

The SWELL Knowledge Work Dataset for Stress and User Modeling Research

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |



This publication was supported by the Dutch national program COMMIT.

SWELL Partners:

Almende, Noldus, Novay, Philips, TNO, Radboud Universiteit Nijmegen, Roessingh Research and Development, Sense-OS, Universiteit Twente,

This document presents 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. Our dataset 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 suitable for several research fields, such as work psychology, user modeling and context aware systems.   
  
The SWELL-KW dataset was collected within the [SWELL project](http://www.swell-project.net/). The collection of this dataset was supported by the Dutch national program [COMMIT (project P7 SWELL)](http://www.commit-nl.nl/projects/swell-smart-reasoning-systems-for-well-being-at-work-and-at-home).

The dataset is suited for two types of research:

## Behavioral research:

The SWELL-KW dataset contains data from 25 participants (~3 hours each), for working under 3 conditions: neutral, interruptions and time pressure (plus a relax phase). The participant information and order of conditions can be found in ‘SWELL-KW - overview available data.xlsx’. The listed raw and pre-processed sensor data, as well as a feature dataset (aggregated per minute) are available:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **directory** | **0-Raw data** | **1-Sequential data** | **2-Minute data** | **3-Feature dataset(#features)** |
| **0 - Subjective experience -** | Questionnaire start  Questionnaire condition[[1]](#footnote-1) |  |  |  |
| **A - Computer interaction** | uLog xml-files |  | uLog per condition, per participant aggregated per minute | Mouse (3)  Keyboard (7)  Applications (2) |
| **B - Facial expressions** | Facereader detailed-log |  | Facereader per condition, per participant aggregated per minute | Head orientation (3)  Facial movements (10)  Action Units (19)  Emotion (8) |
| **C - Body Postures** | Kinect - Joint coordinates | Kinect - Joint angles | Kinect – per condition, per participant aggregated per minute | Distance (1)  Joint angles (10)  Bone orientations (3x11)  (as well as stdv of the above for amount of movement (44)) |
| **D - Physiology** | Mobi-data (s00-files) |  | Mobi – aggregated per minute | Heart rate (variability) (2)  Skin conductance (1) |

Besides sensor data, we provide the questionnaire ratings of the participants on task load (NASA-TLX), mental effort (RSME), emotion (SAM) and perceived stress for each working condition.

The full feature-dataset can be found in ‘Behavioral-features - per minute.xlsx’

When using the SWELL-KW dataset for behavioral research, please refer to this paper:

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M., & Kraaij, W. (2014). **The SWELL Knowledge Work Dataset for Stress and User Modeling Research**. To appear in: *Proceedings of the 16th ACM International Conference on Multimodal Interaction (ICMI 2014)* (Istanbul, Turkey, 12-16 November 2014). [[pdf]](http://cs.ru.nl/~skoldijk/Papers/ICMI%202014%20paper_final_cr.pdf)

## Information Science Research

Part of the SWELL-KW dataset is suited for research into information behavior in context. For this dataset we focus on the computer interaction features, collected from the uLog keylogger. The participants were asked to write reports, prepare presentations and answer e-mails for a total of 8 topical tasks (see ‘Participant\_Instructions.doc’). The raw uLog interaction events are combined with the file access history (IeHistoryView). Moreover the events are aggregated to event blocks. Events collected within the same active window (application or tab of application) belong to the same event block. These event blocks are labeled with the topical task that the participant was most likely executing.

The labeled event block dataset (‘Content-features - Labeled-EventBlocks.csv’) consists of the following features:

* ActualLabel : topical task assigned by annotator
* Certainty: certainty of annotator about label (scale of 1-5 with 5 being very certain)
* Id: identification number of event block
* Person: participant identifier
* BeginTime: begin timestamp
* end time: end timestamp
* duration in seconds
* clicks: number of mouseclicks in block
* key: concatenation of key events in block (key combinations and special keys are represented by a space)
* application: active application
* windowtext: visible windowtext
* caption: data collected from captions of interface elements (for example the caption of a button that is pressed)
* url: active url
* domain: domain of the url
* query: the query that was executed (in case of a google-url)

Furthermore the following data is available:

* directory *0 - Raw data\A - anonimized Computer interaction - raw data uLog*
  + directory *\Written Documents and File History*
    - The documents and presentations produced by each participant
    - The complete file access history collected with IeHistoryView (History\_pp#.txt)
  + The complete log of interactions collected with uLog (a\_pp(articipant)#\_c(ondition)#\_uLog\_\*date\*\_\*time\*.xml)
* directory *Provided documents* : the documents that were on the pc of the participant at the start of the experiment

When using the SWELL-KW dataset for information science research, please refer to this paper:

Maya Sappelli, Suzan Verberne, Saskia Koldijk, Wessel Kraaij(2014) **Collecting a dataset of information behaviour in context**. In: *Proceedings of the 4th Workshop on Context-awareness in Retrieval and Recommendation (CARR @ ECIR 2014)* .[[pdf](http://sverberne.ruhosting.nl/papers/CARR2014_SWELL2.pdf)]

1. Task load (NASA-TLX), Mental effort (RSME), Emotion (SAM) and Perceived stress [↑](#footnote-ref-1)