A neurofeedback system to promote learner engagement

James Lockwood
Dept. of Computer Science
Maynooth University
James.lockwood@nuim.ie

Aidan Mooney
Dept. of Computer Science
Maynooth University
aidan.mooney@nuim.ie

Susan Bergin
Dept. of Computer Science
Maynooth University
Susan.bergin@nuim.ie

Abstract
This paper describes a series of experiments that track novice programmer’s engagement during two attention based tasks. The tasks required participants to watch a tutorial video on introductory programming and to attend to a simple maze game whilst wearing an electroencephalogram (EEG) device called the Emotiv EPOC. The EPOC’s proprietary software includes a system which tracks emotional state (specifically: engagement, excitement, meditation, frustration, valence and long-term excitement). Using this data, a software application written in the Processing language was developed to track user’s engagement levels and implement a neurofeedback based intervention when engagement fell below an acceptable level. The aim of the intervention was to prompt learners who disengaged with the task to re-engage. The intervention used during the video tutorial was to pause the video if a participant disengaged significantly. However other interventions such as slowing the video down, playing a noise or darkening/brightening the screen could also be used. For the maze game, the caterpillar moving through the maze slowed in line with disengagement and moved more quickly once the learner re-engaged. The approach worked very well and successfully re-engaged participants, although a number of improvements could be made. A number of interesting findings on the comparative engagement levels of different groups e.g. by gender and by age etc. were identified and provide useful pointers for future research studies.

1. Introduction
Computer Science (CS) non-progression rates in Ireland are alarming with a large number of students failing to progress every year. Currently non progression rates are at 25% in CS, which is significantly higher than the national average of 16% (Mooney, 2010). It is well acknowledged that a main contributor is introductory programming modules, a staple in most first year CS courses. At our university numerous methods have been tried to improve performance and retention to varying degrees of success (Mooney, 2014; O’Kelly et al., 2004; O’Kelly et al., 2004, Nolan et al. 2015). One of our recent initiatives was the development of self-paced video tutorial on threshold programming concepts (Hegarty-Kelly, 2015). These videos were introduced to allow students to revisit core concepts in their own time, at their own pace. The goal of this paper is to measure how well students engage with learning videos using an EEG device to monitor engagement and where a student disengages prompt them to reengage using neurofeedback. To this end a number of experiments were implemented to monitor engagement during a threshold concept video and also during a mind-control game. The rationale and background for this work are provided in the following section.
2. Background

A review of the literature was carried out on (1) the relationship between engagement and learning, (2) the use of electroencephalogram (EEG) devices (especially the Emotiv EPOC used here) for monitoring neural activity during cognitive tasks, and (3) the use of videos as a learning tool.

An EEG is a non-invasive technique to measure activity in the brain. The EPOC is a 14 channel EEG device for research. It is a dry (uses a small amount of saline) high resolution consumer grade EEG system. A number of verification studies have been carried out comparing the EPOC to a clinical EEG but there has been little independent study as it is a relatively new technology. However, proprietary claims suggest its appropriate for research and, of the limited studies, all have found that it performs well and can be used in place of more expensive units. Taking this into consideration the results obtained can be considered accurate for example, Ekanayake (2010) concluded that the EPOC does capture real EEG data but has considerable noise. This noise can be minimised using different techniques such as averaging. It was also found that electrode placement is fairly fixed and thus this reduces the number of studies it can be used for. Andujar (2011) used the EPOC in a study that compared the engagement level of participants studying a paper handout with using a video game to learn about the Lewis and Clark Expedition (www.classbraingames.com). The game disseminated information about the expedition and the control group received the same information in the handout using only text. Twenty-six participants took part (13 in the experimental group and 13 in the control). Each group had 20 minutes to complete the learning task. Participants wore the EPOC during the task and engagement levels were measured. The results of the study suggested that educational video games might not be significantly engaging and found that learning from handouts could be better for retaining information. This is an interesting finding and the closest study we could find on the use of the EPOC to measure engagement. But this study uses a video game and not a video tutorial and it could be argued, that a tutorial video reduces any cognitive load associated with aspects of the games or story-line issues which might impact on learning.

Aside from EEG studies on engagement other empirical studies on engagement have highlighted its importance for learning and comprehension, some examples follow. Carini et al. (2006) found that student engagement is linked positively to desirable learning outcomes such as critical thinking and grades. In a study on the use of clickers in the classroom Blasco-Arcas (2013) found that increased engagement through the clickers improved student learning. Other studies have shown that increased engagement in reading leads to higher comprehension (Guthrie & Wigfield, 2000 and Miller & Meece, 1999).

