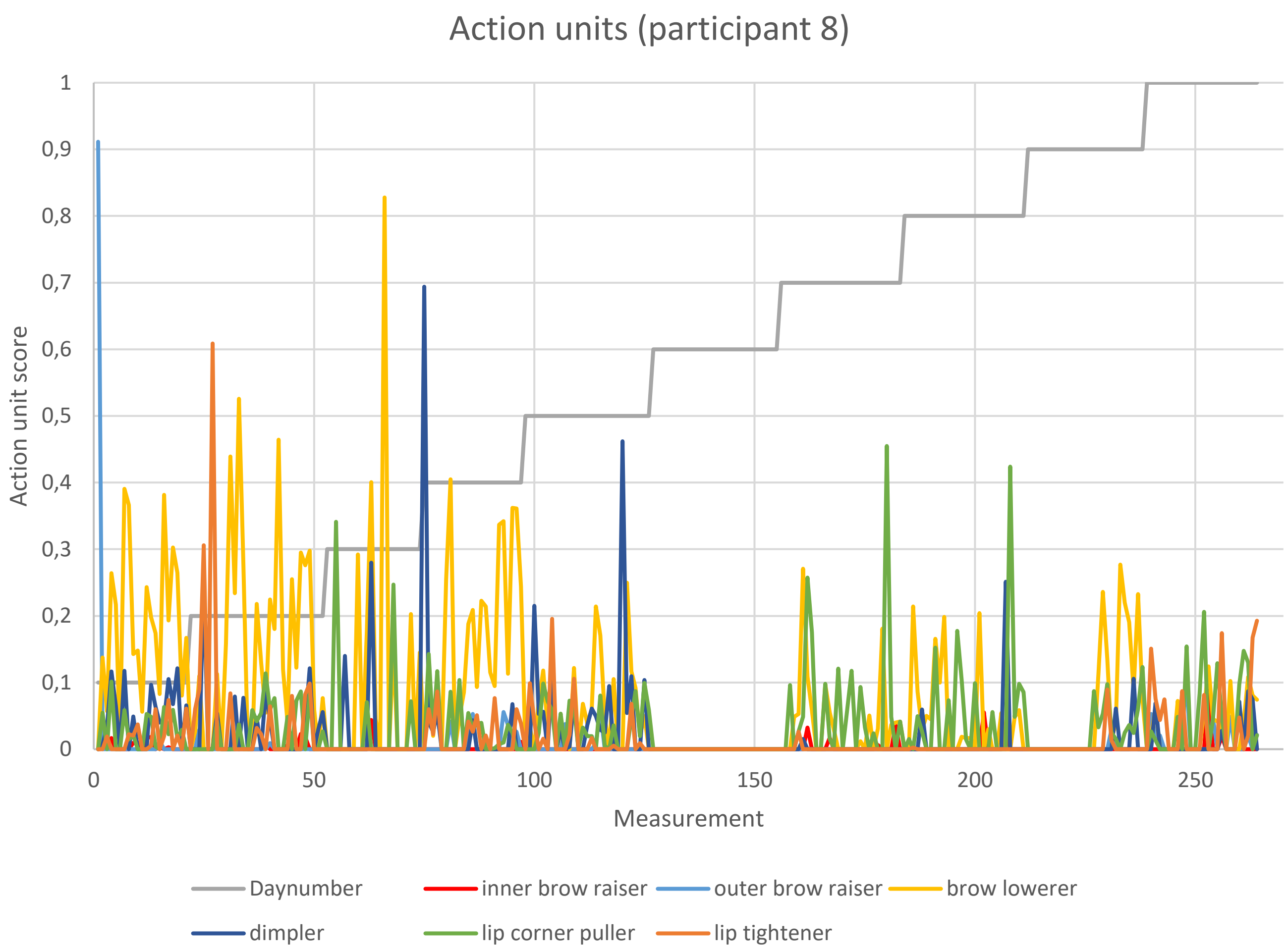


Figure 12. Action unit scores (on a 0 to 1 scale) over total participation period, displayed for participant 8. Participation day number divided by ten is displayed in gray. Measurement number is displayed on the x-axis, action unit score averaged over previous 15 minutes is displayed on the y-axis.

