CLAS: A Database for Cognitive Load, Affect and Stress Recognition

Valentina Markova
Department of Communication Engineering and Technologies
Technical University of Varna
Varna, Bulgaria
via@tu-varna.bg

Todor Ganchev
Department of Computer Science and Engineering
Technical University of Varna
Varna, Bulgaria
tganchev@tu-varna.bg

Kalin Kalinkov
Department of Communication Engineering and Technologies
Technical University of Varna
Varna, Bulgaria
k.kalinkov@mail.bg

***Abstract*— We present the overall design and the implementation of the CLAS dataset, a multimodal resource which was purposely developed in support of research and technology development (RTD) activities oriented towards the automated recognition of some specific states of mind. Although the particular focus of our research is on the states of mind associated with negative emotions, mental strain and high cognitive effort, the CLAS dataset could offer an adequate support to research of a wider scope, such as general studies on attention assessment, cognitive load assessment, emotion recognition, as well as stress detection. The dataset consists of synchronized recordings of physiological signals, such as Electrocardiography (ECG), Plethysmography (PPG), ElectroDermal Activity (EDA), as well as accelerometer data, and metadata of 62 healthy volunteers, which were recorded while involved in three interactive tasks and two perceptive tasks. The interactive tasks aim to elicit different types of cognitive effort and included solving sequences of *Math problems*, *Logic problems* and the *Stroop test*. The perceptive tasks make use of images and audio-video stimuli, purposely selected to evoke emotions in the four quadrants of the arousal-valence space. The joint analysis of success rates in the interactive tasks and the information acquired through the questionnaire and the physiological recordings enables for a multifaceted evaluation of specific states of mind. These results are important for the advancement of research on efficient human-robot collaborations and general research on intelligent human-machine interaction interfaces.**

***Keywords*— HMI interfaces, efficient human-robot collaboration, state of mind, emotions, physiological signals.**

I. Introduction

The advances in computer science and technology made possible the emergency of automated information services and handy applications that ultimately transformed our daily routine and continuously redefine the way we are working and interacting with each other. Numerous studies showed that rigid and old-fashioned designs of HMI interfaces contribute to the accumulation of extra fatigue and stress at the work place and at home. Therefore, the development of intelligent human-machine interaction (HMI) interfaces, which are human centered, adaptive and permit natural and unobtrusive interaction with the user brings numerous benefits to the efficiency of work and the overall wellness of people.

In this regard, numerous studies assess the feasibility of multimodal interfaces that are capable of human intention and behavior interpretation by monitoring the facial expressions, gestures, speech or physiological signals of the user. When concerning the physiological signals, plenty of research efforts were already invested in the combined analysis of the Electrical activity of the heart (ECG), Electrodermal activity (EDA), Respiration Activity, Electrical activity of the brain (EEG), Skin temperature etc. in order to develop functionality for emotion recognition, cognitive load assessment, stress detection etc. In order to support RTD activities aimed at the development of such functionalities, the research community continuously invests time and efforts aimed at the development of appropriate resources. Among these are DEAP [1], MAHNOB-HCI [2], DECAF [3], ASCERTAIN [4], SLADE [5] and other dataset, which consist of recordings of physiological signals, collected while healthy subjects were watching purposely selected audio-visual stimuli. These stimuli aim to evoke physiological responses, which are next evaluated along the arousal, valence, and dominance dimensions. Furthermore, a number of other resources were created in support of RTD activities aiming at cognitive load assessment, attention assessment and other aspects of cognition [6-9]. The CLAS dataset presented here complements these resources with an HMI-oriented setup which can support a wider range of applications.

In the present paper, we outline the design and the development of the CLAS dataset, a new resource that was purposely created in support of research on the automated recognition of certain states of mind, associated with negative emotions, mental strain and high cognitive effort. Furthermore, the CLAS dataset could also offer an adequate support to research of a wider scope, such as general studies on attention assessment, cognitive load assessment, emotion recognition, and stress detection. In brief, the CLAS dataset consists of time-synchronized recordings of physiological signals, such as ECG, PPG, and EDA, as well as accelerometer data, and metadata. The CLAS dataset differs from other available resources, such as the speech and video resources in the AVEC-2019 dataset [10], mainly in the scope of RTD tasks it could support. In Section II, we outline the conceptual design of the CLAS dataset, and in Section III provide a description of the data collection setup. A concise statistical description of the dataset is offered in Section IV. Indicative baseline results for several application-oriented tasks are discussed in Section V. Finally, in Section VI we briefly summarize this work and provide a brief note on the utility and limitations of the CLAS dataset and its potential areas of use.

