INF2260

Satisfaction measured

Mats Jørgensen, Tina Bakland, Espen Thorsen
Universitetet i Oslo 25/11/2012
INF2260

Table of Contents

About the group

The group consists of Espen Thorsen, Tina Bakland and Mats Jørgensen – all BA students on the third semester of Informatics: design, use and interaction.

The collective experience of the group encompass novice to intermediate programming experience, game design and general teamwork. None had any prior knowledge of the use of Emotiv and EEG measurement.

Design brief

Our assignment was to work with Schlumberger using the Emotiv EPOC EEG technology, with their software Petrel, which allows users to interpret seismic data, perform well correlation and build models and maps over reservoirs.

Schlumberger is a multi-national corporation that describes itself as "*the world's leading oilfield services company supplying technology, information solutions and integrated project management that optimize reservoir performance for customers working in the oil and gas industry*"[1]

The initial brief was:

"Schlumberger wants to measure emotional indexes while executing software testing on ribbon and traditional modes.
Our goal is to map emotional responses to different types of stimuli, to measure this we will use the Emotiv EPOC, and see if we can use this information to determine user satisfaction for different operations in Schlumberger`s Petrel software."

However, because of unforeseen events described later in this report, we had to depart from the original brief and change our plans:

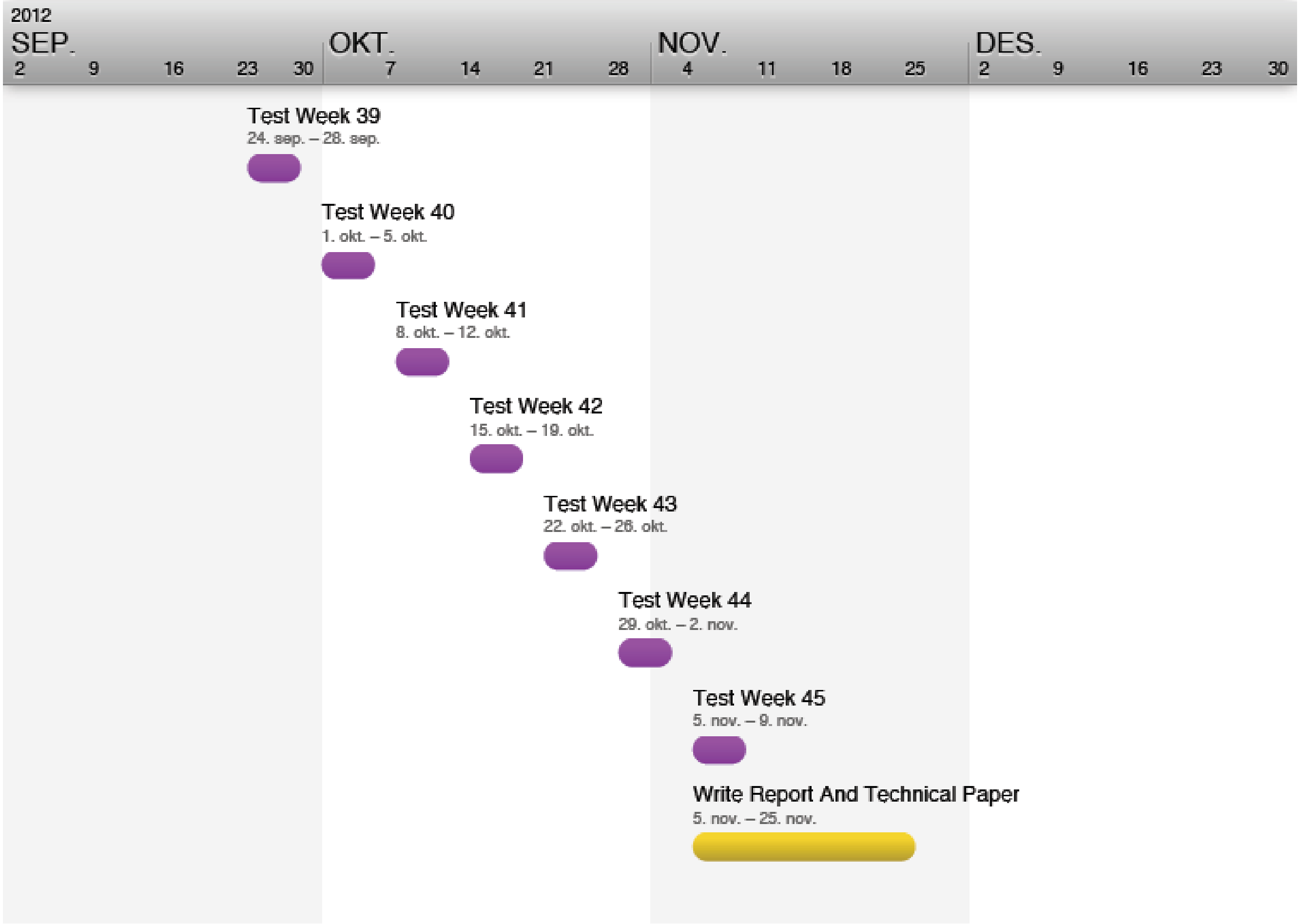
To test the Emotiv headset itself to see if the EEG-measurements can be used to obtain any coherent