Appendix 6: FaceReader basic emotions over time

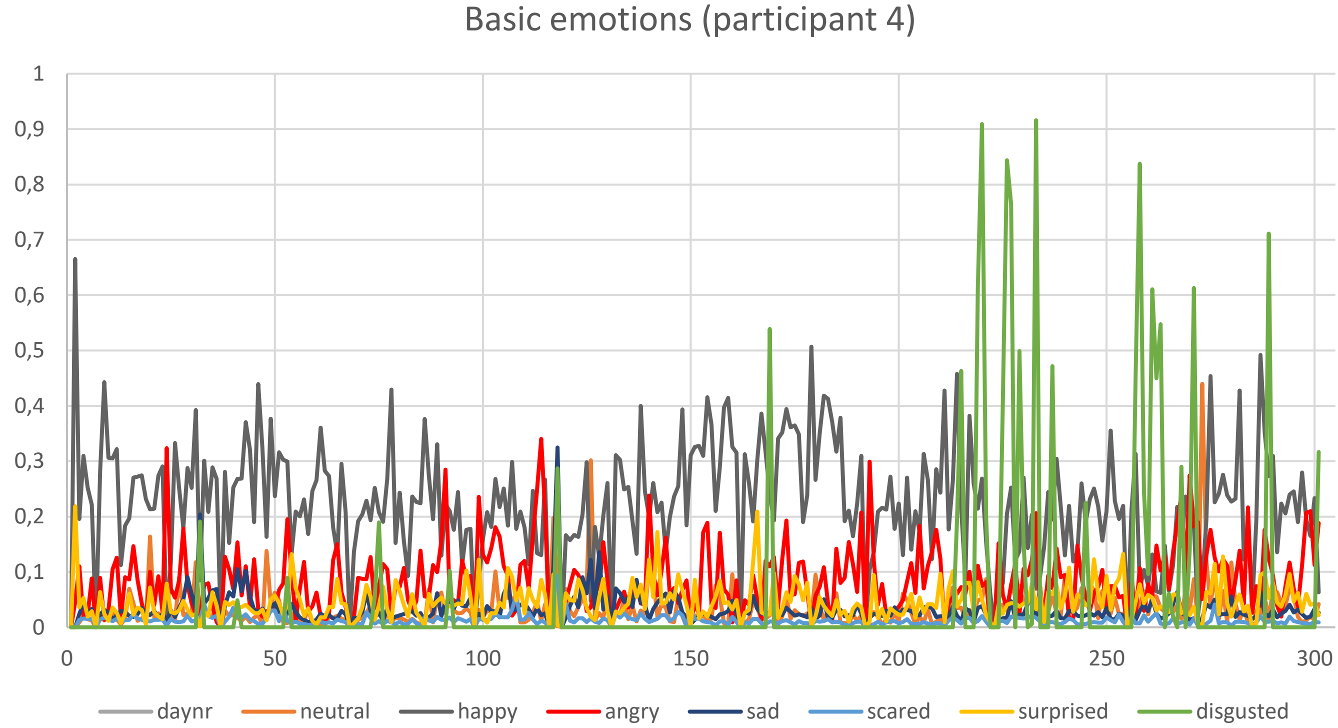


Figure 13. Basic emotion scores (on a 0 to 1 scale) over total participation period, displayed for participant 4. Participation day number divided by ten is displayed in gray. Measurement number is displayed on the x-axis, basic emotion score averaged over previous 15 minutes is displayed on the y-axis.

