SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 15, No. 3, October 2018, 339-351

UDC: 004.414.22/.23:616.3-073 DOI: https://doi.org/10.2298/SJEE1803339A

The Potential for the Use of EEG Data in Electronic Assessments

Miloš Antonijević[1], Goran Shimic[2], Aleksandar Jevremović[1], Mladen Veinović[1], Slađana Arsić[3]

Abstract: One of the most important goals of electronic assessments is to achieve the smallest measurement error with tests that are as simple and short as possible. The psychological state of an examinee is typically ignored, both in the process of designing the tests and during the exam itself. Using the developed framework, we tested 35 participants in an experiment to obtain as much data as possible about the emotional states of the students depending on the different types of question posed. In this paper, we present our current results from an examination of the potential of using EEG data towards applying artificial intelligence for improvement of electronic assessments, as well as a technical platform for this purpose.

Keywords: EEG, Electronic assessment, Online testing, Human-computer interaction, Analytics.

1 Introduction

Although researchers have often tried to find a way to use neuroscience to improve e-learning, there is still a great deal of space to make more progress in this area. One of the biggest challenges of e-learning is how to carry out electronic assessments with results that reflect real learner competences (knowledge and skills) [2].

The focus of modern electronic assessment systems has changed from technical and security aspects to the content of tests. For instance, today there are more than 20 (basic) types of questions that instructors and teachers use to carry out testing. However, the testers usually ignore the participants' psychological state during assessment. Low importance is also placed on this aspect by teachers in the design phase of the questions and the test. This has

[1]Informatics and Computing Department, Singidunum University, Belgrade, Serbia;
E-mails: mantonijevic@singidunum.ac.rs; ajevremovic@singidunum.ac.rs; mveinovic@singidunum.ac.rs
[2]Center for Simulations and Distance Learning, Military Academy, University of Defense, Belgrade, Serbia;
E-mail: gshimic@gmail.com
[3]College of applied health sciences, Ćuprija, Serbia; E-mail: sarsic101@gmail.com

An earlier version of this paper was presented at the 5th International Conference IcEtran 2018, June 11 – 14, 2018, in Palić, Serbia. [1]

produced numerous complaints from the tested students that stressful conditions prevent them from achieving better results.

These facts motivated us to develop an experiment to collect information on how psychological states change during assessment and how different types of test questions influence the factors that form this psychological state: interest, engagement, excitement, stress, relaxation and focus. These factors represent derived information that is formed by capturing brain activity via electroencephalography equipment during assessments. We carried out assessments of almost thirty students in the computer science domain. During this experiment, we measured their emotional states at half-second intervals, recording the questions posed, the students' answers and the time elapsed simultaneously, and placing these data together in common storage for later analysis.

2 EEG device

The EMOTIV EPOC+ is a wireless EEG device with 14 channels, and is designed to measure brain cortex activity [3]. The possibility of accessing raw EEG data makes this device suitable for use in the development of brain-computer interface (BCI) applications. The data are sent using wireless Bluetooth technology.

According to technical documentation [4], the 14 channels of the EMOTIV EPOC+ device are referred to as AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 (Fig. 1). The names and positions of the electrodes are standardised using the *International 10–20 System* [5] convention to ensure the comparability and repeatability of the results.

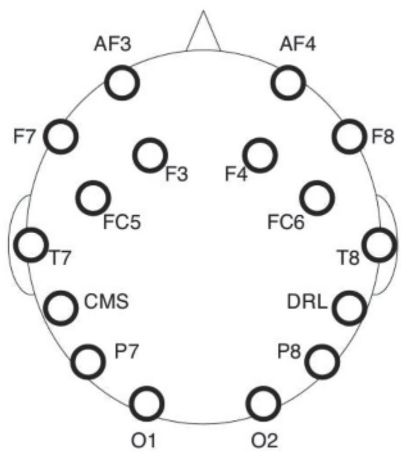


Fig. 1 – *Names and positions of EMOTIV EPOC+ channels.*

Based on the measured raw EEG values and rigorous experimental analysis, EMOTIV deploys three different types of algorithms for detecting facial expressions, emotional performances and mental commands [6]. Using widely available software tools, it is possible to determine six emotional and subconscious states: excitement, interest, stress, engagement, focus and relaxation.