1 http://www.slb.com/about.aspx, "About us", downloaded 23.11.12

information at all about the users state of mind in a testing situation on a more general level. This plan was again changed towards the end of the project because of further unforeseen events which we will describe in further details in this report.

Project plan

This is the project plan we made for this semester. The weeks before these test weeks were used for the experimental phase.



Test week 39: Here we planned to do tests on calibration and reception in Mind Workstation.

Test week 40: Will do tests on light/dark room, frustrating game and music.

Test week 41: Game test with 3 different levels of difficulty and music test.

Test week 42: Image test and game test.

Test week 43: Image test and game test continues.

Test week 44: Taste test and physical stimuli test.

Test week 45: Taste test and physical stimuli test continues.

Week 46 – end of project: Finish report and presentations.

In week 44, when we were going to start the taste and physical stimuli tests, we could no longer get any data from the headset. We used the next 3 weeks working with the usability tests in Petrel instead.

The Technology

Emotiv EPOC was developed and marketed by the Australian company Emotiv Systems, and is, according to them, based on "[...] *the latest developments in neuro-technology* [...]".[2] It was originally designed to be used with computer games, allowing players, in theory, to control computer games by using nothing but their minds.

The Emotiv headset consists of a set of arms made of flexible plastic, where each tip houses a copper-plated saline sensor covered with a felt pad. There are 16 of those sensors (making 8 pairs), arranged in a pattern that, in theory, corresponds to key brainwave activity. The sensors, like in standard EEG-equipment utilized on hospitals and in research, captures electrical activity from the surface of the brain (the voltage fluctuations caused by firing neurons). Those signals is sent wireless from the headset to a computer via an USB receiver, and further analyzed by a (secret) algorithm in the Emotiv software. The interpretations of those signals can then be read and utilized by the operator, making it possible to e.g. link different signals to different operations on the computer, interpret the users brain activity or even link facial expressions to specific actions. Furthermore, the headset is fitted with a gyroscope, potentially allowing for the use of broader motions of the head, like tilting and nodding.

To function properly, the sensors must be placed on exactly the right place on the scalp, and the felt pads must be wet with a sterile saline solution for better contact and signal purity. Also, the signals interpreted by the software must be allowed time to settle down (as the algorithm is kept secret, it is unclear what this entails, but it could have something to do with pattern recognition and "garbage collectors" in the code). On a further note, the code interpreting the signals is based on tests specifically designed with gamers in mind. Such a test could, for example, consist of the test subject

2 http://www.emotiv.com/store/hardware/epoc-bci/epoc-neuroheadset/, "Epoc Neuroheadset", downloaded 23.10.12

playing two variants of a standard first person shooter game – one with a lot of action, and another without action at all. The tests seem to be rather poorly designed from a scientific view, not necessarily lacking in scientific methodology, but more lacking in the understanding of how the human mind works. They were too simple for the complexity of the nature of the technology, it seems, only allowing the developers to base their interpretations of the complex human mind on a few (dynamic) emotions brought on by very limited input.

The mass-marketed bundle consists of the headset with sensors, wireless connection via USB, and a software package allowing for training, playing and development. Two versions of this bundle exists; the EPOC Emotiv and the EEG Emotiv. The only difference between these bundles seems to lay in the software, with the EEG-version offering more freedom in further development for the consumer. The hardware and materials in both bundles is identical.

