

# Measuring Concentration While Programming with Low-Cost BCI Devices: Differences Between Debugging and Creativity Tasks

Víctor M. González<sup>1</sup>(✉), Romain Robbes<sup>2</sup>, Gabriela Góngora<sup>1</sup>,  
and Salvador Medina<sup>1</sup>

<sup>1</sup> Department of Computer Science, Instituto Tecnológico Autónomo de México,  
Mexico City, Mexico

victor.gonzalez@itam.mx

<sup>2</sup> Computer Science Department (DCC), Universidad de Chile, Santiago, Chile

**Abstract.** Computing devices have become a primary working tool for many professional roles, among them software programmers. In order to enable a more productive interaction between computers and humans for programming purposes it is important to acquire an awareness of human attention/concentration levels. In this work we report on a controlled experiment to determine if a low-cost BCI (Brain Computer Interface) device is capable of classifying whether a user is fully concentrated while programming, during three typical tasks: creativity (writing code), systematic (documenting the code) and debugging (improving or modifying the code to make it work). This study employs EEG (Electroencephalogram) signals, to measure an individual's concentration levels. The chosen BCI device is NeuroSky's Mindwave due to its availability in the market and low cost. The three tasks described are performed in a physical and digital form. A statistically significant difference between debugging and creativity tasks is revealed in our experiments, in both physical and digital tests. This is a good lead for following experiments to focus on these two areas. Systematic tasks might not bring good results due to their nature.

**Keywords:** BCI · EEG · Concentration levels · Programming tasks · Debugging · Creativity · Systematic

## 1 Introduction

Computing devices have become a primary working tool for many professional roles, among them software programmers, whose work can be regarded as requiring discipline, creativity, expertise, but, perhaps most importantly, extended periods of sustained concentration. In order to create a more seamless and productive interaction between computers and humans for software programming we believe that it is important to open mechanisms and create new technologies that make computers aware of the level of attention/concentration level of the user.

EEG (Electroencephalogram) devices have a wide range of applications [1], ranging from control of computers, wheelchairs, and videogames or even determine the cognitive state when performing tasks in the real world. All these applications foresee a

positive future for EEG devices; their unfortunate downside is their high cost thereby limiting the research to large laboratories with enough funding. This is why noninvasive, low cost EEG devices such as NeuroSky, Emotiv, Epop, amongst others, have become so popular in the past years [2, 3]. They have expanded the applications of Brain-Computer Interfaces (BCI), for example controlling day-to-day device such as mobile phones [4].

There are two different types of sensors to capture neurological bio-signals: dry (e.g. used by NeuroSky's [5] products) and wet sensors, used, for instance, by the Biopac equipment [3], popular equipment in the research field, which brings high quality readings but at a great cost [6]. When it comes to electrodes, we also have two types: monopolar and bipolar. The monopolar electrodes collect signals at the active site and compare them to a reference point, which is usually located at the earlobe [2].

A comparison of the of the NeuroSky's single dry electrode against the Biopacs nineteen (19) wet electrodes have been made showing a great similarity in the results demonstrating the reliability of their product at a lower cost [6], therefore our chosen BCI device is NeuroSky's Mindwave. The signal obtained from the EEG by NeuroSky's products ranges from 1-100  $\mu\text{V}$ , which are read as signals through time. The signals are classified according to their frequency bands as Alpha, Beta, Gamma, Delta and Theta signals [5], where each class corresponds to different mental and emotional states of the individual.

## 2 Related Work and Justification of the Study

Studies in this field usually examine interruptions presented to working people, in particular in the information-centric context, either external or self-inflicted [9]. Interruption and workflows can be better understood if we can characterize individual's concentration stages during different types of tasks. This study is focused on user's awareness and concentration, by trying to determine if a low- cost EEG signal reader device befits the task and is able to determine if users are concentrated on their task or not. The chosen BCI device, as already mentioned, is NeuroSky's Mindwave due to its availability in the market and having the lowest cost at the moment of this study. The tasks to be performed by the users are three basic and typical activities in programming workflows: creative (thinking and writing code), debugging (improving or modifying the code to make it work) and systematic tasks (commenting and documenting the code).

