The SWELL Knowledge Work Dataset for Stress and User Modeling Research

Saskia Koldijk
ICIS, Radboud University
& TNO, The Netherlands.
s.koldijk@cs.ru.nl

Maya Sappelli
ICIS, Radboud University
& TNO, The Netherlands.
m.sappelli@cs.ru.nl

Suzan Verberne
ICIS, Radboud University, The Netherlands.
s.verberne@cs.ru.nl

Mark A. Neerincx
II, Delft University of Technology
& TNO, The Netherlands.
mark.neerincx@tno.nl

Wessel Kraaij
ICIS, Radboud University
& TNO, The Netherlands.
wessel.kraaij@tno.nl

ABSTRACT

This paper describes the new multimodal SWELL knowledge work (SWELL-KW) dataset for research on stress and user modeling. The dataset was collected in an experiment, in which 25 people performed typical knowledge work (writing reports, making presentations, reading e-mail, searching for information). We manipulated their working conditions with the stressors: email interruptions and time pressure. A varied set of data was recorded: computer logging, facial expression from camera recordings, body postures from a Kinect 3D sensor and heart rate (variability) and skin conductance from body sensors. The dataset made available not only contains raw data, but also preprocessed data and extracted features. The participants' subjective experience on task load, mental effort, emotion and perceived stress was assessed with validated questionnaires as a ground truth. The resulting dataset on working behavior and affect is a valuable contribution to several research fields, such as work psychology, user modeling and context aware systems.

Categories and Subject Descriptors

E.0 [**Data**]: General; H.1.2 [**Models and Principles**]: User/ Machine Systems—*Human factors, Human information processing*; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms

Human Factors, Experimentation, Measurement

Keywords

Dataset; stress; mental state; facial expressions; body postures; computer interaction; physiology

ICMI '14, November 12 - 16 2014, Istanbul, Turkey

ACM 978-1-4503-2885-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2663204.2663257.

1. INTRODUCTION

Nowadays most work involves computer usage and information processing. People that use and produce information as their main task are called knowledge workers. They typically experience all sorts of demands during their work days, such as several tasks that need to be finished before a deadline (high workload, temporal demand). For this they need to combine different information sources, for example from the internet (requiring mental effort). Incoming emails may be an important source of distraction during a task (potentially causing frustration). In case people feel they cannot handle the demands posed upon them, they can experience stress [3]. Stress is a broad concept referring to psychological and biological processes during emotional and cognitive demanding situations. We follow a pragmatic approach and define stress in terms of: (1) the task load, which poses demands on the worker, (2) the mental effort, which the worker needs to handle a task and (3) the emotional response to a task. In this paper, we focus on short term effects of stressors that can be measured within a 3 hour work session.

Stress is a well-known experience in our connected environments. Ruff [16] speaks of 'hurry sickness' as "the belief that one must constantly rush to keep pace with time" and 'plugged in compulsion' as "the strong need to check mail and the internet to stay in touch". Mark, Gudith and Klocke [13] investigated the cost of interruptions and came to the conclusion that "after only 20 minutes of interrupted performance people reported significantly higher stress, frustration, workload, effort and pressure". Stress from time to time with enough room for recovery is no problem [2]. However, when stress builds up this can be a danger to well-being, in the worst case resulting in burn-out.

In our project SWELL (Smart Reasoning for Well-being at Home and at Work)[1] we aim to develop ICT tools that help knowledge workers to cope with stress and gain more well-being at work [8]. We want to interpret recordings in the office real-time in terms of stress and the context in which it appears. Based upon this information, we aim to develop coaching software that can help knowledge workers to gain a more healthy work style. Moreover, we want to develop smart information support tools that assist the knowledge worker in handling the large amount of (incom-

[1] `http://www.swell-project.net`

ing) information he has to work with. In this way, we extend traditional approaches (e.g. questionnaires or department wide interventions, [11, 12]) by empowering individual users to self-manage their own well-being.

To be able to develop the ICT tools we envision, research communities like work psychology, user modeling and context aware systems are in need of a good dataset. This dataset should ideally have the following characteristics: Data should be recorded in a realistic office setting. Stressors should be manipulated in a systematic way and subjective experience should be assessed with validated questionnaires, to be able to investigate the effects of stressors. A multimodal set of sensors from different research fields should be used, to enable multidisciplinary research. The focus should lie on sensors that are readily available in office settings, to make the to be developed system usable outside the lab. To our knowledge no such dataset existed.