The unique aspect of the Emotiv technology lies not in the technology itself, but the mass appeal it receives from its low cost. The EPOC bundle ships for 300USD, and the EEG edition ships for 750USD. When compared to the much higher cost of standard EEG technology, which can be in the hundreds of thousands USD, the Emotiv stands out as an interesting opportunity for both consumers and researchers to utilize the "new" technology on a grander scale than what is possible today.
The low cost, however, seems to reflect some serious drawbacks in quality – something we will discuss later in this report.

Affective suite and Mind Workstation

Epoc Emotiv has three different functionalities that utilize different forms of input

- Cognitive suite: links the users conscious thought to an action that will be performed on a 3d model
- Expressive suite: Detects the users facial expressions, a particular facial expression can then be linked to performing a specific action
- Affective suite: Measures the frequency of the user's brainwaves in order to determine their mood. This is the functionality we used in our experiments.

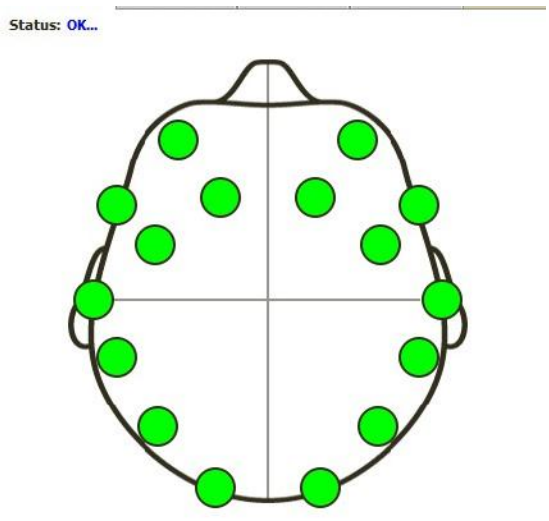
The purpose of the affective suite is to measure the emotions of the user; this is achieved by measuring the users brainwaves, the headset measures four types of brainwaves, beta waves, alpha waves, theta waves and delta waves.

The affective suite has four emotional states that are calculated depending on the frequency of the different brainwaves, “Excitement”, “Frustration”, “Meditation” and “Engagement/Boredom”, Emotiv admits that the names of the emotional states may not accurately reflect what the emotion is, for instance if the user is startled or surprised this may increase the value of the “Excitement” state.[3]

As mentioned above the brainwaves are measured through 16 saline sensors on the headset and then put through an algorithm in the software; out of this algorithm each emotional state is given a numerical value between 0-1.

To achieve precise readings of brainwaves the software relies on good signal quality from the headset. The signal quality is grouped into five different categories:

- Black: No signal
- Red: Very poor signal
- Orange: Poor signal
- Yellow: Fair signal
- Green : Good signal



Sensor map, showing ideal signal quality

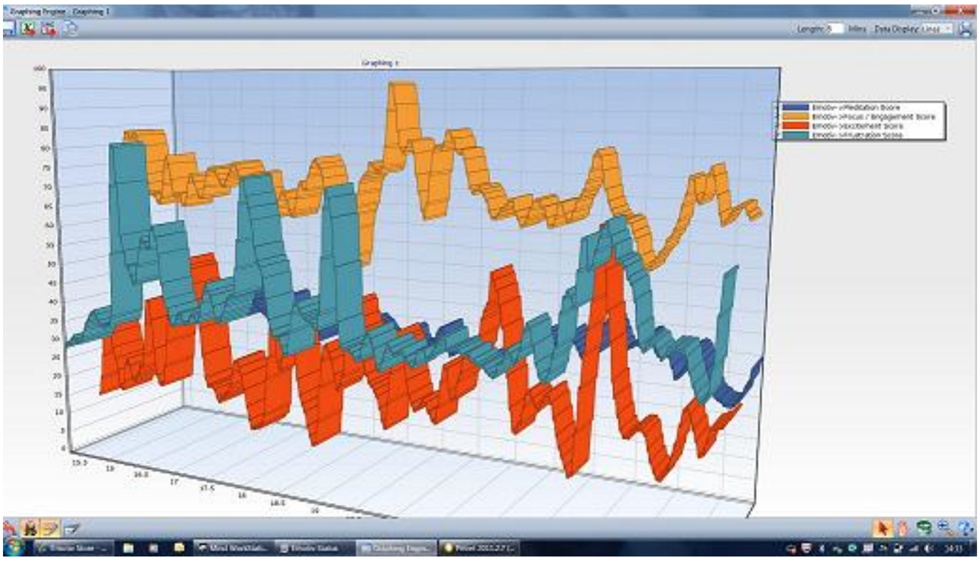
3

http://www.emotiv.com/forum/messages/forum4/topic556/message3099/?phrase_id=479774#message3099 “