As an initial hypothetical scenario we thought of a system that could obtain the user's concentration as a numerical input, enabling the system to aid the user to concentrate in a faster way and for longer periods of time by adapting his or her environment i.e. reducing any possible distraction. Distractions such as: notifications in the GUI on screen, notification sounds, telephone rings and even face-to-face unplanned type of interactions.

Goldberg et al. [7] tried to determine the user's engagement on certain task by using an electroencephalogram (EEG) device and a galvanic skin resistance (GSR) reader. Their results invite to the development of ubiquitous devices. This could also be applied for learning techniques, helping the lecturer with a real-time

measurement of their students. Additionally they stated that the interruption analysis is an important issue while trying to determine the attention of the user.

This assumption is confirmed by Mark et al. [9] and derived in a series of studies that determined for how long a person is able to complete a task after being interrupted while trying to achieve a certain task. In that study they determined that information workers in average focus, on a given task for 3 min before switching to another related task, and interrupt themselves or are interrupted by external sources every 11 min. However, studies like that have been made using observational approaches and a quantitative measure is required to extend their results and conclusions.

For this reason we decided to make an experiment as a first step, attempting to measure the concentration of a group of individuals which are aiming to complete programming tasks, both in a digital and a physical contexts. We aimed at exploring a typical programming scenario by studying a digital programming context, but to also a physical programming context as we are seeing more and more physical component programming scenarios emerging (e.g. Little Bits components).

### 3 Methodology

The experiments conducted in this study were divided into two one-hour sessions, for a total of 24 participants. Each session consists of three different types of tasks: systematic, debugging and creative. In the first session all tasks require a physical action, while in the second session they are done on a computer. The sessions were conducted on different days in order to avoid individuals' fatigue.

At the beginning of each experiment, the individuals were asked to watch a relaxing video in order to set the same starting mental basis on all the users. Afterwards, they were asked to fill out agreement forms and allow the evaluator to place and adjust the NeuroSky Mindwave device on them. The device had to be synchronized with the computer in order to record a total of seven brainwaves from the participant while performing the given tasks. The captured signals were the alpha, beta, gamma, delta, theta and NeuroSky's proprietary compound signals: attention and meditation.

During all tasks a video of the participants' actions on the screen was recorded. For any given task the time established was 10 min, with a 3–5 min interval between tasks in order to give sufficient time for reading instructions and answering doubts that might come up. The tests were conducted mostly on engineering university students (sophomores or older), ranging from specialized on computer engineering, telecommunications engineering, mechatronics engineering or industrial engineering. The requirement was to have passed the basic programming courses offered by the university in which they learn data structures and Java language.

#### 3.1 Physical Tasks

The session where the participants were asked to accomplish physical (programming) tasks started with the systematic task, where the individual was asked to organize a set of Legos by color and size, while filling an inventory sheet with the amount of Legos

corresponding to each type. For the creativity task, with the separated Legos from the previous task, the individual had to build a house with certain characteristics. The how, shape and color was for each individual to decide on, as long as it used the provided Legos and fulfilled the given characteristics such as the amount of doors or windows it should have. The third and final task simulated debugging, where the participant was given a tower of Uno-Stacko (similar to Jenga but with colors and numbers) and he or she had to move one piece at a time in order to build the pattern provided to them in a separate piece of paper i.e. physically debug the plastic tower of sticks.