In this paper we present a newly collected rich dataset which has these characteristics. Our dataset overcomes three drawbacks that are typically observed in related work:

- Instead of a rather artificial task, participants perform natural office work with systematically manipulated stressors.
- Instead of expensive and/or obtrusive equipment, we decided to combine a variety of sensors that can easily be deployed in real-world office settings.
- Instead of only collecting data for our own use, we will make the anonymized dataset available for access by the scientific community, for benchmarking of techniques and algorithms. We will not only provide raw data, but also data in preprocessed and interpreted form.

With our new dataset, we aim to bring research on psychology and computer science together. With this dataset, research questions from several fields can be answered, for example:

- Work psychology: What effect do stressors like time pressure have on the working behavior of knowledge workers? What is the effect of an incoming email? What effect do stressors have on subjective experience of task load, mental effort, emotion or perceived stress? What is the relation between what people mean when they feel 'stressed' and the concepts of arousal, mental effort and valence? Do we see effects of stressors in physiological sensor data?
- User modeling: Can we estimate the mental state or emotion of knowledge workers from unobtrusive sensor data? Do knowledge workers show particular affective expressions during computer work? Are there typical facial expressions or postures that are indicative of mental effort, high workload or stress?
- Context aware systems: Can we automatically determine the task or topic someone is working on? Is there a relation between stress and the context in which it occurs? Can we filter irrelevant emails? Can we make information retrieval more context aware?

In this paper we mainly focus on work stress and user modeling. For more details about the data regarding context recognition and information support, we refer to [17].

This paper is structured as follows. We first present some related work (Section 2). We then outline our experimental setup in which the dataset was collected (Section 3). We describe the dataset in detail in Section 4. Some example analyses are presented in Section 5. We finish with a Discussion and Conclusion (Section 6 and 7).

2. RELATED WORK

In this section we present some related research on work psychology and user modeling, in which sensor data is used to estimate stress, mental or affective states. We specifically address the type of sensors used, the context in which data has been collected and the kind of inferences that have been made. This gives a theoretical framework for research on our collected dataset.

In work psychology, questionnaires are commonly used to get insight in the general working experiences (e.g. [20]). Advances in sensing, as well as the quantified self movement make it possible to extend such an approach with on-site measurements.

Work in the area of affective computing investigates the possibility of inferring stress and emotion from sensor data. Most often, physiological sensors are used and data are collected in experimental environments. In research by Riera et al. [15], for example, electroencephalography (EEG) and facial electromyography (EMG) data were collected. The authors show that EEG and EMG can be used for monitoring emotion (valence and arousal) and stress. Although its great potential, we think deploying EEG in a daily office setting is not yet realistic. Other common measurements in stress research are pupil diameter and electrocardiogram (ECG). Mokhayeri, Akbarzadeh-T and Toosizadeh [14], for example, collected such data in context of the Stroop colorword (SCW) test. They state that pupil diameter and ECG have great potential for stress detection. However, the question that arises is: can we also make an estimate of affective and mental states outside the lab? We see some potential for ECG measurements, with the rise of wearable sensors, which are becoming more and more integrated into devices as watches and bracelets. But, besides measuring the physiological stress response directly, we also see great potential in measuring outward characteristics, such as facial expressions, postures or computer interactions as indicators for the user's state.

In related work, facial expressions are widely used for inferring emotions. The data are often recorded while emotions are induced in participants. The publicly available multimodal dataset described by Soleymani et al. [18], for example, was collected in context of watching emotion inducing video clips and consists of: face videos, audio signals, eye gaze data and physiological signals (EEG, ECG, GSR, respiration amplitude, skin temperature). Although this dataset is very interesting, emotions in a daily computer work context are probably less intense than the valence or arousal experienced during watching a movie clip. An interesting question is whether people show facial emotions during computer work, and whether their facial expressions are indicative of mental states. Preliminary results by Dinges et al. [4] suggest that high and low stressor situations could be discriminated based on facial activity in mouth and eyebrow regions.

Regarding postures, Kapoor and Picard [7] present research in which posture data was collected together with facial expressions and computer information while children solved an educational computer puzzle. Sensors in the chair were used to extract posture features (like leaning back, sitting upright) and activity level (low, medium, high). Posture information yielded the highest unimodal accuracy (80%) for estimating interest (vs. uninterest). Performance was further improved by adding facial expression and computer information. We conclude that posture information and movement are an interesting source for estimating the users' mental state. We see potential for posture measurements in the office, as with the Kinect recently an affordable 3D camera with skeleton detection has entered the market.