The Affective suite in the Emotiv software package presents a graphical representation of each emotional state (see picture below), but we quickly saw that it was difficult to get the exact numerical value of each emotion from these graphs, having the exact value is crucial in order to analyze and interpret the collected data correctly.

Affective suite in action

To solve this problem we did some research online to find another program we could use to collect and interpret that data from the headset, we eventually found a software called Mind Workstation which is developed by The Transparent Corporation, this software allowed us to get the exact numerical value for each emotional state throughout a testing session and also made it easy to export this data to Microsoft Excel to further work with and analyze.



Graphing functionality in Mind Workstation

As mentioned earlier the Emotiv EPOC was originally intended to be marketed towards the gaming industry, this is reflected in how these emotional states were trained and tested, several volunteer subjects were exposed to different types of stimuli while wearing the headset, and measuring other biometric monitors such as heart and breathing rate, in order to evoke specific emotional reactions. The stimuli mostly consisted of playing a FPS (First Person Shooter) game, the Engagement/Boredom state was measured by looking at the contrast between being involved in a firefight with a manageable number of enemies and walking around in an empty landscape with no enemies, while the frustration state was measured while the user was involved in an unmanageable firefight and the controllers were faulty (the user would press left, but the character in the game would move right). The Excitement state was also based on data collected during these gaming sessions.[4]

Who can use the Emotiv headset?

After testing the Emotiv headset, we have found that the headset does not give as good signals when the user has a smaller than average head size. The headset slips off the head and the sensors do not make sufficient contact with the skin. This leads to poor signal quality. The users we have tested who have had the best signals, have been those with larger than average heads.
Another important factor is the length of the hair of the user. With very thick, long hair it is more

4

http://emotiv.com/forum/messages/forum15/topic612/message3425/?phrase_id=475873#message3425), downloaded 23.10.12

difficult for the sensors to make contact with the skin of the head. Short hair is the easiest to work with. It is possible to get good signals with long hair, but it usually takes some amount of time to part the hair and find the correct placement for the sensors.
The ideal user of the Emotiv headset is a person with an average/large head size and very short hair. Users with long, thick hair or small size of the head must expect to spend more time adjusting the headset before getting satisfactory quality of the signals.

This table shows how long it took for some of our participants to get good signals.

Test person	Hair length	Head size	Time before good signals
1	Long	Small	Not achieved
2	Short	Large	45 seconds
3	Short	Small	07 min 10 seconds
4	Long	Small	16 min 02 seconds
5	Short	Average	04 min 11 seconds

Experimental phase

The first thing we did when we got the headset was to play around with it and figure out how the different functionalities worked, then we focused in on the Affective suite since that was the functionality we would be working with throughout the project.
Everyone in the group took turns trying the headset on, this produced discouraging results as none of the group members were able to get stable enough signals to where the Affective suite could continuously measure the mood of the user. Due to insufficient signal quality the graphs in the Affective suite would drop out every few seconds, this produced unreliable results.
After doing some reading on the Emotiv forums we saw that it was normal for the signals and readings in the Affective suite to be unreliable the first few times a new user wears the headset because the software needs to get used to each person's brainwaves.
The first time a user puts on the headset the user is asked to create a user profile, at this point the software has no reference values for the newly created user profile to determine what the mood of the user is, so it assigns each emotional state a numerical value based on a baseline value that represents the average user. As the user wears the headset over a longer period of time and the headset stores more data on the user profile, the scale and values for that user profile are

continuously updated in order to get more reliable results for the current user profile.
In other words the current results are based on earlier results with the same user profile that are continuously updated as new data comes in, the more data the software has on a user profile the more reliable the current readings are. The same principle goes for each emotional state, so the first time the software detects a specific emotion in a user profile this causes a much exaggerated result in the Affective suite graphs, but as the software collects more data on that same emotion it has more to compare the current readings to, so the results settle down and become more reliable.
Due to the amount of time it takes for the headset to calibrate to each individual user we decided that it would be best to use a within-group testing approach and focus on a small group of testers for the future tests in the project period.[5]

Testing phases

After the experimental phase, we consulted with Schlumberger and decided that prior to testing the Petrel software with the Emotiv technology, we should establish thorough knowledge whether or not we actually can get reliable data using the EPOC. This, we found, should be done through a series of smaller tests where test subjects are exposed to stimuli in order to evoce a specific emotional response, we should also increase the number of test subjects in order to get statistically significant data.