With respect to using educational videos to improve student learning, Merkt et al. (2011) conducted two complementary studies, one in the laboratory and one in the field, comparing the usage patterns and effectiveness of interactive videos and illustrated textbooks when German secondary school students learned complex content. They used two separate videos of differing degrees of interactivity (the second of which they made themselves) and an illustrated textbook. Both studies showed that the effectiveness of interactive videos was at least comparable to that of print, in contrast to previous studies working with non-interactive videos. More recently, Willmot et al. (2012) show that there is strong evidence that creating an on-going video report during projects
can inspire and engage students when incorporated into student-centred learning activities through: increased student motivation, enhanced learning experience, higher marks, potential for deeper learning of the subject, enhanced team working and communication skills, and by providing a source of evidence relating to skills for interviews. These studies indicate that videos can be used as an enhancement to a lesson, or unit of study. Furthermore, Guo et al. (2014) provided a list of recommendations on what makes a video engaging and beneficial. To do this they used both data analysis of 6.9 million video watching sessions and interviews with video production staff. These recommendations and findings include: shorter videos are much more engaging, videos where instructors speak relatively fast and with high enthusiasm are more engaging and that videos should incorporate motion and continuous visual flow in tutorials, along with extemporaneous speaking. Pre- and post-production planning is essential as topics can be segmented into smaller sections. They also argue that filming in an informal setting is more engaging than a recording studio. The recommendations provided above influenced the development of the threshold video described in Hegarty-Kelly et al. (2015).

The goal of the work presented here is to monitor, track and respond to changes in engagement whilst cognitive tasks are being attended to by gathering neuro-physiological data from the student. A neurofeedback system like this, which can detect, track and respond to these changes in an online learning environment could be beneficial to learning, especially online learners in the absence of a teacher to monitor and gauge attention. This is especially true and timely with the large growth of MOOC’s (Massive Open Online Courses) in recent years.

This paper describes a neurofeedback system which monitors and responds to engagement levels in two cognitive tasks. First, whilst watching a tutorial video the feedback system would pause the video if a learner disengaged and only restart after re-engagement had occurred. This system could be of huge benefit in MOOC’s and long-distance learning courses which use videos as a major learning tool. The second task is a maze game in which a caterpillar moves around a maze, and the speed at which this happens is related to a participant’s engagement level. This application has the potential to be used in therapy settings for conditions such as Attention deficit hyperactivity disorder (ADHD). This engagement therapy taps into the brains ability to re-wire its own connections (neuroplasticity) and has been shown to be a superior method of treatment for ADHD than classical medicated interventions of the past (Fuchs et al. 2003).

3. Method
3.1 Hardware and software Description

The Emotiv EPOC was designed for practical research applications and comes with a software suite that measures several emotional states including engagement, as well as boredom, excitement, frustration and meditation level in real time. Two applications were developed. The first monitored a user watching a video whilst wearing the EPOC and allowed interventions to take place depending on a user’s engagement level. The second was a maze game, where an object would travel around a maze and the speed at which it moved was determined by the user’s engagement level. The idea for this game was based on similar neurodfeedback games developed for children with ADHD such as that of Fuchs et al. (2003). Their version of pac-man was developed for children with ADHD where if they start to lose focus on the game pac-man starts to fade. The idea of the maze game developed here,
in the Processing language (Foundation, 2016), was a very simplistic idea in which a caterpillar-like object moves around a maze. The speed of the caterpillar is related to the participants level of engagement, the higher the engagement level the faster the caterpillar moved, the lower the the engagement, the slower it moved.

3.2 Participants, Experimental design and pilot study
Ethical approval was sought and granted for the study: all participants signed a consent form and an information sheet was provided prior to the session with questions. In total 21 undergraduate (final year) students along with 5 postgraduate students were recruited to participate in the study (16 of the undergraduate and 4 of the postgraduates students studied CS). A video of 8 minutes duration on if statements, was used. This is material that all 4th year CS students would be extremely familiar with and so would potentially increase the likelihood that the participants would disengage and show the capabilities of the system.