3 Related works

The results of relevant works in which the EMOTIV EPOC+ device was used point to several important conclusions.

This device is capable of isolating a P300 low voltage signal (2–5 μV), which is considered to be associated with a stimulus evaluation or categorisation [7, 8]. Its low power compared to an EEG means that the device can distinguish this signal from the background noise that occurs during the measurement. The overall conclusion is that this device can be used as a reliable BCI.

In the study in [9], the authors investigated the emotional states of students, represented in the form of frustration and excitement that occurred as a result of feedback information gathered from intelligent tutoring systems (ITS). By analyzing the obtained data, the authors developed a system that enabled student emotions to be anticipated and the feedback information to be modified accordingly.

Similarly, emotional reactions to different visual stimuli were examined in two independent works [10, 11]. The first of these studied the ability to recognise EEG patterns in a state of relaxation and while imagining two different types of pictures (faces and houses), and a precision of 48% was achieved. In the second, the authors developed a method of interpreting EEG values in order to discriminate between mental patterns when participants observed pleasant and unpleasant pictures compared to neutral content.

The goal of the experiment described in [12] was the detection of the level of pleasure. The authors also tested a method for correcting the robot's behaviour in order to increase the pleasure level. Although the experiment was carried out on a small number of participants (four males), a correct classification was achieved in an average of 79.2% of cases.

4 Experiment framework

4.1 MyEmotivator application

We developed the *MyEmotivator* application to measure and record the variability in emotional performances depending on various external factors, with the EMOTIV EPOC+ device (Fig. 2). Using algorithms to calculate the values of six emotional states (interest, engagement, excitement, stress, relaxation and focus) based on gathered EEG data, this application is capable of displaying, processing and saving the measured information for use in real-time analysis by other services or humans. Within a wider platform, we implemented an option for synchronised pairing with other human-computer interaction (HCI) components to obtain a complete picture of the user–environment interaction.

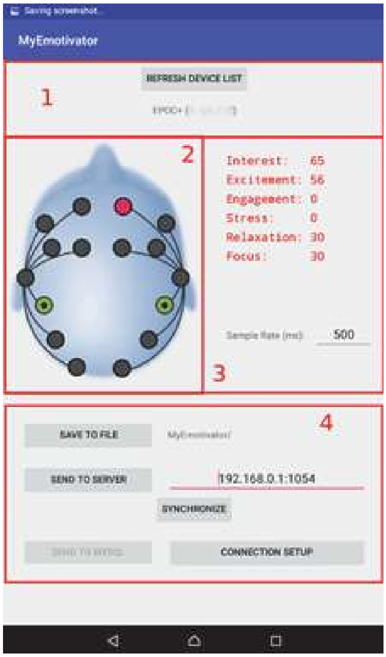


Fig. 2 – *Sections of MyEmotivator app Main window.*

In order to allow it to be used in the widest possible experimental conditions, the *MyEmotivator* application was created for use on mobile devices. In this way, it can be used in external environments even in cases when the participants in the experiment are in motion. Data are sent from the EPOC+ device using wireless Bluetooth technology. For this reason, a basic requirement for the device running the application is a Bluetooth SMART capability with a theoretical communication range of about 100 meters.

The recommended procedure for connection to the EMOTIV EPOC+ device involves turning on the device and searching for active devices in the Bluetooth domain (Fig. 2, Section 1). When the desired EPOC+ device name is selected by the user, the application begins the process of pairing (connecting) with the device. Under normal conditions, this process takes 1–2 seconds, although this may be increased when there is a large physical distance between the EPOC+ and the mobile device or if the battery is low. After successful pairing, the application stops the scanning process and starts the quality testing phase of the channel contact.

The application can uniquely determine the name of the channel that is sending the signal (in accordance with the International 10-20 standard) and can visually display its position on the head. There are three levels of signal strength: no signal, bad signal and good signal. Based on the detected level of signal quality, the connector positions are shown in a different colour in the corresponding picture on the main window: red, orange or green (Fig. 2, Section 2). Data about the emotional performances are gathered simultaneously with the quality information. The default reading frequency is set to twice per second (500 ms). Each of the six values examined (interest, engagement, excitement, stress, relaxation focus) fall into the range 0–100, where 100 is the maximal emotional level for the user, and 0 represents a theoretical minimum (Fig. 2, Section 3).