Phase 1:

The first tests were designed to reveal eventual noticeable manifestations of emotions when test subjects were exposed to simple stimuli. We used five random test subjects for a series of simple within-group tests. The test subjects were all young (20-30), but had nothing specific in common.

We fitted the test subjects, one by one, with the headset, getting mostly good signals on all of them. After the fitting, we let them play with the functionality of the Emotiv's training program (controlling a 3D cube) while we waited for the EEG signals to stabilize somewhat. After reading

5

http://emotiv.com/forum/messages/forum16/topic1436/message8443/?phrase_id=475873#message8443, downloaded 23.10.12

the Emotiv forums, we found that we had to allow at least 20 minutes for this stabilization to occur. According to the creators of Emotiv, several hours of stabilization would be preferable to achieve trustworthy readings. This, however, would be impossible time-wise in a research situation like ours.

We formulated a hypothesis, which we used throughout the project:
Our hypothesis was: Different stimuli will have reliable and significant impact on the emotional readings
Our null hypothesis was: Different stimuli will have no reliable or significant impact on the emotional readings

After 20 minutes of stabilization, we exposed the subjects for the following stimuli:

- Darkness
- Music and noise
- A challenging computer game

1: The subjects were placed alone in a room. On irregular intervals, the lights were turned off, and then on again after about half a minute. We observed the graphs in the emotion suite (standard Emotiv software) and noted the fluctuations at the moment of changes in light levels.

2: The subjects listened to classical music for about a minute, before we switched the music with a piece of non-musical noise. The idea was to first see if the calm, classical music could induce a positive effect on the “meditation curve”, while negatively affecting the “frustration curve”. Then, we speculated that the sudden onset of noise would have the opposite effect. As with the light-test, we observed the signals and noted changes.

3: Here, the subjects were given the task of playing a challenging computer game without instructions. We observed and noted the changes.
During this test we expected the frustration curve to increase.

Reviewing our notes from the testing, we could find no consistent evidence to prove our hypothesis. However, our collected data was lacking – being only hand written notes from subjective observations. In the future tests, we focused more on collecting evidential data.

Phase 2:

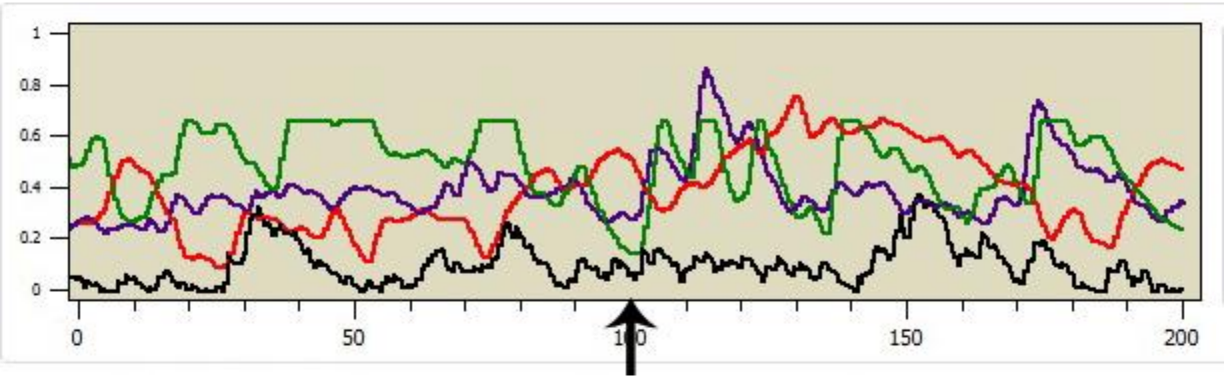
In this phase, we continued the tests from phase 1, with the exception of the light-test and the addition of storing screen captures of the graphs for future analysis. Five subjects were chosen randomly, and do not represent special groups, background, ethnicity, sex or occupation.