3.3 Instruments:
A post-experiment questionnaire was used to gather feedback on perceived engagement for both the maze and video tasks based on a survey by Wigfeld & Guthrie (1997) (further adapted by Hyun-Gyung Lee (2012)), which all participants filled out directly after doing both the video and maze tasks (see Lockwood & Bergin 2016). All data was anonymised and each participant was given a unique identifying code. This code was used by participants in another survey that gathered demographic and academic background information.

3.4 Experimental Design:
All participants completed the survey and took part in all of the tasks in the following order: Calibration, Video, Video Survey, Maze and Maze Survey. Although counter-balancing an experiment that involves discrete components can be important, the video was always shown first as we were most interested in it.

Before the main study a three-person pilot study was carried out to ensure that each of the experiments worked correctly and to test whether the selected baseline engagement value (set at 0.5, the mid-point value for engagement in the proprietary software) was reasonable. After two participants it was clear that this crude approach to measuring engagement was insufficient and an individualised calibration procedure was needed as engagement levels differ greatly from person to person. As such a calibration process was carried out where (1) the participant fixated on a black cross on a grey screen for sixty seconds, and (2) the participant closed their eyes for sixty second. During the latter, participants were instructed to try to relax and to not focus on anything in particular. Whilst their eyes were closed the program took a running average of their engagement level for that minute and this value was used as each participants baseline value. In theory, when the participants eyes are closed they should have been disengaged and this was confirmed in their levels. This baseline was then used as the threshold for whether the video should pause and also as to how slow the caterpillar would move in the maze game. This means that the video would pause and would not restart and, similarly, the caterpillar would continue to go slowly, until the level was above the threshold value. The formula used for the caterpillar was as follows: if the engagement level was above the threshold the caterpillar moved at 10*Engagement level +1 , if it was below this level then it moved at 1*Engagement level + 0.1. After the application had been designed, a third and final
participant took part in the pilot study; all three experiments were conducted on this participant. This calibration task was found to work well and interacted correctly with both the maze and video applications, however it is a simplistic technique and future work should seek to validate or improve upon it.

3.5 Experiments

After building the experiments and running the pilot study, the next step was to carry out a main study using the developed applications. Nineteen CS students and six non-CS students participated in the study. This resulted in twenty usable data sets (16 CS, 4 non-CS). The experiment took ~30–40 minutes to complete. Below is a full demographic breakdown of the participants from who data was successfully collected.

The experiments were run on a Lenovo IdeaPad U330p laptop (8GB RAM, Intel i5 processor running Windows 10) which participants had on a table in front of them. The first author remained in the room to monitor the headset’s signal as occasionally it would drop out, that is, one or more of the contacts would move out of ideal placement. After reading the information sheet the participants were instructed, and helped as appropriate, on the correct placement of the EPOC on their heads. Once a good signal was achieved (i.e. the 14 channels were detecting a good signal) participants took part in the calibration procedure. After an individual baseline had successfully been obtained through the calibration system, they then viewed the video, being advised to watch the video and that if it paused, to try to refocus on the task. After completing the video, the participants completed an engagement questionnaire. If needed, extra saline solution was applied to the sensors of the EPOC. The participants then completed the maze task and associated engagement questionnaire.

Twenty full data sets were successfully obtained over a period of four weeks. Six more participants took part but their data was either corrupt, too noisy or not obtainable. Four of these were female participants and the cause was most likely the amount and thickness of their hair prevented the EPOC’s sensors establishing a good connection. Another contributing factor may have been head size, that is, they had slightly smaller heads than other participants and so sensor contact may have been sub-optimal, this is down to the restricted movement of the sensors and flexibility available of the EPOC compared to medical EEG devices.

4. Analysis

Statistical analysis including frequency tests, t-tests and P-values were performed on the data collected. Due to the small sample size, and even smaller breakdown, anything with a p-value of less than 0.15 (i.e. Alpha = 0.85) was treated as potentially interesting. Although the sample size is small (n=20), some of the findings are
significant enough to warrant further investigation. It should be taken into account that this was an initial experiment and project and so the findings and methods can be built upon and improved in future studies.

A point worth noting is that the uninteresting nature of the video content and the production quality were almost unanimously confirmed in post-video response forms (19/20 participants responded to the question “What did you find that was boring or too easy?” (participants were instructed to answer this question and all others in regards to the task completed, either the video or maze) with comments that the video itself or the content was boring or easy). At the beginning of each experiment participants were told what would occur if they disengaged, but were not informed to try harder than normal to engage, just to watch the video/maze and try to re-engage if the video paused or the caterpillar slowed.