Finally, in some research, stress or emotions are estimated from computer interaction data. Vizer, Zhou and Sears [19], for example, investigated the effect of stress on typing patterns. Participants first performed a mentally or physically stressful task (e.g. remembering digits or exercising) and were then asked to write an email. Results indicate that stress can produce changes in typing patterns. This makes computer logging a valuable sensor for user state modeling. We think not only typing patterns, but also more general computer behavior might be indicative of mental states, like the amount of window switching, number of typos or time spent browsing.

Besides inferring stress or particular mental or affective states, the context in which they appear can be interesting. Computer interactions give rich insight in the user's current working behavior. Research by Koldijk et al. [10] shows that it is possible to infer the task someone is working on from computer interaction data. Moreover, one could add analysis of contents worked on.

To conclude, research from various related fields shows the potential of using sensors for estimating stress, mental and affective states and the context in which they appear. In each field, a particular setup of sensors is used. We decided to combine several of these in our unique dataset: computer interactions, video for facial expressions, Kinect 3D for postures and body sensors for heart rate and skin conductance.

3. DATA COLLECTION CONTEXT

In this section we present the experimental setup that was used to collect data.

3.1 Design

In our experiment we manipulated the conditions under which our participants worked:

- Neutral: the participant was allowed to work on the tasks as long as he/she needed. After a maximum of 45 minutes the participant was asked to stop and told that enough data of 'normal working' was collected.
- Stressor 'Time pressure': the time to finish all tasks was 2/3 of the time the participant needed in the neutral condition (and maximally 30 minutes).
- Stressor 'Interruptions': 8 emails were sent to the participant during the task. Some were relevant to one of the tasks, others were irrelevant. Some emails required a reply, others did not. Examples are: "Could you look up when Einstein was born?" or "I found this website with lots of nice pictures for presentations.".

All participants worked under all 3 conditions. The neutral condition was always the first condition, in order to collect an uninfluenced baseline of normal working. The order of the two stressor conditions was counterbalanced, see Figure 1. The within-subject design included relaxation breaks to start each condition in a well-rested state.

Order		Block1	Q		Block2	Q		Block3	Q
A	Relax	Neutral		Relax	Stressor interruptions		Relax	Stressor time pressure	
B		Neutral			Stressor time pressure			Stressor interruptions	

Figure 1: Design. For 13 participants order A was used, for 12 participants order B.

3.2 Tasks

The participants performed knowledge worker tasks on a desktop computer in a controlled lab setting. We asked them to write reports and make presentations on predefined topics (in English). We selected 6 topics on which people with various backgrounds could work:

- 3 opinion topics: Experience and opinion about 'stress at work', 'healthy living' and 'privacy on the internet'.
- 3 information topics: 'describe 5 Tourist attractions in Perth (West Australia)', 'plan a coast to coast road-trip in the USA' and 'write about the life of Napoleon'.

Some detail on what to include in the report was also given. Participants' were allowed to look for information on the internet and use documents that we previously stored on the computer. This setting is typical for knowledge work as available information can be combined with the worker's own input in a coherent way, with the purpose of generating a new information product. During the task the email program Outlook was running and participants were told to make use of information from incoming emails and reply when necessary. In this way, a realistic office work scenario was created.

We wanted to ensure that the participants worked on the tasks seriously. Therefore we told them that it was important to finish all required tasks for receiving the full subject fee. Moreover we told them that they would have to give one of the prepared presentations. After the experiment, we debriefed all participants and informed them that they did not need to give a presentation and would get the full subject fee.

3.3 Procedure

To be able to record stress responses as a result of our experimental manipulations, we instructed the participants to not smoke or drink caffeine 3 hours prior to the experiment, as these are possible confounders. Before the experiment started, the experiment and recordings were explained and all participants signed a consent form to confirm that the recorded data may be used for research purposes. Body sensors were applied and while the experimenter checked the recordings, the participant read the experiment instructions and filled in a general questionnaire.

The experiment was divided into three blocks for the different stressor conditions, each taking approximately one hour. Each of the experimental blocks started with a relaxation phase of about 8 minutes (which is typical for stress research) in which a nature film clip was shown. Then the participants received instructions on the tasks to work on. In each block the participants were provided with 2 of the 6 topics, which were randomly selected from the list, in such a way that always an opinion topic was combined with an information topic. The participants were instructed to write 2 reports, one on each topic, and make 1 presentation on one of the topics (participants could choose the topic). To prevent learning effects, the participants were provided with different topics in every block. In both stressor conditions, participants were provided a count-down clock for showing them the remaining time.