II. Conceptual Design

The CLAS dataset was conceived as a repository which is purposely developed to support research on the automated assessment of certain states of mind and the emotional condition of a person. This resource is intended to support RTD activities aiming at the development of intelligent human computer interaction (HCI) interfaces that incorporate functionalities allowing for the automated recognition of human emotions, the automated detection of stress conditions, the automated assessment of the degree of concentration, cognitive load, and momentary cognitive capacity, and can account for some personality traits related to the ability to quickly solve logical and mathematical

978-1-7281-4754-3/19/$31.00 ©2019 IEEE

problems under strict time constraints. These and other related functionalities open additional opportunities in support of the advancement in research on: (i) intelligent human-machine interfaces in entertainment, e-Health, e-Learning, e-Government and other applications; (ii) improving the efficiency of human-robot collaboration systems; (iii) risk assessment and management systems in robotics and in high-risk professions (rescue crews, military, miners etc.); and (iv) a wide range of personal applications aiming at the enhancement of health-related quality of life and wellbeing.

The CLAS dataset is a resource of labeled physiological recordings, which contain synchronized ECG, PPG, and EDA signals captured while the test subjects are involved in some purposely designed *interactive* or *perceptive* task. Specifically, the *interactive* tasks are designed to evaluate different aspects of the momentary cognitive load, the degree of attention and concentration, or the cognitive capacity of a person. In order to obtain a quantitative and qualitative assessment of these aspects, the test subject is presented with a set of purposely designed mathematical and logical problems with the strict requirement for prompt response within a short time window. In the *perceptive* tasks, the test subjects are exposed to a series of carefully selected stimuli which aim to elicit emotions. In the *perceptive* tasks the test subjects typically receive instructions to simply watch the images or to follow the video clips. The emotional conditions and the states of mind elicited during the various interactive and perceptive tasks are captured in the CLAS dataset by means of the variations in the physiological signals of each person. The dataset design permits both person-specific and person-independent modeling.

In Fig. 1, we illustrate the conceptual design of the perceptive subset of the CLAS dataset. This design implements the controlled exposure of each person to the 4 groups of stimuli representing the quadrants of the arousal-valence emotion space. Each group consists of 4 stimuli with identical tags, as such a design provides the minimum number of stimuli required for the subsequent data modeling and classification stages. These stimuli tags are meant as person-independent ground-true labels for the corresponding segments of the ECG, PPG, and EDA recordings in the perceptive tasks.

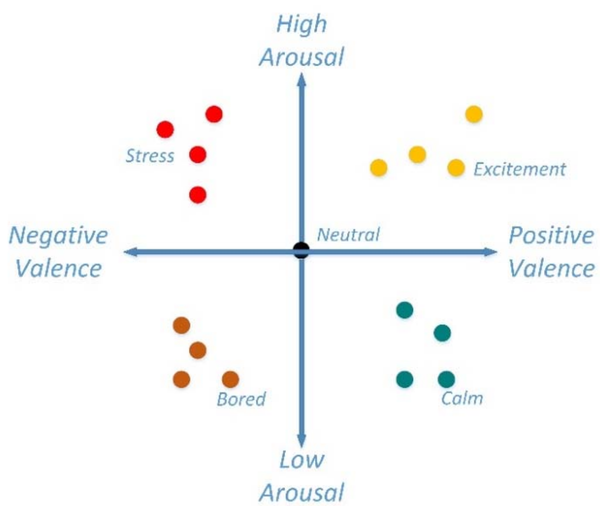

Fig. 1. A conceptual illustration of the stimuli distribution in the four quadrants of the arousal-valence emotion space.

