

CLAS – Cognitive Load, Affect and Stress Database

Dear user,

This document describes the structure of the CLAS database and provides information about the purpose, recording procedure, stimuli used and other supporting information accompanying the database.

The CLAS dataset was created in support of RTD activities focusing on the development of functionalities related to the automated recognition of specific human emotions, detection of stress-related conditions and negative emotional states, as well as the automated assessment of the degree of attention and concentration, cognitive load and momentary cognitive capacity. However, the dataset design has the potential to support RTD activities of a wider scope, mainly in human-robot collaboration scenarios and the development of advanced human-machine interfaces that are aware of intrinsic human aspects, associated with communication, productivity, and efficiency of collaboration.

In brief, each of the 62 persons recorded in the dataset is represented with a 30-minute recording of physiological signals (ECG, PPG and EDA), 3D accelerometer data, along with the corresponding tags and metadata. The dataset consists of five subsets, which correspond to the recordings collected while the volunteer participants were performing the abovementioned perceptive and interactive tasks. For detailed information please read the full paper accompanying the database.

If you decide to use the database in your RTD activities, please, do not hesitate to cite the database description paper:

V. Markova, T. Ganchev and K. Kalinkov, "CLAS: A Database for Cognitive Load, Affect and Stress Recognition," 2019 International Conference on Biomedical Innovations and Applications (BIA), Varna, Bulgaria, 2019, pp. 1-4. DOI: 10.1109/BIA48344.2019.8967457

URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8967457&isnumber=8967447>

1. Database structure (CLAS.zip)

The database is organized in 4 folders:

- *Answers* – answers of the questions in the interactive tasks (Math problems, Logic problems and the Stroop test) for each person.
- *Block_details* – metadata for each block (1 block per task) for every participant.
- *Data* – raw signal recordings for the individual participants.
- *Documentation* – accompanying documents.

A detailed tree of the folders is shown in Figure 1.

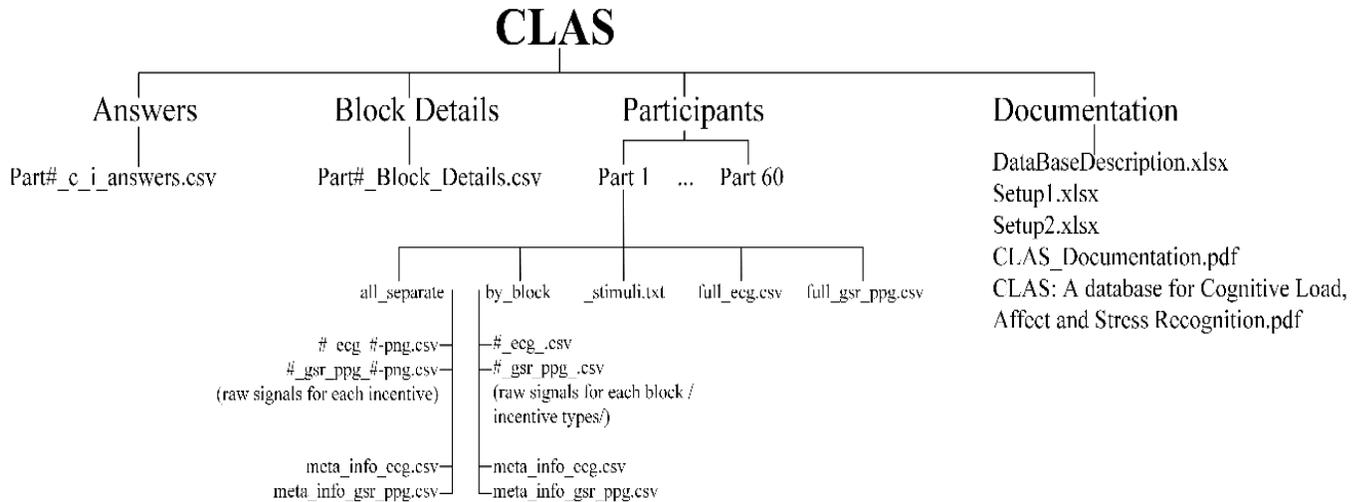


Figure 1 – Structure of the CLAS database

2. File description

- **Part#_c_i_answers.csv** – provides information for the correctness of answers per participant N (#).
- **Part#_Block_Details.csv** – metadata for each block (type, files, length and signal quality).
- **#_ecg_#-png.csv** – timestamps and raw data for the ecg2 channel (Shimmer ECG Unit). The ecg1 channel is not used. The files are separated per stimuli.
- **#_gsr_ppg_#-png.csv** – timestamps, raw signals of GSR, PPG and three axis accelerometers. The data from the accelerometers is not calibrated. Calibration information can be found in the Shimmer GSR Unit documentation. The files are separated per stimuli.
- **#_ecg_.csv** - timestamps and raw data per block for the ecg2 channel (Shimmer ECG Unit). The ecg1 channel is not used.
- **#_gsr_ppg_.csv** - timestamps, raw data per block of GSR, PPG and three-axis accelerometers. The data for the accelerometers is not calibrated. Calibration information can be found in the shimmer sensor documentation.
- **meta_info_ecg.csv** – information for the stimuli group, stimuli names, types, tags and the filename with the raw ECG signal.
- **meta_info_gsr_ppg.csv** - information for the stimuli group, stimuli names, types, tags and the filename with the raw PPG and GSR signal.
- **DataBaseDescription.xlsx** – information for the database structure, file structures, stimuli types and source.
- **Setup1.xlsx** – describes the block types, their number and length in seconds for Participants 1 to 11.
- **Setup2.xlsx** - describes the block types, their number and length in seconds for Participants 12 to 60.