Appendix 7: Computer usage parameters over time

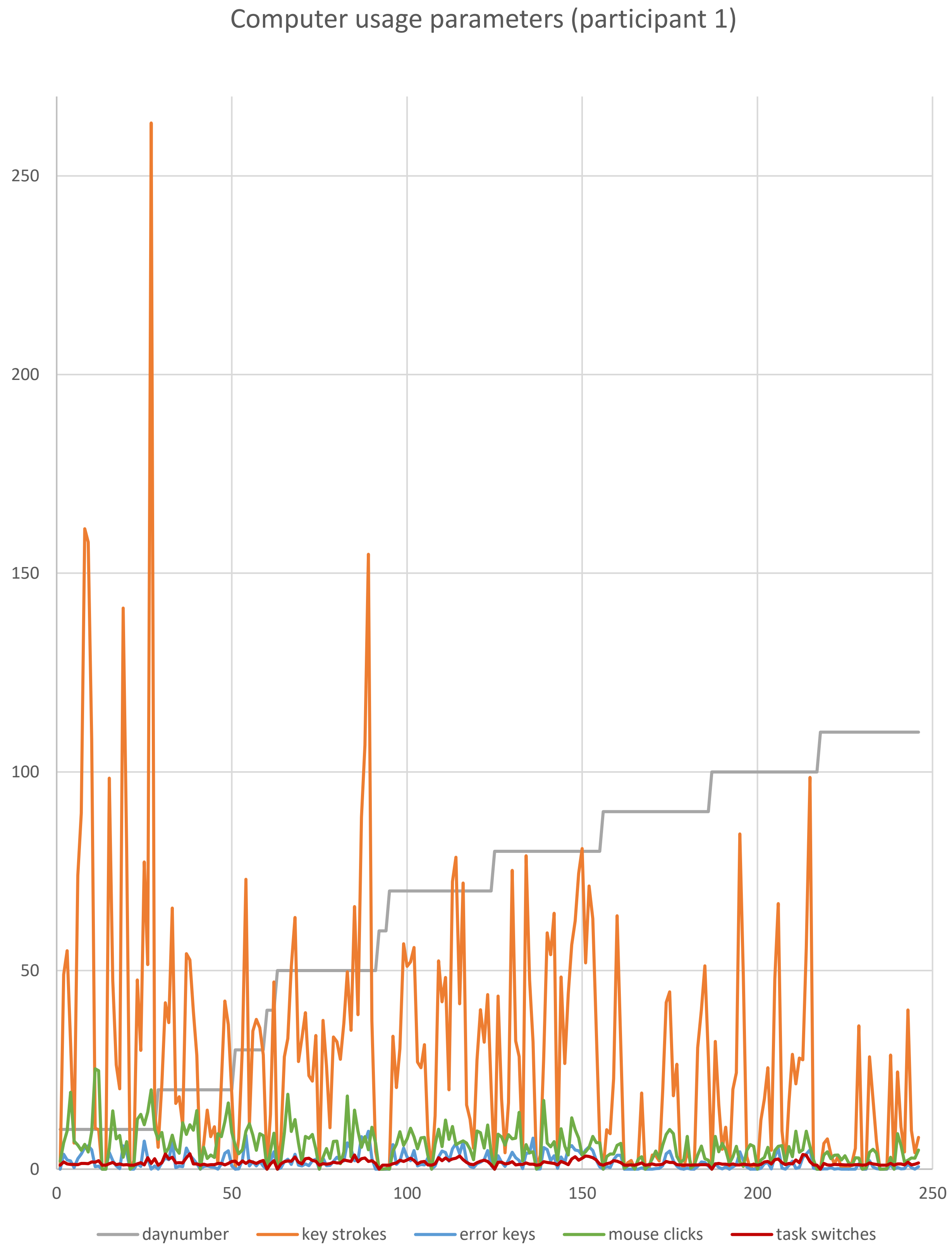


Figure 14. Computer usage parameters over total participation period, displayed for participant 1. Participation day number multiplied by ten is displayed in gray. Measurement number is displayed on the x-axis, mean computer usage score averaged over previous 15 minutes is displayed on the y-axis.