After completion of the tasks, the participants were asked to fill in a questionnaire about the current block. This procedure of relaxation, tasks execution and questionnaire was then repeated for block 2 and 3 (see Figure 1). Between the conditions the subjects were allowed a short break and the total experiment took about 3 hours. After the experiment the participants were debriefed.

3.4 Apparatus

Participants performed their tasks on a computer (Dell Latitude E6400) with Windows 7 Professional with a 17 inch screen and mouse and keyboard (see Figure 2). Office 2010 was installed, which the participants used for email (Outlook), report writing (Word) and making presentations (Powerpoint). As a browser, Internet Explorer was used with Google as default search engine. The start page of Internet Explorer was www.google.nl.

Figure 2: Experimental set-up.

3.5 Subjective Ratings

To collect a ground truth of the subjective experience after each block, we used a combination of validated questionnaires. Task load (in terms of mental demand, physical demand, temporal demand, effort, performance and frustration) was determined with the 'NASA-Task Load Index' [6]. Mental effort was assessed with the 'Rating Scale Mental Effort' [21]. Emotion response (in terms of valence, arousal and dominance) was determined with the 'Self-Assessment-Manikin Scale' [1]. Moreover, we asked participants to report their perceived stress on a visual analog scale from 'not stressed' to 'very stressed' (10 point scale).

Furthermore, we asked the participant's to fill in the 'Internal Control Index' questionnaire [5]. People with an internal locus of control tend to praise or blame themselves, whereas people with an external locus tend to praise or blame external factors. This might be of influence on participants' stress perception or behavior. Moreover, participants were asked to rate their interest in the topics, as well as how difficult they found it to write a report or make a presentation on a topic on a 7-point Likert scale (from 'not interesting / difficult' to 'very interesting / difficult').

3.6 Sensors

Computer logging. Computer interactions were logged with the key-logging application uLog (version 3.2.5, by Noldus Information Technology), which ran as a background application on the users' computer.

Video. Video recordings of the participants face and upper body were made with a high-resolution USB camera (iDS uEye UI-1490RE, 1152x768) which was positioned below the participants' monitor. The AVI files from the USB camera were further analyzed using the facial expression analysis software FaceReader (version 5.0.7 RC 4.5 (Beta)). An additional webcam (Philips SPC 900NC, SVGA resolution) was placed above the participants' monitor.

Kinect 3D. The participants body posture was recorded with a Kinect (for Windows, model 1517) depth camera. The camera was placed in front of the participants at a distance of about 2 meters, such that their whole body, including their legs under the desk were visible (see Figure 2). Besides 3D depth video, Kinect also recorded normal RGB video. Recordings were made with Kinect Studio (v1.7.0), which resulted in xed-files. From the recorded Kinect data the depth image and information on the skeletal model were extracted using the Windows Kinect SDK (v1.7). We smoothened the data with several predefined filters.

Body sensors. ECG was recorded using a Mobi device (TMSI) with self-adhesive electrodes. The electrodes were placed across the heart, one below the participants right collar-bone, the other left below the chest, with a grounding electrode below the left collar-bone. Some preprocessing was programed into the recording software Portilab2. To record skin conductance, Mobi was used with finger electrodes. These were fixed with Velcro tape around the lower part of the thumb and ring finger of the participant's non-dominant hand. Recording frequency was 2048 Hz. All signals (ECG and skin conductance, raw and preprocessed) were stored together in S00-files.

Additional Lab Recordings. The lab's ceiling camera and microphones were used for making records of the lab during the whole experiment, as well as a screen capture of the participant's screen. The video files are encoded in AVI-format with a codec specific to the labs recording software (GeoVision's CCS5). Audio is encoded in separate wav-files.

3.7 Participants

25 students participated in our experiment, of which 8 were female and 17 male. The average age was 25 (standard deviation 3.25). Most participants were native Dutch. They were interns from TNO and students from Delft University of Technology who were approached by advertising. Since these interns and students are experienced in handling (large amounts of) information for their courses, and often use computers as their most important tool, they are assumed to be representative of knowledge workers. Additionally, they are experienced with the knowledge worker tasks we have chosen: writing reports and preparing presen-

tations. The participants received a standard subject fee for their participation in the experiment.