The tests were, as in phase 1, as follows:

1: The subjects listened to classical music for about 1 minute, before being exposed to intense noise.

2: The subjects played a challenging game without instructions.

This is an example of the readings during these tests:



The arrow indicates the change from classical music to loud noise. Since we could get no numbers to work with, we were left with the only option; to interpret the data visually. As you can see there is a sharp increase in the frustration curve when the music is changed, but about 30 seconds earlier (while the classical music was playing) the frustration curve is at the same level as after the change and the curve also goes back down during the time the test subject was exposed to the noise. Looking at the results we could find nothing indicating reliable changes in emotional state – although we should stress the fact that none of us have any knowledge in the field of neuro psychology. Again, we found no evidence to prove our hypothesis.

Phase 3:

Finding that the standard Emotiv software was not enough to analyze the data, we found and installed a far more extensive software called "Mind Workstation" - designed to conduct tests and record data with EEG-equipment. As this program had good compability with the Emotiv hardware, we found it to be perfect for our needs. The most valuable function was the ability to export the data from the readings as Excell- or XML-files, giving us access to an extensive set of numbers (as the program updates the reading each every second, we got a set of numerical values from each of the emotional state for every second, ordered in collumns and exported) for detailed statistical analysis.

This time, we designed the tests more carefully, focusing more on visual stimuli. We wanted to find out whether we could evoke stronger emotions with the use of images, and whether those potential emotions would show up on the readings. We also designed a more comprehensive game-test where we would measure eventual changes in the readings as the subject embarked upon several different grades of difficulty when playing a game. Five test subjects were chosen randomly, and again do not represent special groups, background, ethnicity, sex or occupation.

The tests were as follows:

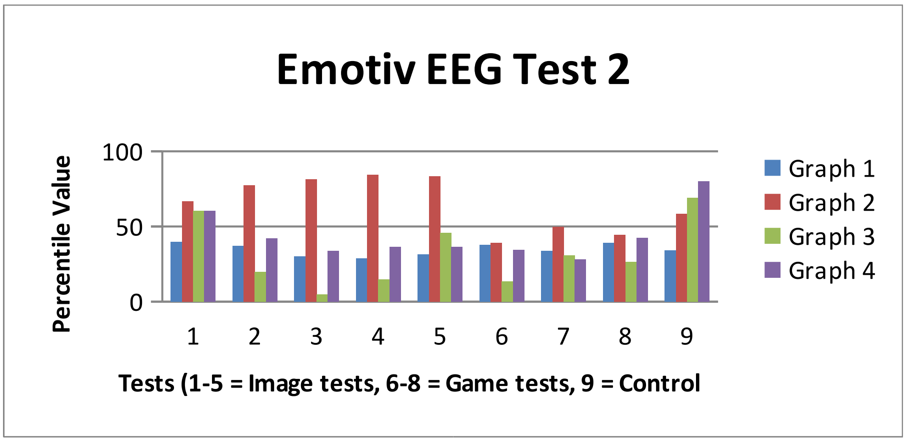
1: Subjects were shown a slideshow of five images, each lasting ten seconds. The images consisted of:

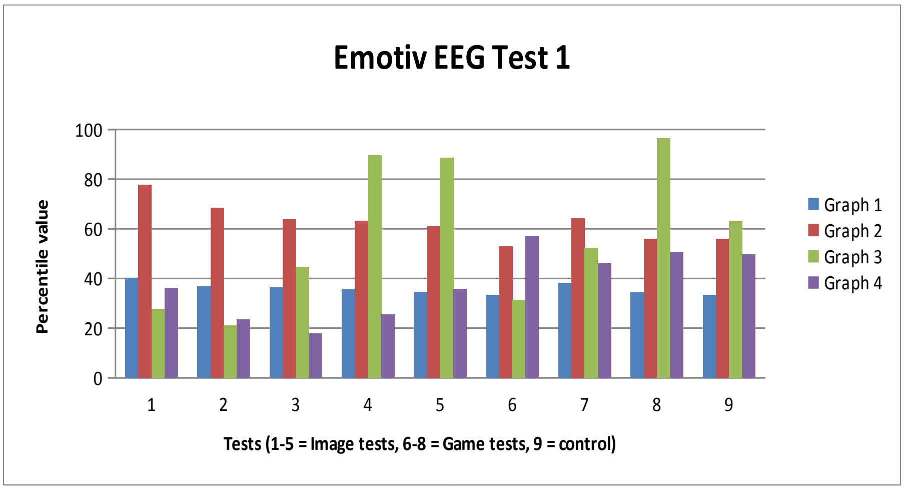
- A child being beaten by the police
- A cold beer
- A cute kitten
- A funny fase
- A calm and exotic beach

As the subject watched the images, we noted the timecode when the pictures changed for future analysis.