### 3.2 Digital Tasks

The second session was formed by a set of three programming tasks. During the first task the individual had to debug a certain code in Java (using NetBeans IDE). The requirement was to be able to compile the code and print an expected output, of which each individual was informed of, prior to the test. Afterwards, the individual had to program, in Java language, small problems extracted from Project Euler [11]. The problems obtained vary in difficulty according to what is taught in the required programming course. The participant could choose which ones and in which order to solve the problems given. For the last test the individual was asked to document the code he or she generated in the previous task. With these we again have the same three categories of tasks as in the previous session: debugging, creativity and systematic.

## 4 Participant's Characterization

The experiment described was conducted on 24 participants, mostly engineering students, sophomores or older, as long as they fulfilled the prerequisites. The initial experiment was re-designed and corrected with the help of an additional 8 participants. From the other 24 individuals, only 20 had useful signals and complete physical and digital tasks. From the total 32 participants, 26 of them were male students and 6 female students. The latter reported consistent problems with the Mindwave, as it kept falling off their heads.

**Table 1.** Demographics of participants

Measure	Age	Semester	GPA
Min.	20.00	3.000	7.500
1st Qu.	20.00	5.000	8.000
Median	21.00	7.000	8.775
Mean	21.04	6.179	8.679
3rd Qu.	21.25	7.000	9.125
Max.	24.00	11.000	9.900

A more detailed characterization of the students was made in order to understand the population at hand and the results obtained. Table 1 shows the range of age, semester and GPA of all the students that participated in the experiment. The GPA is measured under a 10-point scale, where 5 is a fail, 6 is a passing grade and 10 is the highest grade possible. All our participants have ages 20 to 24 and in average are more than halfway through their bachelor's degree.

**Table 2.** Participant's area of study (Mayor)

Degree	Number of students
Computer Eng.	21
Mechatronics	6
Telecommunications	1

The majority of the students Computer Engineers, Table 2 shows the distributions of BSc degrees.

**Table 3.** Participant's with a double degree

Degree	Number of students
None	16
Administration	1
Industrial Eng.	2
Applied Mathematics	2
Mechatronics	2
Telecommunications	5

Some of the individuals study two degrees and Table 3 accounts for them. The variety in the second degree is more than in the first one, although only 16 out of 32 participants have double mayors.

## 5 Signal Processing

One of the main problems presented while working with the Mindwave headset was its recording resolution. The maximum sample rate of the signals is 1 Hz. This forced the experiment to sample long periods of time in order to obtain relevant results.

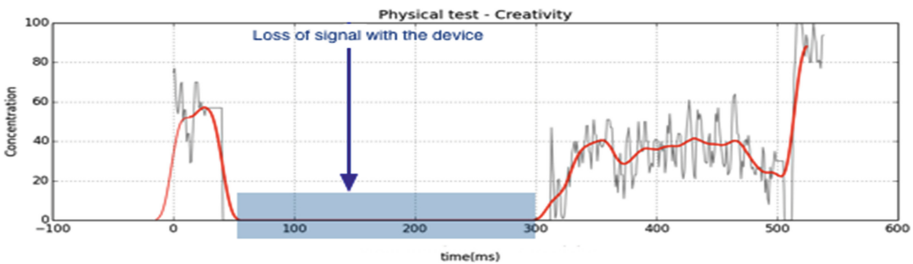
The Mindwave device was capable to sample the alpha, beta, gamma, delta and theta waves, as well as two composite signals: attention and meditation. All the signals aforementioned were taken into consideration for the statistical analysis, although our main focus was the Mindwave's attention proprietary signal. This signal brought the most reliable and stable results.

### 5.1 Cleaning the Results

The cleaning process of the results obtained required the elimination of six individuals due to the loss of recording during their testing period.

The loss of the recording signal can be due to many factors such as: a lousy bluetooth signal from the Mindwave device; the participant’s head size where the Mindwave would be too big; the participants forward tilting their heads while making some of the physical tests. This last factor should be taken into consideration in future tests by trying to avoid that the individuals move their head during their test.

The graphic representation of how a signal was lost during the test looks like Fig. 1.