To assess whether the participants worked on the tasks seriously, we checked the quality of the written reports and presentations. As the quality was satisfactory, none of the subjects needed to be excluded from the corpus. Of the participants, 2 were left handed and 8 wore glasses (which could be of concern for the software analyzing facial expressions). Results of our pre-questionnaire showed that none of the participants indicated to have a heart disease or take medicine which could have influenced their heart rate. About half the participants indicated that they were physically active before the experiment as they came by bike. 4 participants indicated that they had experienced stress prior to the experiment. None of the participants smoked, drank caffeine or alcohol 3 hours prior to the experiment. The participants scored on average 3.67 on on the internal control index (scale from 1 to 5, with higher scores indicating more internal control; stdv = 0.29).

4. DATASET

In this section we present the public SWELL-KW dataset in more detail. We collected data from the following sensors: computer logging, video, Kinect 3D and body sensors. Handling this data requires expertise in different fields. We preprocessed this data to get an aggregation of computer interactions, extraction of facial expressions, postures, heart rates and skin conductance levels. We finally aggregated this data into features per minute. For an overview of all available data see Table 1. We now fist describe the available fully preprocessed and aggregated feature data. Then we describe the available raw data and preprocessing.

4.1 The Feature Data

The feature dataset contains our completely preprocessed data, aggregated per minute, for all 25 participants. It contains the following features: 12 computer interaction features, 40 facial expression features, 88 body posture features and 3 physiology features as listed in the right column of Table 1. The feature dataset is annotated with the conditions under which the data was collected. Per participant three times 6 minutes relaxation data are included, ca. 45 minutes of working under normal conditions, ca. 45 minutes working with email interruptions and ca. 30 minutes working under time pressure.

Moreover, we provide the scores on our questionnaire items as ground truth for the subjective experience in each condition, see Table 3. As 25 participants each rated 3 conditions, this yields 75 ratings in total.

4.2 The Raw Data and Preprocessing

Besides the completely preprocessed and aggregated data, we also provide some raw data and files resulting from our preprocessing, as listed in the middle column of Table 1.

Computer logging. The computer logging software recorded detailed timestamped information in XML format about each computer event. Examples of computer events are mouse clicks, mouse scrolls and application changes. Moreover we parsed the files and printed them in a more intelligible timestamped table format, which will also be made available. Finally, we computed several relevant mouse, keyboard and application characteristics per minute (listed in Table 2), which are contained in the feature dataset.

Table 2: Computer interaction features (aggregated per minute).

Type	Feature	Description
Mouse	MouseActivity	Number of all MouseEvents
	LeftClicks	Number of left clicks
	MouseWheel	Number of mouse wheel scrolling
Keyboard	KeyStrokes	Number of all KeyEvents
	ShortcutKeys	Number of shortcut keys (Ctrl+ c/x/v/z/s/a; Shift+Tab)
	DirectionKeys	Number of direction keys (arrow left/right/up/down)
	Characters	Number of characters (a-z)
	CharactersRatio	#characters devided by #keyStrokes
	ErrorKeys	Number of error keys (Backspace, Delete, Ctrl+Z)
	ErrorKeyRatio	#errorKeys devided by (#characters + #spaces)
Applications	AppChanges	Number of application changes
	TabfocusChange	Number of tab focus changes

Facial expressions from video. We do not include the fully recorded videos in our dataset to keep our participants anonymous. Instead, we provide data files with the analysis of facial activation. These were extracted from the video per timeframe using the software FaceReader. The characteristics that are included in the dataset are: quality, estimates on the orientation of the head, some global features of the face like looking direction and the amount of activation in several facial action units. Moreover, FaceReader provides an estimate of the subjects emotion, which is also available in our dataset. We parsed these files to get a more intelligible timestamped table format, which will also be made available. Besides data per video frame, we also calculated averages per minute for all characteristics (see Table 1), which are contained in the feature dataset.

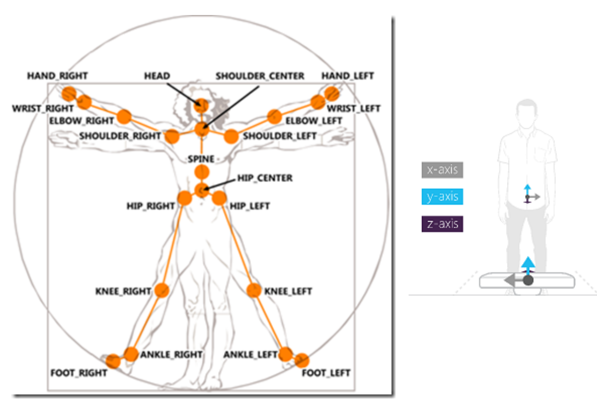

Figure 3: Kinect joints and Kinect orientation from Kinect SDK. Images from: http://msdn.microsoft.com

[2]`http://www.noldus.com/human-behavior-research/products/ulog`
[3]`http://www.noldus.com/human-behavior-research/products/facereader`
[4]`http://www.tmsi.com/en/products/mobi.html`

Table 1: Our dataset contains data from 25 participants (3 hours each). The listed raw and preprocessed sensor data, as well as a feature dataset (aggregated per minute) will be made available.