In Fig.2 we illustrate the conceptual scheme of the method we use for an indirect assessment of the momentary state of mind of a person. This method relies on the concentration-capacity space in order to evaluate the concentration, the cognitive load, the success rate of each person by means of some purposely designed interactive tasks. In the CLAS dataset the focus is on the evaluation of the level of concentration, the cognitive capacity and certain personality traits, which are related to the ability of a person to quickly solve logical and mathematical problems under strict time constraint and psychological pressure. Therefore, we rely on series of low-complexity logical or mathematical problems which a person with average Intelligence quotient (IQ) scores and basic level of math skills can solve easily when time is unrestricted. However, here we allow short time for response (of only few seconds) in order to increase the cognitive load. After the response deadline expires, the person is quickly presented with the correct answer and then is immediately introduced with the next problem of the sequence.

Here we evaluate the degree of concentration indirectly based on the success rate during the *Stroop test*. The cognitive effort is assessed separately based on the success rates obtained in the *Math test* and the *Logics test*, which consist of long series of logical and math problems. The Logics test is a subset of the widely used geometric-figure similarities discovery problems, which are often used in the IQ tests. The ground true labels for the corresponding ECG, PPG, and EDA recordings are assumed “high cognitive load” during the presentation of the logic and math problems and “low cognitive load” during the neutral stimuli in the beginning of each session and in-between the tasks.

The average success scores are computed for the entire dataset (59 persons) in order to estimate and accordingly reposition there the center of the coordinate system (Fig.2). Depending on the success rates of the different persons, four (more or less) compact groups of scores are observed. The corresponding labels of the physiological recordings allow us to distinguish between people with high concentration (high score on the Stoop test), who were observed to perform low or high on the Logics and Math problems, and people who have low success on the Stroop test and perform with low or high success on the Logics and Math problems.

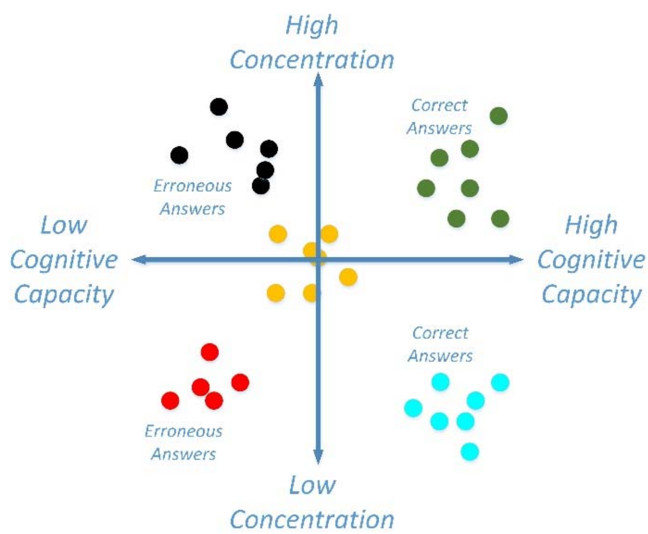

Fig. 2. Evaluation of the momentary state of mind conditions in the concentration-cognitive capacity-success rate space

III. DATA COLLECTION

The data acquisition campaign implemented the synchronized recording of physiological signals (ECG, PPG, EDA), accelerometer data and user responses in a common experimental protocol. The data collection setup was common for the interactive and perceptive tasks described in Section II. In the interactive tasks, the tags of the physiological recordings were set according to the correctness of the corresponding answers, while in the perceptive tasks we have used the stimuli tags as ground-truth labels.

A. *Data collection setup*

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 stimulators 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. These fall in two distinct categories:

- In the video clips test we made use of 16 emotionally tagged video clips retrieved from the DEAP database [1]. In our setup these are organized in 4 blocks with 4 video clips each. Between blocks a neutral stimulus (a video clip with duration of 30 sec) was played;
- In the pictures test we used 16 emotionally tagged IAPS pictures grouped in 4 blocks with 4 images each separated by neutral video stimulus (5 sec) [11]. Between blocks a neutral stimulus was presented.

The three sequences of cognitive stimuli, referred to as the Math problems test, the Stroop test and the Logic problems test, consist of various multiple-choice assignments, followed by a display with the correct answer:

- The *Math problems test* consists of a sequence of 24 assignments. Each assignment is presented for very short time (4 sec), then the person has 2 sec to select the correct answer, and afterwards the correct answer is displayed for 1 sec. Then the test continues with the next assignment of the sequence.
- The *Stoop test* consist of 30 assignments. The time for showing each assignment was 3sec, the time for answering – 2sec, and for showing the correct answer – 1sec.
- The *Logic problems test* consists of 20 assignments. The duration of each assignment was 10 sec, followed by 4 sec for answering, and 1 sec for displaying the correct answer.