4.1 Calibration

Finding 1 – During the calibration task, participants who wear prescription glasses had lower average engagement levels during the fixation cross section

Participants who wear prescription glasses (n=6) had a statistically lower average engagement level during the minute in which they focused on a fixation cross than those who don’t (p value = 0.148). These same participants also had a greater range in their average values (eyes open/shut) than those who don’t wear glasses (p-value = 0.131). This tendency towards an effect may suggest something interesting such as, for example, that glass-wearers have learnt to concentrate harder when instructed than non-glass wearers but ultimately did not concentrate as well as non-glass wearers. Such a hypothesis would have to be tested for merit however.

Finding 2 – Participants over the age of 23 had a lower average baseline during the fixation cross section

Mature participants (over 23) had a significantly lower average baseline engagement level when staring at the fixation cross than participants < 23 years old (p-value = 0.115). This could be an interesting finding as mature participants seem to disengage with both the video and maze much less than non-mature participants.

Finding 3 – Those with previous programming experience had significantly higher ranges than those who had no previous experience

Those who have previous programming experience had a larger range (eyes closed compared to eyes open) on average than those with none (p-value = 0.057) during the calibration task. They also had a higher average level with their eyes shut, however this was not as statistically significant (p-value = 0.275).

4.2 Video Data

With regards to the video, related data initial analysis was focused on the number of times the video paused during the duration of the video. This indicated a drop below the baseline level recorded during the calibration task, referred to as a “drop” or “dropping” in the next few sections. Of the 20 participants, 7 watched the whole video without it pausing once. Of these seven participants, a brief summary of their breakdown is follows: 1 female; 6 males, 2 mature, 5 non-mature, 2 postgraduate students, 5 undergraduates, 1 non CS student, 6 CS students.
Across the whole population the mean number of drops was 12.55, the median was 9 and the mode was 0, the maximum value was 46 (see Figure 2).

Some points of interest from this data include the following:

**Finding 1 – Participants with lower baseline engagement levels paused the video significantly less than those with higher baseline levels**

On both tasks those who had lower engagement levels with their eyes shut (the value taken as a baseline) caused the video to pause less and had similar results with the maze game i.e. caused it to move slowly for less of the time. To compare the two groups, the average of the whole populations’ baselines was determined ($\bar{x} = 0.501407$) and the group was then split using this value. The baseline values were the unique values that were calculated for each participant during the calibration process. It should be noted that the sample size was small and the calibration task was crude thus more work is required to determine the validity of these findings.

All participants with a baseline below the population average ($n = 9$, referred to here as the Low Group (LG)) were compared with those above it ($n=11$ referred to as High Group (HG)). The LG dropped significantly less than the HG, with the LG pausing on average 2.333 times whereas the HG paused on average 20.9 times (p-value = 0.002). This is interesting as it could suggest those who are able to disengage before completing the tasks are more likely to succeed at the cognitive task. This suggests that learning meditative techniques and applying these before a cognitive task (test, studying etc.) can improve performance (Zeidan et al., 2010).

**Finding 2 – Participants who rated their Java level as lower paused significantly more than those who rated their level as high**

Participants who rated their Java level as 1-3 ($n=10$) (with 1 being “Very poor” and 5 being “Very good”) paused significantly more than those who rated it 4-5 ($n=10$). On average those who rated their Java level as lower paused 18.1 times with those rating it higher pausing on average 7 times (p-value = 0.0852). The video is designed for beginner programmers, and so in theory those with greater knowledge and experience of the language “should” find it less engaging. However, most of the participants responded that the video was boring. This finding also holds true if you remove those who have never programmed before; for four of the participants this was true and they rated their level as 1 (no-one rated their level as 2). If you take those participants ($n=6$) who rated their Java
level as 3 and compare it to those rated 4-5, those rated 4-5 paused significantly less, on average 19.33 times for level 3 participants compared to 7 for 4-5 (p-value = 0.125). It would be interesting to take this experiment and run it on those learning to program (possibly in two groups, one who want to continue CS and one just making up credits) to see if this holds for the target population.

**Finding 3 – Mature participants (over 23) caused the video to pause less**

Mature participants (n=4) caused the video to pause less on average than non-mature. Although not statistically different it was numerically quite different, with matures averaging 6.5 drops compared to 14.06 with non-mature. Further investigation in this area is warranted.