Type	Available raw and preprocessed data	Available features (#features)
Computer interactions	uLog output[2] (i.e. xml-logs of all computer events) Parsed data (i.e. txt-file with timestamped data)	Mouse (3) Keyboard (7) Applications (2) (for details see Table 2)
Facial expressions	FaceReader output[3] (i.e. txt-logs with facial information and emotions) Parsed data (i.e. txt-file with timestamped data)	Head orientation (3) Facial movements (10) Action Units (19) Emotion (8)
Body postures	Joint coordinates extracted with Kinect SDK (i.e. txt-file with timestamped data) Angles of the upper body (i.e. txt-file with timestamped data)	Distance (1) Joint angles (10) Bone orientations (3x11)) (as well as stdv of the above for amount of movement (44))
Physiology	Data from Mobi[4] (i.e. S00-files with raw and filtered signals)	Heart rate (variability) (2) Skin conductance (1)

Body postures from Kinect 3D sensor. We do not include the recorded 3D Kinect files in our dataset to keep our participants anonymous. Instead, we provide data files with analysis of the participant's body posture per time-frame. These were extracted from the 3D Kinect recordings using the Kinect SDK. By fitting the Kinect skeletal model (see Figure 3, left), we got coordinates of all body joints per frame. This data will be made available. We further used these joint coordinates to determine joint angles between bones of the upper body, for example the angle between the upper and lower arm. Moreover, we determined bone orientations of the upper body relative to the x, y and z axis (see Figure 3, right), for example the angle between the left shoulder and the up pointing y axis.[5] This information on angles per frame will be made available. From the depth image the average distance of the user was also determined. Finally, we determined average angles per minute, which are contained in the feature dataset. We also calculated standard deviations for each minute, to determine features that indicate the amount of movement and changes in joint angles. These are also contained in the feature dataset.

Physiology from body sensors. We provide raw and preprocessed ECG data. The raw ECG signal was filtered as described in the TMSI[6] manual: First a high pass filter (8Hz) was applied to filter out large fluctuations in the signal. A 15ms second delay was added, together with a delta filter to let the low frequency parts of the signal disappear. To be independent of the direction of the QRS complex (due to morphology of the ECG), we took the absolute signal. Finally, a moving window averager (0.1sec) was added to get the envelope of the signal. This yielded a filtered signal with clear peaks. The raw and preprocessed ECG data will be made available.

We also calculated the heart rate and heart rate variability. Therefore, we processed the filtered data further in Matlab. First of all we applied a peak detection algorithm to the filtered signal. To determine the heart rate, the found peaks were counted per 1 minute time-frame. Then we calculated the distance between the found peaks (R-R). To determine the heart rate variability we took the root mean square of all these peak distances (RMSSD). Due to some remaining noise in the signal, the peak finding algorithm sometimes failed to accurately detect peaks. Therefore we excluded all 1-minute time frames in which more than one peak distance appeared unusual. We defined an unusual peak distance as a distance larger than 1.2 seconds where probably a peak was missed (or otherwise the HR would be below 50bpm) or a distance smaller than 0.5 seconds where probably an extra peak was detected (or otherwise the HR would be over 120bpm). The resulting heart rate and heart rate variability are contained in the feature dataset.

Moreover, we provide raw skin conductance data. We also calculated the average skin conductance level by averaging the raw signal per minute, which is contained in the feature dataset.

5. EXAMPLE ANALYSES

In this section we present some research that was done based on our dataset, as an example of its use.

Work stress. To find relations between the measured concepts, we performed a correlation analysis on the questionnaire data. We found that perceived stress is moderately related to high task load in terms of mental demand, temporal demand and frustration. Moreover, stress is related to emotion in terms of negative valence and high arousal. For more details on these results, see [9].

To investigate the effect of our stressors on the participants' subjective experience, we compared the questionnaire ratings of the neutral baseline condition with the time pressure and email interruption conditions (see Table 3). Under the stressor time pressure, participants experienced significantly higher temporal demand and higher arousal. The stressor email interruptions yielded reports of more mental effort, more positive valence and more dominance. We found that perceived stress did not differ significantly between the stressor and neutral conditions. Stress might be a too complex concept to measure in a short-termed work task. For more details on our results, see [9]. These analyses show the potential of using the dataset for research on the effect of

[5]We use a projection to the plane to distill only variance in one direction.