The data acquisition protocol was identical to all 62 people recorded to this end. In brief, each data recording session began with a one-minute baseline recording, during which a neutral audio-visual stimulus was displayed. Next, the three interactive tests follow, different groups of stimuli were presented in blocks, separated with self-assessment questions and neutral audio-visual stimuli (30-second video-clip). The tasks are separated with neutral stimuli, in order to relax the volunteers and restore their emotional state back to neutral.

```
1) Start or recording
2) One-minute baseline recording
   (1 min emotionally neutral stimuli)
3) Math problems test
   (problem, answer, correct answer)
4) Self-assessment question
5) Display of a neutral music video clip
   (30 seconds)
6) Stroop test
   (task, answer, correct answer)
7) Self-assessment question
8) Display of a neutral music video clip
   (30 seconds)
9) Logic problems test
   (task, answer, correct answer)
10)Self-assessment question
11)Display of a neutral music video clip
   (30 seconds)
12)Pictures test
13)Display of a neutral music video clip
   (30 seconds)
14)Video clips test
15)Q: "Did you find the tasks difficult?"
16)End of recording
```

Fig. 3. The sequence of tasks in the CLAS data collection protocol

In summary, each recording session consists of sixteen steps (cf. Fig. 3), during which we collected physiological signals of the participant, and 3D accelerometer data. A detailed description of the CLAS dataset is provided in Section IV.

B. *Physiological signals and sensors*

The three types of physiological signals: ECG, PPG, and EDA were recorded by means of wearable sensors, such as the Shimmer3 GSR+ Unit and Shimmer3 ECG Unit [4,5]. The Shimmer3 GSR+ Unit measures electrodermal resistance between two Ag/AgCl electrodes attached to two fingers of one hand. In addition, the Shimmer3 GSR+ Unit has an interface to the optical pulse sensor that captures the PPG signal from the ear lobe. We used Lead I configuration of the Shimmer3 ECG Unit to measure the ECG signal. The 3D accelerometer data are captured by the Shimmer3 GSR Unit.

All physiological signals were acquired with sampling rate of 256 Hz and resolution 16-bits per sample. The data collection from the two Shimmer units and the stimuli displayed on the PC monitor were synchronized with a purposely designed custom-build software application.

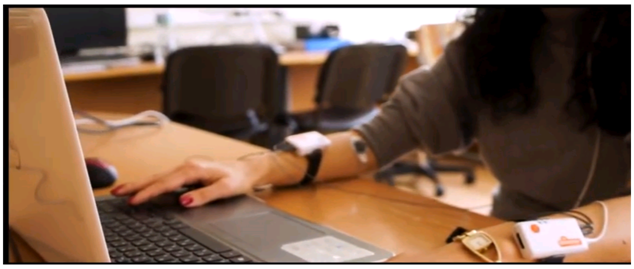

Fig. 4. The CLAS data collection setup, showing the two Shimmer units attached to the left and right arms

IV. DATASET DESCRIPTION

Following the data collection protocol outlined in Section III, 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 lefthanded.

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. Each person is represented in the dataset 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. The CLAS dataset is available free of charge for academic use[1].

V. Baseline results

We report the averaged recognition accuracy for the detection of low/high Arousal, Valence, High-Arousal Negative Valence (HANV) condition, and low/high Concentration. The averaged results reported here were obtained with person-specific SVM-based detectors, which we refer to as baseline results. In all experiments we made use of the recordings of 59 people from the CLAS dataset, which have a sufficient number of GSR and PPG/ECG recordings. In all experiments we used the leave-one-out data split approach. The SVM detectors make use of a polynomial kernel trained with the Sequential Minimal Optimization (SMO) method. In the perceptive tasks, the ground-true labels C_n were set to the corresponding stimuli tags. In the Concentration detection task, the ground true labels were set in accordance with the correctness of each answer in the Stroop test. The adjustable parameters of the SVM detectors were fine-tuned through the grid-search based on three parameters of the classifier: the box constrain C, tolerance ε, and polynomial order, p. The search range was set as follows: box constrain C ϵ [10e-6, 1] with step 10e+0.2, tolerance $\varepsilon \epsilon$ {10e-8,10e-7}, and polynomial order p ϵ {1,2,3}. The classifiers were fed with a set of 39 features [12] computed from two combinations of physiological signals: ECG+GSR vs. PPG+GSR. The former, ECG+GSR, requires larger number of electrodes (in our case 3) and higher computational demand for signal preprocessing. Fishers' discriminant ratio with threshold 0.3 was used to obtain the person-specific feature subsets in each classification task.