- **CLAS_Documentation.pdf** – the current document.
- **CLAS: A database for Cognitive Load, Affect and Stress Recognition.pdf** – accompanying BIA-2019 paper with a description of the database.

3. Recording procedure and signals

All participants were briefly introduced with the purpose of this data collection campaign and then presented with additional information to read. Before the beginning of the recording procedure, each participant signed a consent form and filled out a questionnaire. The questionnaire collects general information about their current health status, sleeping habits, the frequency of use and recent intakes of drugs, alcohol, tobacco, beverages containing caffeine and other stimulants of body and mind. Next, the participants were provided with a set of instructions explaining the experimental protocol and the self-assessment questions.

The CLAS data collection campaign relied on a well-established protocol, which makes use of stimuli purposely selected to evoke emotions in the 4 quadrants of the valence-arousal space. The recording procedure is illustrated at Figure 2.

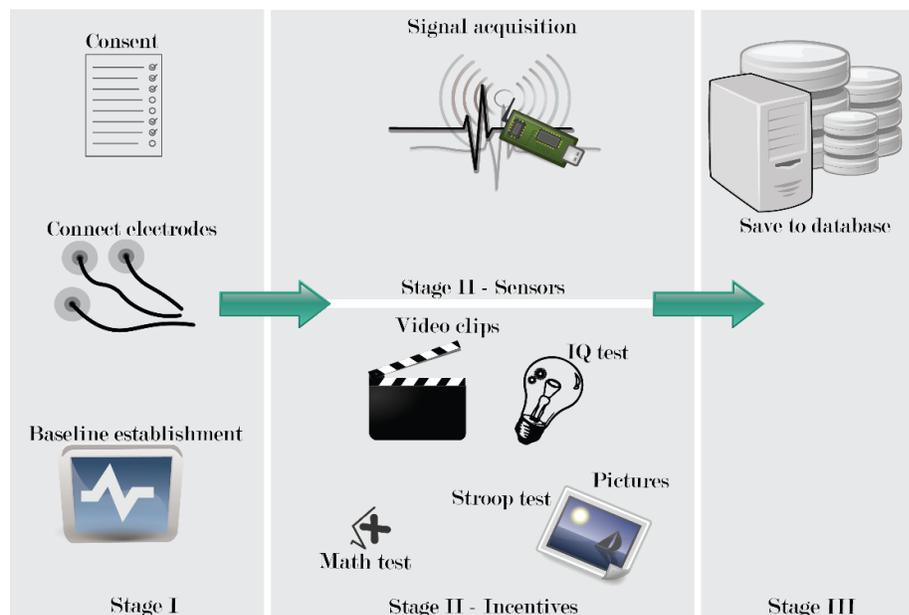


Figure 2 – Experimental Protocol for the creation of the CLAS database

The dataset consists of recordings of physiological signals and accelerometer data:

1. ECG recordings.
2. PPG recordings.
3. EDA recordings.
4. Three-axis accelerometer data logs.

4. Stimuli

Six groups of stimuli were used in the creation of the CLAS database: neutral stimuli, three types of stimuli sequences used in the interactive tasks (Math problems, Logic problems, Stroop test), and two types of stimuli sequences used in the perceptive tasks (emotional pictures and videos). These six groups of stimuli can be characterized as:

- 1) The first group consists of neutral stimuli. It aims to establish the baseline and to relax the participant in-between the different perceptive and interactive tasks.
- 2) The second group consists of stimuli that comprise the Math problems test. The Math problems test sets limited time to answer of simple mathematical problems.
- 3) The third group consists of the stimuli for the Stroop test. It provides limited time for the user to match the text color with the meaning of the word.
- 4) The fourth group consists of the stimuli in the Logic problems test, where the participant is expected to solve simple logical problem similar to those used in the IQ tests for a short time.
- 5) The fifth and sixth sequences of stimuli do not require interaction, as the participant is only required to watch pictures or watch and listen multimedia clips. The pictures used in the creation of CLAS database are taken from the International Affective Picture System (IAPS) dataset.

5. Volunteer's statistics

In total, we recorded 62 healthy volunteers, which were recruited without special care for gender balance. Among these are 17 women and 45 men, including 1 woman and 7 men left-handed.

Most of the recruited people were students in their twenties (20-27 years old). Still, in the database there is one person in his thirties, one in her late forties and one man who is 50 years old.

6. License

The CLAS dataset is available free of charge for academic use¹ under the provided EULA. Please contact Assoc. Prof. Valentina Markova (via@tu-varna.bg) for queries and orders.

¹The CLAS dataset is available at the web-site of Sensor Networks Lab: <https://www.sensornetworkslab.com/database/CLASdataset>