[6]`http://www.tmsi.com`

Table 3: Subjective experience data (one rating per block). Average values for the Neutral, Interruption and Time pressure condition can be found in the last 3 columns.

Type	Feature	Description	N	I	T
TaskLoad (NASA-TLX)	MentalDemand (0: low - 10: high)	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)?	4.9	5.4	4.9
	PhysicalDemand (0: low - 10: high)	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)?	1.9	2.3	2.7
	TemporalDemand (0: low - 10: high)	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred?	5.7	5.9	7.1
	Effort (0: low - 10: high)	How hard did you have to work (mentally and physically) to accomplish your level of performance?	5.2	5.9	6.1
	Performance (0: poor - 10: good)	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?	4.8	6.1	6.0
	Frustration (0: low - 10: high)	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	3.5	3.6	3.5
Mental Effort (RSME)	MentalEffort (0: no - 10: extreme effort)	How high was the mental effort for the tasks you just finished?	5.5	6.5	6.3
Emotion (SAM)	Valence (1 - 9)	How do you feel at this moment? (unhappy - happy)	4.8	5.7	5.3
	Arousal (1 - 9)	How do you feel at this moment? (calm - excited)	3.3	3.9	4.6
	Dominance (1 - 9)	How do you feel at this moment? (submissive - dominant)	5.2	6.2	5.9
Stress (VAS)	Stress (0: not - 10: very stressed)	How stressed do you feel?	2.9	3.2	3.8

work stressors on experience and behavior.

User modeling. We are aiming to develop algorithms that can estimate the level of workload and stress that a knowledge worker is experiencing from sensor data, in order to unobtrusively model their mental state. To investigate whether the stressors affected the participants' behavior, we compared computer interactions and facial expressions in the two stressor conditions with the neutral baseline condition. Under time pressure we see significantly more key strokes than in the neutral condition, and under interruptions we see more application changes and left clicks. So both stressors create typical behavioral patterns.

Explorative correlation analysis on questionnaire and facial expression data showed that moderate correlations were found for mental effort with several facial features. When working in a condition with higher mental effort, participants looked more disgusted and sad and they showed more activation in the facial action units LidTightener, UpperLipRaiser, BrowLowerer and CheekRaiser. So mental effort might be estimated based upon video information. For more details on our results, see [9]. These analyses show the potential of using the dataset for research on user modeling.

Context recognition and Information support. The tasks of writing reports and preparing presentations enabled us to use the dataset in the fields of context recognition and search behavior as well. For this purpose, the raw events were aggregated in event blocks, which were labeled for their task content. Details can be found in [17], where also some initial analyses of the search behavior in the data are described.

6. DISCUSSION

To our knowledge, our dataset is the first in which a set of unobtrusive sensors from different research fields was used to collect data in a realistic office context, while stressors were manipulated.

In collecting and preprocessing the data, we encountered a number of challenges. First, simulating a realistic work setting and inducing stress was challenging. We do think that we succeeded in simulating a realistic work setting. Some participants noted afterwards, that although they knew that the emails were fake, they felt responsible to reply. We also think we were able to manipulate the working conditions with stressors: the questionnaire data showed that participants' experience of task load, mental effort and emotion changed in the stressor conditions. Due to our experimental design of always starting with the neutral baseline condition, subjects might have experienced order or fatigue effects. We do, however, think that the relaxation phases helped participants to start each condition in a well-rested state. We did not find significant effects of our working conditions on perceived stress. Real-world stress might be complex, involving worries or thing outside work and stress building up over days. Therefore, a limitation of this dataset is that only short term effects of stressors can be investigated. For longitudinal research on temporal (stress) patterns over days or weeks, we are currently recording the presented sensor suite in a real-world office.

A second challenge was synchronization of all data. Different sensors were recorded via different computers. We synced computer clocks, and most sensors made exact starting timestamps upon hitting the record button. Nevertheless, we cannot guarantee second-precise synchronization among modalities (especially the uEye camera start times may be somewhat unprecise).

Finally, using different sorts of sensors requires multidisciplinary expertise, like knowledge (and software) for processing physiological, image or Kinect data. Our contribution is to provide a dataset that not only contains raw data, but also preprocessed and aggregated data, which makes it easier for other researchers to use the data.

The strength of this dataset is its richness in terms of modalities and its size in terms of the amount of data per participant. Although limited in size of participants (25), initial experiments have shown that the dataset has sufficient power to detect significant differences. With the presented new dataset, automatic inference of a rich set of context information of a user in the office can be studied. It can be used to develop context aware systems to support knowledge workers during their work. Moreover, it provides ample resources for stress and work style related studies.

7. CONCLUSION

We identified the need of a rich dataset and its desired characteristics. In this paper we described how we collected such a new dataset that overcomes drawbacks common in related work: We used a realistic office setting while stressors were manipulated systematically. We used a varied set of sensors: computer logging, video, Kinect 3D and body sensors. We preprocessed the data and extracted features per minute. The resulting dataset SWELL-KW will be shared with the scientific community. We presented a selection of research questions that could be answered with this dataset. As demonstrated, analyses of the data can yield insights in the effects of stressors at work, or on the relation between subjective ratings and the sensor data. The presented new affective and behavioral dataset is a valuable contribution to research fields like work psychology, user modeling and context aware systems.

More information on the dataset and its access can be found at `http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-kwrv-3e`

8. ACKNOWLEDGEMENTS

This publication was supported by the Dutch national program COMMIT (project P7 SWELL).

9. REFERENCES

[1] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[2] E. Demerouti, A. B. Bakker, S. A. Geurts, and T. W. Taris. Daily recovery from work-related effort during non-work time. *Research in Occupational Stress and Well-being*, 7:85–123, 2009.

[3] E. Demerouti, A. B. Bakker, F. Nachreiner, and W. B. Schaufeli. The job demands-resources model of burnout. *Journal of Applied psychology*, 86(3):499, 2001.

[4] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, 76(Supplement 1):B172–B182, 2005.

[5] P. C. Duttweiler. The internal control index: A newly developed measure of locus of control. *Educational and Psychological Measurement*, 44(2):209–221, 1984.

[6] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[7] A. Kapoor and R. W. Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682. ACM, 2005.

[8] S. Koldijk. Automatic recognition of context and stress to support knowledge workers. In *Proceedings of the 30th European Conference on Cognitive Ergonomics*, pages D20–D23. ACM, 2012.

[9] S. Koldijk, M. Sappelli, M. Neerincx, and W. Kraaij. Unobtrusive monitoring of knowledge workers for stress self-regulation. In *User Modeling, Adaptation, and Personalization*, pages 335–337. Springer, 2013.

[10] S. Koldijk, M. van Staalduinen, M. Neerincx, and W. Kraaij. Real-time task recognition based on knowledge workers' computer activities. In *Proceedings of the 30th European Conference on Cognitive Ergonomics*, pages 152–159. ACM, 2012.

[11] L. Koppes, E. d. Vroome, G. Mars, B. Janssen, M. Zwieten, and S. v. d. Bossche. Nationale enquete arbeidsomstandigheden 2012: Methodologie en globale resultaten, 2012.

[12] K. Kraan, S. Dhondt, I. Houtman, R. Nelemans, and E. d. Vroome. Handleiding nova-weba: Hernieuwde versie., 2000.

[13] G. Mark, D. Gudith, and U. Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110. ACM, 2008.

[14] F. Mokhayeri, M.-R. Akbarzadeh-T, and S. Toosizadeh. Mental stress detection using physiological signals based on soft computing techniques. In *Biomedical Engineering (ICBME), 2011 18th Iranian Conference of*, pages 232–237. IEEE, 2011.

[15] A. Riera, A. Soria-Frisch, A. Albajes-Eizagirre, P. Cipresso, C. Grau, S. Dunne, G. Ruffini, and S. aStarlab Barcelona. Electro-physiological data fusion for stress detection. *Studies in health technology and informatics*, 181:228–32, 2012.

[16] J. Ruff. Information overload: Causes, symptoms and solutions. *Harvard Graduate School of Education*, pages 1–13, 2002.

[17] M. Sappelli, S. Verberne, S. Koldijk, and W. Kraaij. Collecting a dataset of information behaviour in context. In *Proceedings of the 4th Workshop on Context-awareness in Retrieval and Recommendation*, 2014.

[18] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, 3(1):42–55, 2012.

[19] L. M. Vizer, L. Zhou, and A. Sears. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009.

[20] D. Zapf. Stress-oriented analysis of computerized office work. *The European Work and Organizational Psychologist*, 3(2):85–100, 1993.

[21] F. Zijlstra and L. van Doorn. The construction of a scale to measure subjective effort. Technical report, Delft University of Technology, 1985.