In Table I and Table II we present the averaged classification accuracy obtained for 59 people. Based on the experimental results we can conclude that no significant gain of performance is observed for the more demanding setup ECG+GSR when compared to the setup PPG+GSR.

Table I. The recognition accuracy in percentages for the two subsets of the CLAS database, corresponding to the perceptive tasks.

	Video clip stimuli			Picture stimuli		
	HALV	Arousal	Valence	HALV	Arousal	Valence
ECG+GSR	88.9 %	70.8 %	71.6 %	84.1 %	75.1 %	74.5 %
PPG+GSR	88.9 %	71.6 %	71.7 %	86.1 %	77.6 %	74.2 %

Table II. The recognition accuracy in percentages for the automated detection of high/low concentration.

	Concentration
ECG+GSR	78.2 %
PPG+GSR	74.2 %

VI. Conclusion

In the description of the CLAS dataset, we focused mainly on the aspects and 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. The limitations linked to the CLAS dataset, which was collected in this first pilot phase of our project, stem from its relatively small size and the static body position during data collection.

Acknowledgment

This research was supported by project NP6/2019, entitled "Development of an integrated platform for research concerning modern health-related technologies", funded by the NSFB and the Technical University of Varna

References

[1] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdan, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras. DEAP: A Database for Emotion Analysis Using Physiological Signals, *IEEE Trans. Affective Computing*, vol.3, no.1, pp. 18–31, 2012.

[2] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic. A multimodal database for affect recognition and implicit tagging, *IEEE Trans. Affective Computing*, vol.3, pp. 42–55, 2012.

[3] M. Abadi, R. Subramanian, S. Kia, P. Avesani, I. Patras, N. Sebe. DECAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses, *IEEE Trans. Affective Computing*, vol.6, no.3, 2015.

[4] R. Subramanian, J. Wache, M.K. Abadi, R.L. Vieriu, S. Winkler N. Sebe. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors, *IEEE Trans. on Affective Computing*, vol.9, no.2, pp. 147–160, 2018.

[5] V. Markova, C. Dicheva, F. Feradov, Y. Kalinin, T. Ganchev. SLADE–stress level and emotional state assessment database:Phase I, *Computer Science and Technologies*, vol.1, pp. 60–68, 2016.

[6] Ballenghein U., Megalakaki O., Baccino T., Cognitive engagement in emotional text reading: concurrent recordings of eye movements and head motion, *Cognition and Emotion*, vol. 33, issue 7, 2019.

[7] Sun, J., Yu, S., Chao, C., Effects of intelligent feedback on online learners' engagement and cognitive load: the case of research ethics education, *Educational Psychology*, vol. 39, issue 10, 2019.

[8] Mijic, I., Sarlija, M., Petrinovic, D., MMOD-COG: A database for multimodal cognitive load classification, ISPA 2019, pp 15-20, 2019.

[9] Mohamed, F., Krishnan, P., Yaacob, S., Validation of Driver's Cognitive Load on Driving Performance Using Spectral Estimation Based on EEG Frequency Spectrum, *Advanced Structured Materials*, vol. 119, pp 55-56, 2019.

[10] Ringeval, F., Messner, E., Song, S., Liu, S., Zhao, Z., Mallol-Ragolta, A., Ren, Z., Soleymani, M., Pantic, M., Schuller, B., Valstar, M., Cummins, N., Cowie, R., Tavabi, L., Schmitt, M., Alisamir, S., Amiriparian, S. (2019). AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. 3-12. DOI:10.1145/3347320.3357688.

[11] Lang, P.J., Bradley, M.M., Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8.University of Florida, Gainesville, FL.

[12] K. Kalinkov, V. Markova, T. Ganchev, "Front-end Processing of Physiological Signals for the Automated Detection of High-arousal Negative Valence Conditions", X National Conference with Intern Participation (ELECTRONICA-2019), Sofia, Bulgaria, pp. 1-4, 2019.

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