Due to its universality, the CSV format was chosen to organise the collected data. As the default option for data storage, the application saves data to the external memory of the mobile device (Fig. 2, Section 4). This is the most reliable method, with the lowest resource consumption and minimal probability of data loss. In addition to the six main emotional states, the application also saves the exact time of the measurement in the UNIX timestamp format for use in later analysis. The features of this recording option means that it is ideal in cases that do not require the simultaneous interpretation of measured data and real-time user interaction.

4.2 Human-computer Interaction Monitoring and Analytics Platform (HCIMAP)

One of the main challenges in using multiple sensors is time synchronization and fusion of gathered data [13]. With simple experiments, all sensors are connected to a single computer and share the same time source, so all data are collected simultaneously (synchronously). However, in case of more complex experiments which are conducted in a distributed environment, not only do those different sensors require more time sources, but they also use different network routes to central servers which have various delays.

Some of the reasons for shifting data processing to remote server(s) are existence of more participants in experiments, but also the need for fast processing of collected data and returning results to the main application in the form of generated feedback. In that way, adaptive behavior can be achieved and, for processing, we can implement more advanced artificial intelligence algorithms that require significant processor (CPU) resources in order to function in real time.

For this purpose we have developed a platform for synchronization of gathered data from different sensors and applications, their fusion and processing in real time and returning of obtained results in suitable formats for further analysis by computer or interpretation by humans (Fig. 3). Currently, EEG, eye-tracker, facial emotion recognition and mouse tracking sensors are supported; but, thanks to the open platform interface, connection to other sensors can be easily implemented. Beside the sensors, using the same interface, a platform can receive information from user applications. The platform supports exporting of all data as a time series (in CVS format), but also as more complexed reports (e.g., recording of user interface with eye position visualization), as well as modular addition of new report types. Data exporting can be done afterward, in stated manner, or in real time through the developed API. The set of algorithms used for data processing is expandable, either by implementation on the platform or by connecting to other systems on the network.

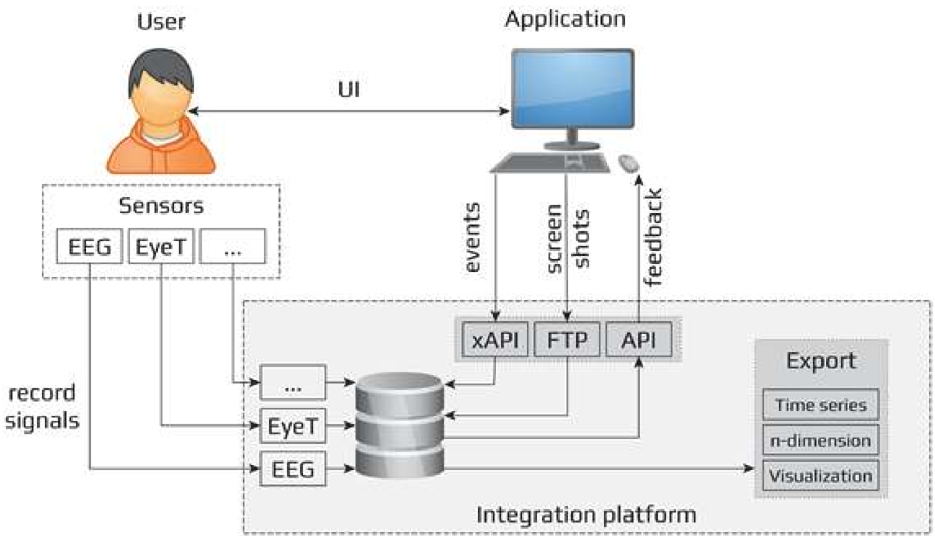


Fig. 3 – *Basic HCI-MAP platform architecture.*

Besides data gathering, the HCI-MAP platform enables data overview (by using Control-Room interface), as well as sending feedback to applications in real time (through API). In this sense, the platform is modularly extendable, i.e., it is possible to embed new algorithms for processing collected sensor data. Thanks to that, and based on EEG sensor data, it is possible to deploy adaptive behavior of the experimentally developed LMS (Learning Management System).

Finally, except for the real-time mode, the platform allows subsequent exportation of data. Data export can be in the form of time series (CSV format), visualized in standard graphics, or in custom format by developing appropriate module (Fig. 4). For example, one of the modules for data exporting that we have developed allows exporting of eye position, mouse cursor, screenshot and EEG states as animation.

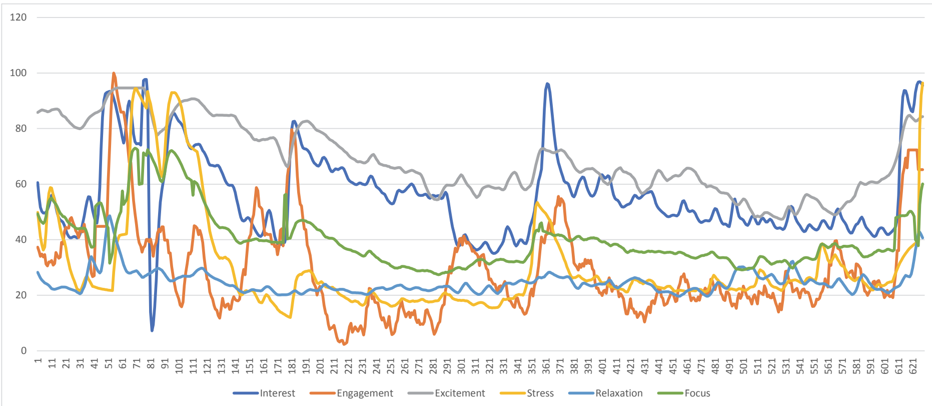


Fig. 4 – *Chart with EEG signals, collected during one of the sessions.*

HCI-MAP essentially uses the TCP/IP network stack and with sensors, i.e., their intermediaries, communicates by HTTP(S) protocol. Communication with some sensors is direct; for example, software that monitors mouse cursor position is realized as an application that runs on the client computer. On the other hand, some sensors communicate with the platform indirectly. In case of *MyEmotivator* application, software that delivers data from the EEG sensor to the platform runs on an Android tablet, which communicates with EEG device itself using Bluetooth protocol (Fig. 5).

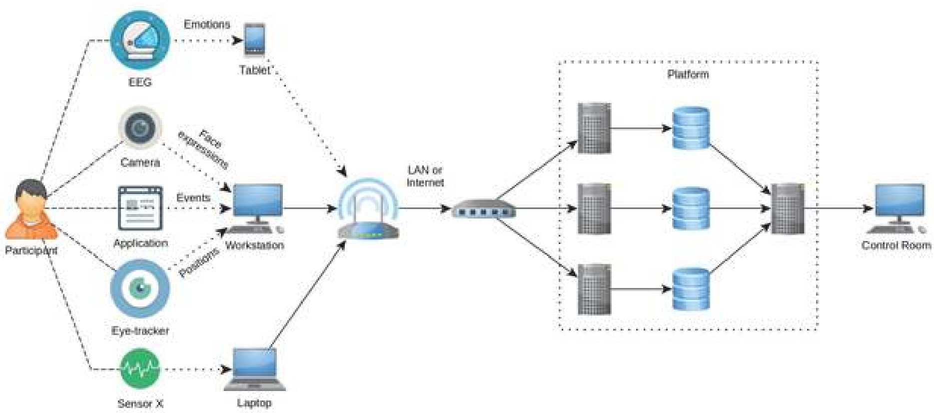


Fig. 5 – *Network organisation of HCI-MAP platform.*

One of the main challenges in the multi-sensor data gathering process is sensor data time synchronization. Sensor states in HCI can change many times per second. Errors of a few hundred of milliseconds can make data results completely useless.

We have decided to implement NTP (Network time protocol) [14] functionality in the HCI-MAP platform itself. It was realized by embedding the simple method (timeOffset) on the server side of the platform, which, like NTP server, sends the local time of receiving the request and sending the response. Formula for calculation the time offset:

$$T = \frac{(t_{sr} - t_{cs}) + (t_{ss} - t_{cr})}{2},$$

where:

tcs is the time when the client sent the request;

tsr is the time when the server received the request;

tss is the time when the server sent the response and

tcr is the time when the client received the response

(local client and server clock times are used).

The client side is done at the application level with software that controls the particular sensor (EEG, eye-tracker, etc.). Query to the server is repeated 10 times in a row and results with the lowest round-trip time are taken:

$$T_{rt} = (t_{cr} - t_{cs}) - (t_{ss} - t_{sr}).$$

Synchronization is done at the beginning of the experiment – i.e., at sensor initialization – so there is a fresh synchronization with every new experiment. With this approach, we achieved the required synchronization precision (<5 ms error margin on local 100Mb/s network and up to 30 ms error margin on ADSL 10Mb/s Internet connection), automated synchronization startup at the beginning of every measurement, as well as operating system and administrator privilege independence.

4.3 E-testing platform - mTutor

The e-testing platform used in this experiment is called *mTutor*, and was developed 12 years ago [15] (Fig. 6). It is a Web application that enables the electronic assessment of students' knowledge using single and multiple choice questions. The platform consists of two basic applications: one for teachers, and one for students. The teacher's application is used to create questions and answers, and to define the parameters of the test.

In the student application, after the user has logged in and selected a test, the questions are displayed. The order in which the questions and their possible answers is displayed is random. After choosing one or more of the answers offered, the student moves on to the next question. Throughout the test, the screen shows the number of questions remaining and the time remaining for the test. If the time runs out before all questions have been answered, the test ends. The evaluation strategy is such that there are no negative points for incorrectly answered questions; for partial answers, a certain percentage of the possible points is given, but in the case of multiple choice questions, the selection of one or more incorrect answers means that no points are given for that question.

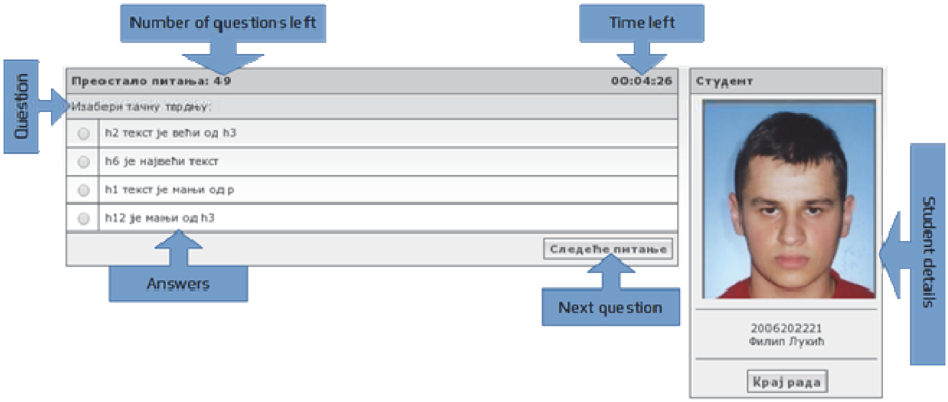


Fig. 6 – *mTutor platform - Student interface.*

5 Description of experiment

A total of 35 students voluntarily participated in the experiment. All participants were male high school students between 17 and 18 years old, and they all attended the same course entitled Basics of WEB Design.

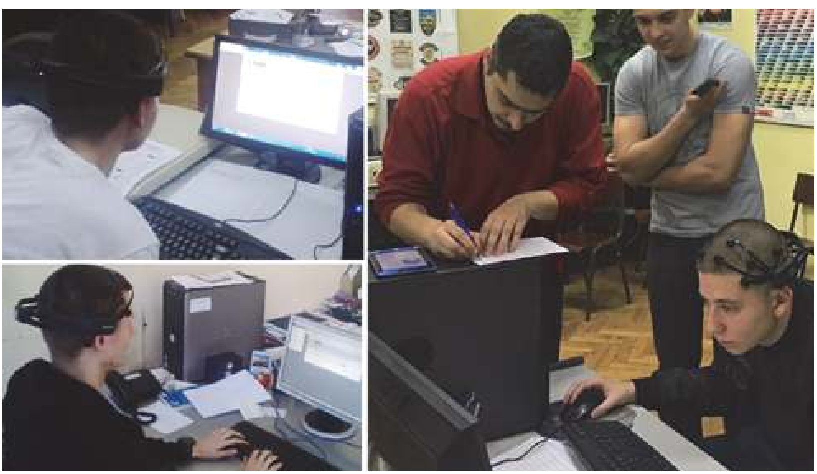


Fig. 7 – *Students held EEG equipment during the experiment.*

We used the content of this course as a basis for preparing the test questions. All participants possessed basic knowledge in computer usage and text editing. The experiment was conducted in a controlled environment in a high school computer laboratory (Fig. 7).

The questions covered material that was familiar to students from previous lectures. Students were told that their work would be graded and that the gathered data were to be used for improving future classes. Participants took the test in turn, wearing an EMOTIV EPOC++ device on their head under the teacher's supervision. Synchronisation problems arising in the experiment caused data collected from eight of the students to be discarded. Invalid data were excluded from further analysis.

The question pool contained 50 single choice questions. A test was individually generated for each participant, consisting of randomly chosen questions from the pool, and the time limit was set to five minutes. Students were instructed to answer as many questions as possible in the time allowed.

6 Results

In order to obtain as much data as possible about the different emotional responses of the participants, depending on the difficulty and content, five groups of questions were used in the experiment. Three groups contained regular questions (with *high*, *medium* and *low* difficulty; HDQ, MDQ and LDQ, respectively). The other two groups contained specific questions: *funny* questions (FQ) that had correct answers, but where the other possible answers were funny, and *impossible* questions (ImQ) with no correct answers at all.

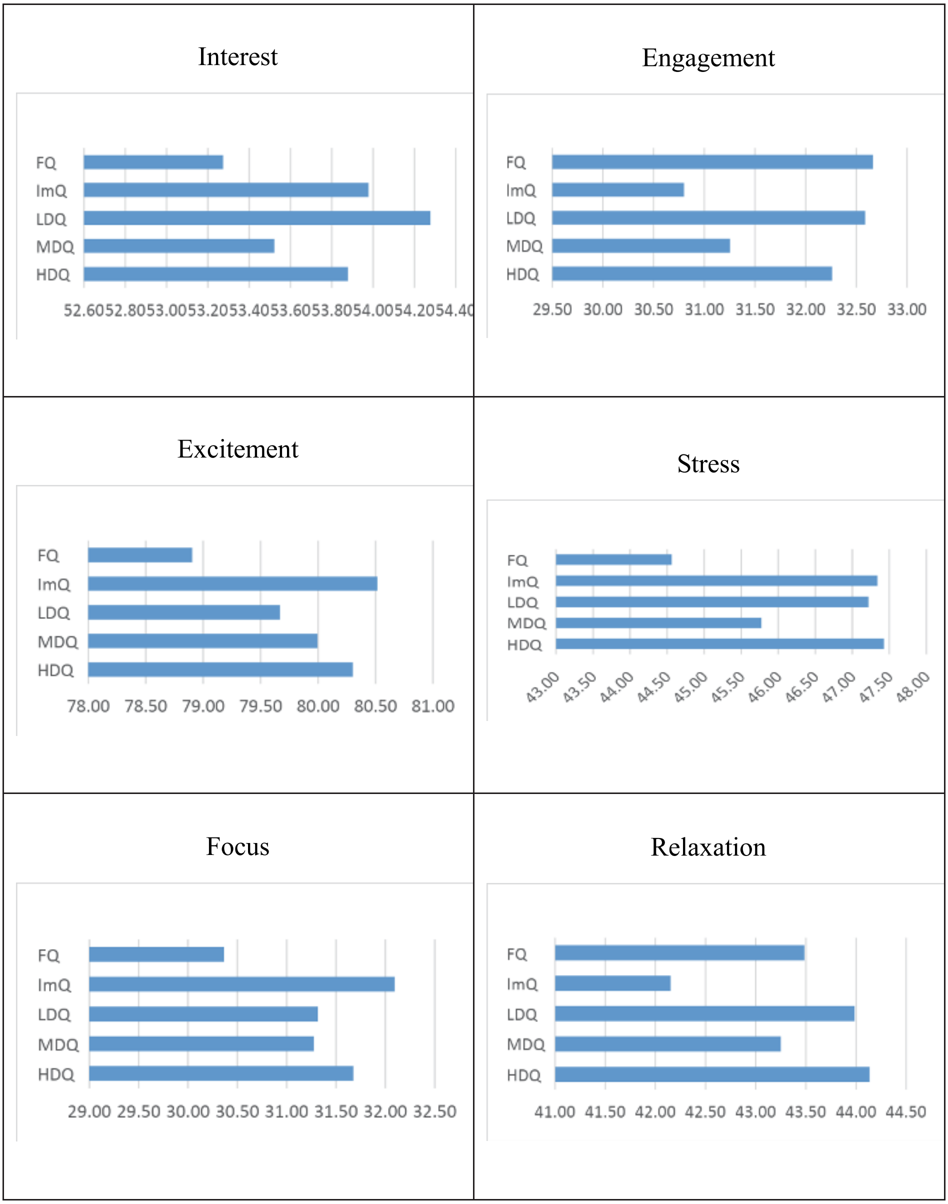
Table 1 shows six charts representing the average distribution of emotional factors measured for different question types. It is interesting that the funny questions (FQ) elicited minimum levels of stress, excitement, interest and focus, while the so-called impossible questions (ImQ) showed minimum levels of engagement and relaxation and highest levels of excitement and focus. The LDQ, MDQ and HDQ questions, as regularly used in exams, showed similar trends, with the HDQ showing the highest values on four out of the six emotional factors.

The second aspect of the statistical analysis was to find the correlations between the different parameters (correctness, time, interest, engagement, stress, relaxation and focus). This analysis was done for each of the questions, for the average parameter values for all participants in the test. Fig. 8 displays bivariate scatter plots (below the diagonal), histograms for each parameter, and Pearson correlations (above the diagonal).

From these results, it can be concluded that there is no strong correlation between any individual parameter and the accuracy of the response, and that the largest (negative) correlation is between the accuracy and the time taken. The greatest correlations occur between the following pairs: engagement–focus (+0.64), interest–relaxation (+0.62), stress–focus (+0.57), stress–relaxation (+0.51), and interest–stress (+0.48).

Table 1

Average emotional parameters distribution depending of question types.



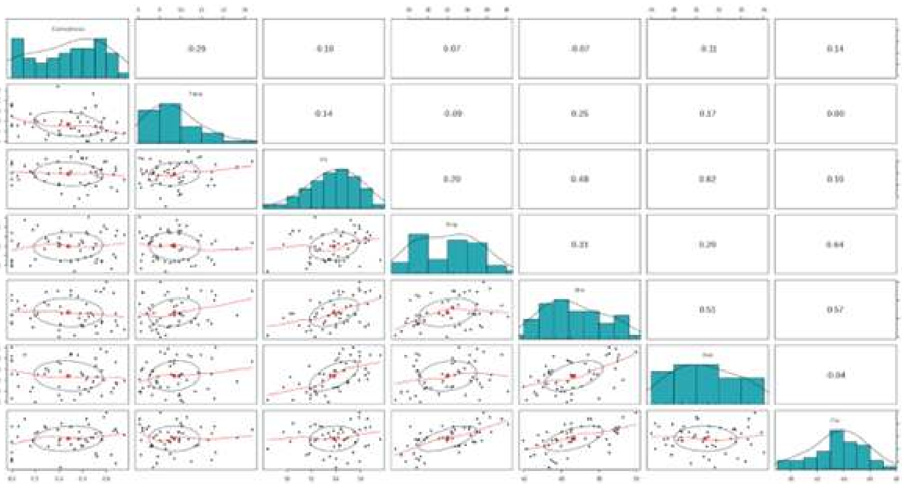


Fig. 8 – *Correlation between different question attributes - correctness, time, interest, engagement, stress, relaxation and focus*.