**Finding 4 – No difference in people whose self-perceived engagement level was low (1-2) or high (3-5)**

On another note, no statistical difference was found between those who responded to how engaged they were during this task with a 1-2 rating or a 3-5 rating (with 1 being “Not at all” and 5 being “ Completely/Always”). Those who rated how engaged they were as 1-2 dropped on average 11.4 times compared to 13.7 times with those who rated engagement as 4-5 (p-value = 0.732). This is interesting as it could mean that people felt they were more engaged than they actually were, or conversely that they were less engaged than they felt. From anecdotal evidence and observation, many participants felt they were focusing and engaging in the video but still caused it to pause, however all agreed the video was boring and unengaging. This could mean that even though they felt they were engaging, the EPOC and associated baseline engagement level did not indicate they were. It could also mean that people are poor self-reporters, however due to the limited number of studies into the accuracy of the EPOC it is hard to know which of these is more likely.

### 4.3 Maze Data

During the maze game if participants’ engagement levels dropped below their baseline level, the speed of the caterpillar dropped dramatically. The speed was constantly changing due to the level but dropped drastically if this occurred. This is again referred to as a drop.

Of the 20 participants, 7 played out the duration of the game without it dropping once. The profiles of the seven participants were: 7 males, 3 mature, 4 non-mature, 3 postgraduate students, 4 undergraduates, 7 CS students. Five of the 7 who didn’t drop during the maze task also didn’t drop during the video task.

The mean number of drops for all was 3.15, the median was 3, the maximum value was 10 (see Figure 3).

![Figure 3 – The number of times the caterpillar slowed (drops) during the maze game](image-url)
Finding 1 – Those with higher baseline engagement levels dropped (slowed the caterpillar) less than those with lower baselines.

As previously noted, a significant difference was found in the number of between the LG and HG. Those in the LG dropped (or slowed) on average 1.11 times compared to 4.81 times for those in the HG (p-value = 0.021).

Finding 2 – Those who rated their Java level as low (1-3) dropped more than those who rated their level as high (4-5)

Similar to the video data, those who rated their Java level as low (1-3) caused the caterpillar to slow more times than those who rated it high (4-5). On average this was 4.3 times compared to 2 times (p-value = 0.126). It is worth noting that those who rated their Java level as 1 (n=4, all the non-CS participants) had the highest average drops in the maze with an average of 5, compared to 3.83 with level 3 (6/20), 2.5 with level 4 (8/20) and 0 with level 5 (2/20). This is interesting and warrants further investigation to identify why this is observed.

Finding 3 – Mature participants (over 23) dropped less than those under the age of 23

Another similarity with the video data, with greater statistical significance, is that mature participants caused the caterpillar to slow less than non-mature. Matures caused it to slow on average 1 time compared to an average of 3.69 for non-mature (p-value = 0.149).

Finding 4 – No statistical difference in those who found the maze more engaging, interesting or challenging than those who did not

In line with the data collected from the video task there was no statistical difference between participants who found the task more or less interesting, more or less challenging or more or less engaging.

4.4 Maze vs Video

In terms of survey feedback for both the maze and video, participants were asked to rate three survey items from 1-5 (see Lockwood & Bergin 2016 for details) as follows: (1) I was challenged by this task, (2) This task was interesting to me, (3) This task engaged me. Interestingly and insightful for future work, all participants found the maze more challenging, engaging and interesting than the video (see Figure 4).

Figure 4 - Number of people who rated the maze/video more or less in each question.
5. Conclusions
This project has shown that a consumer grade EEG device can be used to track and respond to a learner’s engagement level. This background work could be expanded into a more robust system which could help improve both a learner’s experience and their engagement.

The findings from the experiments provide interesting insight into student engagement whilst watching tutorial videos. From the analysis the most significant findings that warrant further investigation are:

- Those with lower baseline engagement levels caused both the video and the maze game to drop less than those with higher baseline levels. This could support evidence that practising meditative methods increases cognitive performance.
- Those who rated their Java level as low caused both the video and the maze to drop more than those who rated their level as high. This is interesting especially given the nature of the video (a beginner Java tutorial) and it would be useful to investigate this with similar tutorial videos.
- Mature participants (those over 23) caused the video and maze to drop less than those under the age of 23. This is another interesting finding as potentially it could show, with further and more intensive study that those over the age of 23 may find it easier to engage with cognitive tasks than those younger.

As stated, given the small sample size, the findings, although interesting, require further investigation.
6. References:


