EEG student engagement dataset of: EEG-based measurement system for monitoring student engagement in learning 4.0

Andrea Apicella [1] Pasquale Arpaia[1] Mirco Frosolone[3,2,1] Giovanni Improta[2] Nicola Moccaldi[1] Andrea Pollastro[1]
Corresponding author: Pasquale Arpaia. pasquale.arpaia@unina.it
[1] Department of Electrical Engineering and Information Technology
[2] Department of Public Health and Preventive Medicine, University of Naples Federico II. [3] Institute of Cognitive Sciences and Technologies (ISTC) - National Research Council.

I. Experimental Paradigm

The dataset contains the EEG data of 22 subjects. Each subject participate with one session recorded. An ad-hoc experimental protocol was implemented to induce in the participants different Emotional and Cognitive Engagement status. The experimental sample was extracted from the population of college students in order to soft the impact of age and educational attainment on performance. The ethical committee of the University of Naples Federico II approved the experimental protocol. All methods were performed in accordance with the relevant guidelines and regulations. Before the experiment, each subject read and signed the informed consent. Each subject was equipped with a mouse to carry out the experimental test. After putting the EEG-cap on, the contact impedance was assessed to guarantee optimal signal-acquisition conditions. Each subject underwent an experimental session composed of 8 trials. Various stimuli to induce high and low levels of emotive and cognitive engagements were equally distributed among the trials. Continuous Performance Test (CPT)79 was used to modulate the cognitive engagement. In particular, a CPT version based on a learning by doing activity on how an interface works was adopted. Whereas, proper background music and social feedback was used to modulate the emotive engagement level. More in detail, the three different stimuli are described as follows:

- *Revised CPT:* a red cross and a black circle on the computer screen were presented to the subject. The red cross tends to run out from the circle on the screen in random directions. The subject was asked to keep the cross inside the circle by using the mouse. For each trial, a different difficulty level was set by the experimenter changing the cross speed. The percentage of the time spent by the red cross inside the black circle with respect to the total time was reported to the subject at the end of the trial (Fig. **??**).
- *Background music:* for each trial, a particular emotive engagement level was favored by proper background music. The music tracks were randomly selected from the MER database where songs are organized according to the 4 quadrants of the emotion Russell's circumplex model. The songs associated with the Q1 and Q4 quadrants (*cheerful music*) were employed in high emotional engagement trials, Q2 and Q3 for the low ones (*sad music*).
- *Social feedbacks:* during each trial, the experimenters gave proper social feedbacks according to the emotive engagement levels under the experimental protocol. The positive and negative social feedbacks consisted of encouraging and disheartening comments respectively, given to subject on his/her ongoing performance. The positive and negative social feedbacks were administrated using sentences composed of words extracted from a validated database proposed by Hermans and De Houwer
  (e.g. *intelligent, game, fast, rule, surprise, applause, good humour, strong, tenacious, skilful, damn, attentive, careless, talented, energetic, music, careless, weak, naive, silly, confused, inexperienced, clumsy, inhibited, great*, etc.). For example, subjects were encouraged and discouraged through comments such as:
  - "*Applause* to you. You did *great*, you achieved a very impressive score in this game. You deserve a round of *applause*. You are a real talent, what a nice surprise."
  - "*Damn*. You didn't do very well. You were careless. Shall we try again?"

  The social feedback effectiveness was also improved by the simultaneous music background effects.

A well-founded metrological reference is ensured by two assessment procedures validating the stimuli effectiveness:

- *performance index*: an empirical threshold was used to confirm that an appropriate CPT stimuli response was given by the participant. The threshold changed according to the trial difficulty level.
- *Self Assessment Manikin questionnaire (SAM)*: the emotional engagement level was assessed by a 9-level version of the SAM. The lower emotional engagement level was associated to the SAM score 1, while the greater one to 9.

The experimental session started with the administration of the SAM to get information about the initial emotional condition of the subject. Then, a preliminary CPT training phase to uniform all the participants starting levels was realized. After this preliminary phase, each trial was implemented by a succession of a CPT stage followed by a SAM administration.

A. Data labelling

45 s acquisition EEG signals were labeled according to two parameters: i) high or low emotional engagement, and ii) high or low cognitive engagement. More in detail, regarding the cognitive engagement, the trials were labeled according to the CPT speed, since the higher was the speed the more the cognitive engagement increased. Vesga et al. correlate directly the cognitive engagement with the cognitive workload. Many studies show how the changes in game difficulty are correlated with cognitive engagement and cognitive load. The concept of *desirable difficulties* is presented in terms of "varying the conditions of learning rather than keeping conditions constant and predictable". This concept is particularly interesting because it connects together the difficulty of the task, the level of involvement and the effectiveness of learning. In this study, the greater difficulty of the task is supposed to induce an increase in the cognitive resources employed by the participant, only if the performances remain compatible. The percentage of the time spent by the red cross inside the black circle with respect to the total time is the performance index used in this study. In detail, for each trial the performance index was analyzed and the subject was assessed as engaged if the final score was within 20 % variation respect the baseline. Otherwise, the trial was not included in the dataset. The trials having speed lower than 150 pixels/s were labeled as low_c, whereas $high_c$ were assigned to the trials having speed higher than 300 pixels/s. As concern the emotional engagement, the trials characterized by cheerful/sad music and positive/negative social feedback were labelled as $high_e$/low_e. For each trial, the SAM results (normalized to the initial pre-session values) were consistent with the proposed stimuli. In fact, a one-tailed t-student analysis revealed a 0.02 P-value in the worst case.

II. Data Recording

The AB-Medica Helmate system Class IIA (certified according to the Regulation on medical devices (EU) 2017/745) is used for the EEG signal measurements (fig.1 a). The device provides 10 dry electrodes disposed according to the International

(a) EEG-signal acquisition (b) Dry electrodes

Fig. 1: (a) EEG-signal acquisition device Helmate8 from abmedica, and (b) examples of its dry electrodes [**?**].

Positioning System 10/20: Fp1, Fp2, Fz, Cz, C3, C4, O1, O2, AFz (ref), and Fpz (Ground). The signals are differentially acquired with respect to the Fpz electrode and grounded to the AFz electrode. The Electrodes (made of a conductive rubber ending with Ag/AgCl coating) are of three different shapes to minimize the contact impedance in each scalp area (fig.1 b). The *Helm8 AB-Medica Software Manager* allows to i) verify the contact impedance level, and ii) apply several digital filters for a real-time signal visual analysis. The EEG signals are acquired with a 512 Sa/s sampling rate and sent via Bluetooth to a computation device.

III. Data file description

All data sets are stored in the .mat extension (one file for each subject) and zipped in one file. Each .mat file contains one structure, namely Subject_## with progressive number of participants. The single .mat struct contains thirteen fields:

- *Subject_##_task_(from 01-08)* are are the NxM matrices for the eight trials, where N is the channels number and M is the 45 s relative samples.
- *y* a vector of length 8 containing the labels. y_emotional_label for emotional engagement and y_cognitive_label for cognitive engagement.
- *Fs* is the sampling frequency;
- *Classes* is a 1x2 cell containing the task associated to the labels, both for emotional engagement (classes_emotional_engagement) and cognitive engagement (classes_cognitive_engagement);

IV. INFORMATION

All others information about data and processing are explained in the article: *EEG-based measurement system for monitoring student engagement in learning 4.0"* available at: https://doi.org/10.1038/s41598-022-09578-y [1].

REFERENCES

[1] Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., Pollastro, A. (2022). EEG-based measurement system for monitoring student engagement in learning 4.0. Scientific Reports, 12(1), 5857.