7 Conclusion

In this paper, we presented our current results from an examination of the potential of using EEG data towards applying artificial intelligence for improvement of electronic assessments. This paper describes the developed platform (HCI-MAP) for synchronised data gathering from multiple sensors and applications (EEG signals and events in the electronic testing system), and exporting (in real time) in formats suitable for further automated and human processing.

The psychological state of an examinee is typically ignored, both in the process of designing the tests and during the exam itself. Using the framework developed here, we tested 35 participants in an experiment to obtain as much data as possible about the emotional states of the students depending on the different types of question posed. The first analyses of the obtained (summarised) data are based on a search for a simple correlation between the various parameters recorded. The obtained results are not sufficient to demonstrate the effective use of the statistical method used here for the correction of students' test results, but they point to the possibility of using certain question types in order to influence the psychological state of students during assessments. For example, by inserting "funny" questions with one obvious correct answer, this system can decrease the students' stress. Furthermore, the most interesting questions are the easy ones, while the focus of the student can be increased by using "impossible" questions with no correct answers. However, an algorithm for rating students must be adopted for use with these new question types.

8 References

[1] M. Antonijevic, G. Shimic, A. Jevremovic, M. Veinovic: Potentials of Using Artificial Intelligence and EEG Data in Electronic Assessments, Proceedings of the 5th International Conference on Electrical, Electronic and Computing Engineering, Palic, Serbia, June 2018, pp. 1253 – 1255.

[2] R. F. Burton: Multiple Choice and True/False Tests: Reliability Measures and Some Implications of Negative Marking, Assessment and Evaluation in Higher Education, Vol. 29, No. 5, 2004, pp. 585 – 595.

[3] Emotiv, Available at: https://www.emotiv.com.

[4] EMOTIV EPOC+ - 14 Channel Wireless EEG Headset,
Available at: https://www.emotiv.com/epoc/.

[5] 10–20 system (EEG) – Wikipedia, Available at:
https://en.wikipedia.org/wiki/10%E2%80%9320_system_(EEG)

[6] Advanced EEG Technology - Backed by Science,
Available at: https://www.emotiv.com/the-science.

[7] H. Ekanayake: P300 and Emotiv EPOC: Does Emtiv EPOC Capture Real EEG? Retrieved from: http://neurofeedback.visaduma.info/emotivresearch.htm, 2011, pp. 1 – 7.

[8] S. Debener, F. Minow, R. Emkes, K. Gandras, M. de Vos: How About Taking a Low-Cost, Small, and Wireless EEG for a Walk?, Psychophysiology, Vol. 49, No. 11, 2012, pp. 1617 – 1621.

[9] P. S. Inventado, R. Legaspi, M. Suarez, M. Numao: Predicting Student Emotions Resulting from Appraisal of ITS Feedback, Research and Practice in Technology Enhanced Learning, Vol. 6, No. 2, 2011, pp. 107 – 133.

[10] P. Bobrov, A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, A. Zhavoronkov: Brain-Computer Interface Based on Generation of Visual Images, PLoS ONE, Vol. 6, No. 6, June 2011, pp. 1 – 12.

[11] M. K. Petersen, C. Stahlhut, A. Stopczynski, J. E. Larsen, L. K. Hansen: Smartphones Get Emotional: Mind Reading Images and Reconstructing the Neural Sources, Proceedings of the Affective Computing and Intelligent Interaction: 4th International Conference (ACII 2011), Memphis, USA, Vol. 2, Part 2, October 2011, pp. 578 – 587.

[12] E. T. Esfahani, V. Sundararajan: Using Brain–Computer Interfaces to Detect Human Satisfaction in Human–Robot Interaction, International Journal of Humanoid Robotics, Vol. 8, No. 1, 2011 pp. 87 – 101.

[13] F. Sivrikaya, B. Yener: Time Synchronization in Sensor Networks: a Survey, IEEE Network, Vol. 18, No. 4, July-August 2004, pp. 45 – 50.

[14] Network Time Protocol – Wikipedia, Available at:
https://en.wikipedia.org/wiki/Network_Time_Protocol.

[15] G. Simic, A. Jevremovic: Development of student knowledge testing system within frames of existing university IS, 15. Telecommunications Forum (TELFOR 2007), Belgrade, Serbia, November 2007, pp. 633 – 636. (In Serbian).