**Fig. 1.** Loss of signal during session

The next table shows the amount of individuals whose captured data was relevant. All of them participated in each of the three tests: Systematic, Debugging and Creativity, in both of their modalities: Physical and Digital.

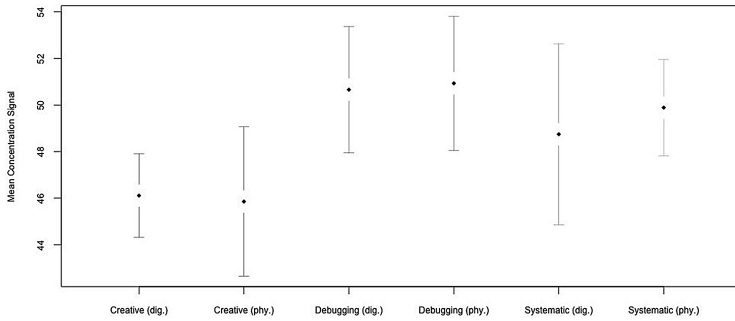
Afterwards, the signals’ noise was removed based on the work done by Hong Tan [9] we decided to apply a FIR filter to smooth the signals as they are impossible to work with at first instance. The FIR filter had a cutoff of  $1/100$  a filter order of 30 and a warm-up of 30 s, as the individuals were opening the required software or being explained the tasks during that time (Table 4).

**Table 4.** Most relevant test subjects

Type of Task	Number of individuals
Systematic	14
Debugging	16
Creativity	16

## 6 Results

An initial sampling of the signals through a means plot, Fig. 2, lets us evaluate a visual difference between the three types of tasks and in particular between creative and debugging tasks. A resemblance amongst physical and digital tests can also be observed from this figure. Further statistical analysis will provide solid evidence and explanation for this outcome.



**Fig. 2.** Means Plot with 95 % Confidence Interval

As a first approach, an ANOVA of means was performed taking into consideration all signals obtained, including composite signals. The analysis was done for the physical tests and the digital tests separately. Then an ANOVA was made to find if a similarity existed among the tasks in the physical world compared to their digital counterparts.

### 6.1 Physical Analysis

The ANOVA test was applied to the means of each signal obtained through out the whole test of each participant. The hypothesis for this particular test is:

$H_0$ : means are the same

$H_1$ : means differ (at least one in the group)

**Table 5.** Anova test performed on physical tasks

	Df	Sum squares	Mean squares	F value	Pr(> F)
(Intercept)	1	109764.59	109764.59	4105.96	0.0000
groups	2	227.94	113.97	4.26	0.0205
Residuals	43	1149.52	26.73		

The *p-value* for the ANOVA test was 0.0205, which is less than 0.05, therefore we can reject  $H_0$  and conclude that at least one mean in our tested population differs from the others. Further testing has to be in order to know which group in particular we are referring to (Table 5).

### 6.2 Digital Analysis

The already explained ANOVA test was employed to the digital tasks, obtaining a *p-value* of 0.044 as shown in Table 6.

**Table 6.** Anova test performed on digital tasks

	Df	Sum squares	Mean squares	F value	Pr(> F)
Intercept	1	108182.02	108182.02	4054.37	0.0000
groups	2	166.81	83.41	3.13	0.0440
Residuals	43	1147.36	26.68		

From these tasks we can also reject  $H_0$  and expect to find differences between the type of tasks (debugging, systematic, creative) when using a post hoc analysis.

### 6.3 Post-Hoc Analysis

When evaluating both physical and digital results the difference between each task becomes even more evident than in previous results shown.

The *p-value* for the ANOVA test for all, physical and digital tasks, was 0.0158, as shown in Table 7, which is less than 0.05, therefore we can reject  $H_0$  and conclude that at least one mean in our tested population differs from the others.