2: Subjects were to play the game of Tetris on three different levels of difficulty (where the difficulty was merely changes in speed of the falling bricks). Again, we noted the timecode with each change in difficulty.

After the tests, we sorted the data in Excel, applying standard statistical calculations and making diagrams showing changes in each emotional state with each test on each test subject.
We could find nothing indicating reliable readings – only mostly random numbers. This could possibly stem from the fact that we only allowed 20 minutes for the algorithm to “get to know” the subjects.





Each graph represents an emotional state. The control was signals captured in a 3-minute time span before testing started.

If we compare the two exampls above, we can find no corresponding behavior between them. This is true for all the results we collected, indicating that a) emotional states are unique for each subject and/or b) the Emotiv EPOC does not produce stable enough data.

In either case, the results would be impossible to interpret on a general basis such as measuring satisfaction with use of a piece of software. One could perhaps get a deeper understanding of the measurements by consulting an expert in neuropsychology, but this would probably mean that a longer stabilization time would be necessary.

Phase 4:

After talking to Alma about what we had done so far in the project period we were advised to move away from the stimuli type tests and instead return to the original assignment, to focus whether or not the EPOC could be used to measure user satisfaction in a software usability test.

This phase was designed to record potential changes in the readings as subjects used a complicated 3D software with little or no instructions. The program to be used was Blender 3D – a software widely known to be extremely difficult to learn even for experienced 3D artists. We chose this program as a substitute for Petrel, as we found it too difficult to get subjects to come to Schlumberger contra the more central and accessible UiO. Blender 3D, being an advanced program for creation of 3D models, is the closest free alternative we found to resemble the functions of Petrel.

The subjects were to try several basic operations, such as creating a cube, moving the cube, rotating the cube and add color to the cube. These are operations that to some degree resemble basic operations in Petrel.
However, when we were to conduct the tests, we found the Emotiv headset to be in a state of extreme wear. The plastic had expanded, no longer fitting the subjects' heads (making the placement of sensors impossible), the sensors had started to rust and could no longer deliver the signal quality needed for successful EEG recording, and one of the sensors was missing a part. The same state of disrepair applied to the headset owned by UiO, making any attempt to conduct further experiments impossible.
Therefore, we had to cancel the experiment and rethink our approach to the whole project.

Usability testing with Petrel

Test plan

After the Emotiv headset stopped providing us with sufficient signals for testing, we talked to Alma to find a way to continue our project. We were advised to conduct a usability test on Petrel without using the headset.

According to the customer, some users of Petrel find it frustrating and far too complicated to use. Our objective was to find the operations that caused the most frustration among the users.
We wanted to test a typical workflow in Petrel. This would help us find which operations caused most frustration for the everyday-user of the software.

Test environment

Since we were not allowed to install Petrel on our own computers, we had to conduct the usability test at Schlumbergers offices at Røa. We were given a meeting room and a computer with Petrel to work with while we were there. They also provided us with a camera to document the tests. We installed a mouse tracking software on the computer to gather data from the tests including number of mouse clicks and a heatmap of where on the screen the user had clicked the most.

Participants

Our first test plan involved experienced users, people who work with Petrel on a daily basis. Unfortunately it was not possible for Schlumberger to provide us with test participants, so we had to perform the tests on ourselves and the members of the other Schlumberger-group, Mind Control. The target group was changed from experienced to novice users and the tests had to be changed accordingly.