**Table 7.** Anova test performed on all tasks

	Df	Sum squares	Mean squares	F value	Pr(> F)
Intercept	1	217943.74	217943.74	8160.27	0.0000
groups	5	397.63	79.53	2.98	0.0158
Residuals	86	2296.88	26.71		

The Holm adjustment method, as post hoc for ANOVA analysis, is useful in our case as it contrasts the lowest *p-value* with a Type I error rate i.e. decreasing it for each subsequent test. The results are shown in Table 8.

The obtained results identify a clear difference between creative and debugging tasks, both in physical and digital tasks. A distinction between creative and systematic tasks can also be established, although such observation is only statistically significant in physical tasks (phy.) and not in digital tasks (dig.).

Another method employed for comparing results is the Tukey Honest Significant Difference (HSD), as it can adjust to correct Type I error rates between multiple comparisons and it is considered as an acceptable statistical technique. The results obtained are shown in Table 9. It can be observed that our results are quite identical (although not mathematically) to the ones using the Holm adjustment method.



**Table 8.** Pairwise comparisons using t tests with pooled SD - Holm adjustment method

	Creat. (dig.)	Creat. (phy.)	Debugg (dig.)	Debugg (phy.)	Syst. (dig.)
Creat. (phy.)	0.89	–	–	–	–
Debugg. (dig.)	0.01	0.01	–	–	–
Debugg. (phy.)	0.01	0.01	0.88	–	–
Sys. (dig.)	0.17	0.13	0.31	0.25	–
Sys. (phy.)	0.05	0.04	0.68	0.58	0.56

**Table 9.** Tukey multiple comparisons of means test, 95 % family-wise confidence level

Task Compared	Diff. in means	Lower end	Upper end	p-value (after adj.)
Debugging (dig.)-Creative (dig.)	4.55	0.12	8.98	0.04
Systematic (dig.)-Creative (dig.)	2.64	-1.95	7.22	0.35
Systematic (dig.)-Debugging (dig.)	-1.91	-6.50	2.68	0.57
Debugging (phy.)-Creative (phy.)	5.08	0.64	9.51	0.02
Systematic (phy.)-Creative (phy.)	4.03	-0.56	8.63	0.10
Systematic (phy.)-Debugging(phy.)	-1.04	-5.64	3.55	0.85
Debugging (phy.)-Debugging (dig.)	0.27	-5.05	5.60	1.00
Systematic (phy.)-Systematic (dig.)	1.14	-4.55	6.84	0.99
Creative (phy.)-Creative (dig.)	-0.26	-5.58	5.07	1.00
Debugging (phy.)-Creative (dig.)	4.82	-0.50	10.15	0.10
Systematic (phy.)-Creative (dig.)	3.78	-1.74	9.29	0.35
Systematic (phy.)-Debugging(dig.)	-0.77	-6.29	4.74	1.00

Debugging and systematic tasks did not differ much, as shown in light blue in Table 9, they have *p-values* equal or close to 1.0 i.e. their means hardly differ and can be considered as equally demanding tasks in terms of concentration brainwaves. Debugging and creative tasks do show a statistical difference in their means, as can be appreciated in light yellow in Table 9, as their *p-values* are less than 0.05.

## 7 Conclusions and Future Work

We observed a statistically significant difference between debugging and creativity tasks no matter if the test was physical or digital. This is a good lead for following experiments to focus on these two areas. Systematic tasks might not bring good results due to their nature: a systematic task does not require the participant's entire focus to achieve its completion and as such may lead the individual to a state of complete dispersion. Finally, we observed that the data generated by the BCI was rather noisy; research is being performed to analyze the performance of other higher-end BCI devices in our subsequent studies.

One of the main contributions of this paper is that it gives some insights into the differences between creative, debugging and systematic programming tasks, both in the digital and physical worlds. In the long-term, these results could be used to implement better software development environments.

Nonetheless this study holds some limitations. First of all, the programming tasks given throughout the experiments are very short in time, for example the digital creative tasks usually took between 3-9 min (depending on the programmer's skills). A full-time programmer could spend 8-12 h a day working on the same programming problem, thereby having longer spaces of time to concentrate. This limitation can also be viewed as an advantage, since we could consider that a working programmer is, in many cases, subject to various interruptions throughout the day and has to get back into the work flow as quickly as possible i.e. this limitation is actually an insight into how fast a programmer can get concentrated in each type of task.

Future work could show us if the 3-9 min in creative digital tasks are a true sample of a full-time programmer's awareness level, and the insights could be expanded to cover more situations resulting from this process. Other devices (e.g. heart rate monitors) and camera recordings could be of aid in this.

NeuroSky's MindWave has a proven to be a reasonable device for the experiments performed, as it was easy to use, below a \$200 dollar budget and non-intrusive to wear. The only three problems reported throughout the process were the loss of connectivity (computer-device), the interference of other Bluetooth devices around the individual and headband falling off some individuals. Therefore we are currently continuing the study with an Emotiv EPOC device, as to have more data that will backup the MindWave and to obtain further characterizations of programmer's concentration during debugging and creative tasks aside from evaluating how each programmer gets back into its work flow. Future work plans also include the incorporation of interruptions into the study, an in-depth profiling of the programmers participating and longer tasks.

**Acknowledgements.** This work has been supported by LACCIR (Project ID: RFP1212LAC004) the Asociación Mexicana de la Cultura A.C. and the Consejo Nacional de Ciencia y Tecnología de México (CONACyT). We thank Frida Rojas and Farid Fajardo Oroz for their help editing and proofreading the document.

## References

1. Tan, D., Nijholt, A.: Brain Computer Interaction: Applying our Minds to Human-Computer Interaction, vol. Ch. 1. Springer, London (2010)
2. Stamps, K., Hamam, Y.: Towards inexpensive BCI control for wheelchair navigation in the enabled environment – a hardware survey. In: Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N., Huang, J. (eds.) BI 2010. LNCS, vol. 6334, pp. 336–345. Springer, Heidelberg (2010)
3. Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cathy, D.S., Seetharaman, K., Cheron, G.: A P300-based quantitative comparison between the Emotiv EPOC headset and a medical EEG device. In: Proceedings of the 9th IEEE/IASTED International Conference on Biomedical Engineering. Acta press, Innsbruck (2012)

4. Campbell, A.T., Choudhury, T., Shaohan, H., Hong, L., et al.: NeuroPhone: brain- mobile phone interface using a wireless eeg headset. In: Proceedings of the Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds, MobiHeld 2010. ACM, New York (2010)
5. NeuroSky Brain-Computer Interface Technologies. Brain Wave Signal (EEG) of NeuroSky, Inc. NeuroSky, Inc. (2009)
6. Crowley, K., Sliney, A., Pitt, I., Murphy, D.: Evaluating a brain-computer interface to categorize human emotional response. In: 10th IEEE International Conference on Advanced Learning Techniques. Cork, Ireland (2010)
7. Goldberg, B.S., Sottolare, R.A., Brawner, K.W., Holden, H.K.: Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions. In: D’Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 538–547. Springer, Heidelberg (2011)
8. Tan, B.H.: Using a Low-cost EEG Sensor to Detect Mental States. Diss. Carnegie Mellon University (2012)
9. Mark, G., Gudith, D., Klocke, U.: The cost of interrupted work: more speed and stress. ACM, Proceedings of the SIGCHI conference on Human Factors in Computing Systems (2008)
10. Project Euler. <https://projecteuler.net/>
11. Banerjee, A., Kanad B., Aruna C.: Prediction of EEG signal by digital filtering. In: Proceedings of International Conference on Intelligent Systems & Networks, Jagadhri, India (2007)
12. Oikonomou, V.P., Tzallas, A.T., Fotiadis, D.I.: A Kalman filter based methodology for EEG spike enhancement. *Comput. Methods Programs Biomed.* **85**(2), 101–108 (2007)