Test materials

Because none of us had any knowledge of Petrel, we were given a demonstration and explanation of how the software is to work with. Because the software is so extensive, we only focused on a small part of it when we designed the tests. This part we called the well section, where we could work with different well tops. These are the tasks the test participants had to complete in the test:

- Select wells
- Create well tops
- Select a well top
- Move well top
- Flatten well top
- Delete well top

Test sessions

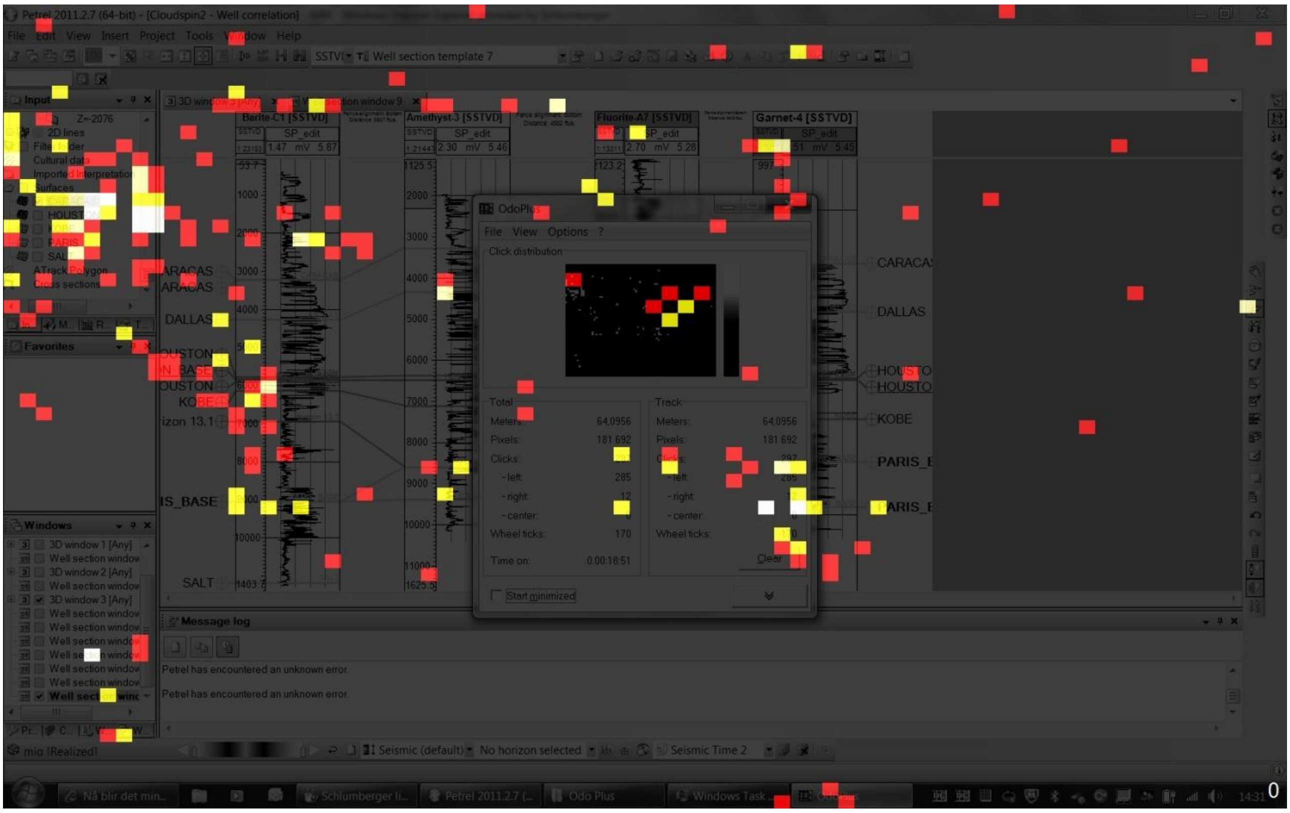
We first performed the test ourselves. Those in the group who did not do the test were filming and timing the user. Although we had seen the tasks done by someone who knows Petrel before, it was difficult to remember how it was done. All the test participants spent around 20 minutes to complete all tasks. After everyone in the group had performed the tasks, we had a member of Mission Control do the same test. This user had not seen the demonstration and had no knowledge of Petrel. The user struggled to complete any of the tasks without help and after 11 minutes we stopped the test. We did not see the point in helping the user with every task to complete the test. We did not test any of the other members of Mission Control after this session.

After the test was complete, we asked the users to fill in a form and explain what they found the most difficult and frustrating of the operations in the test.

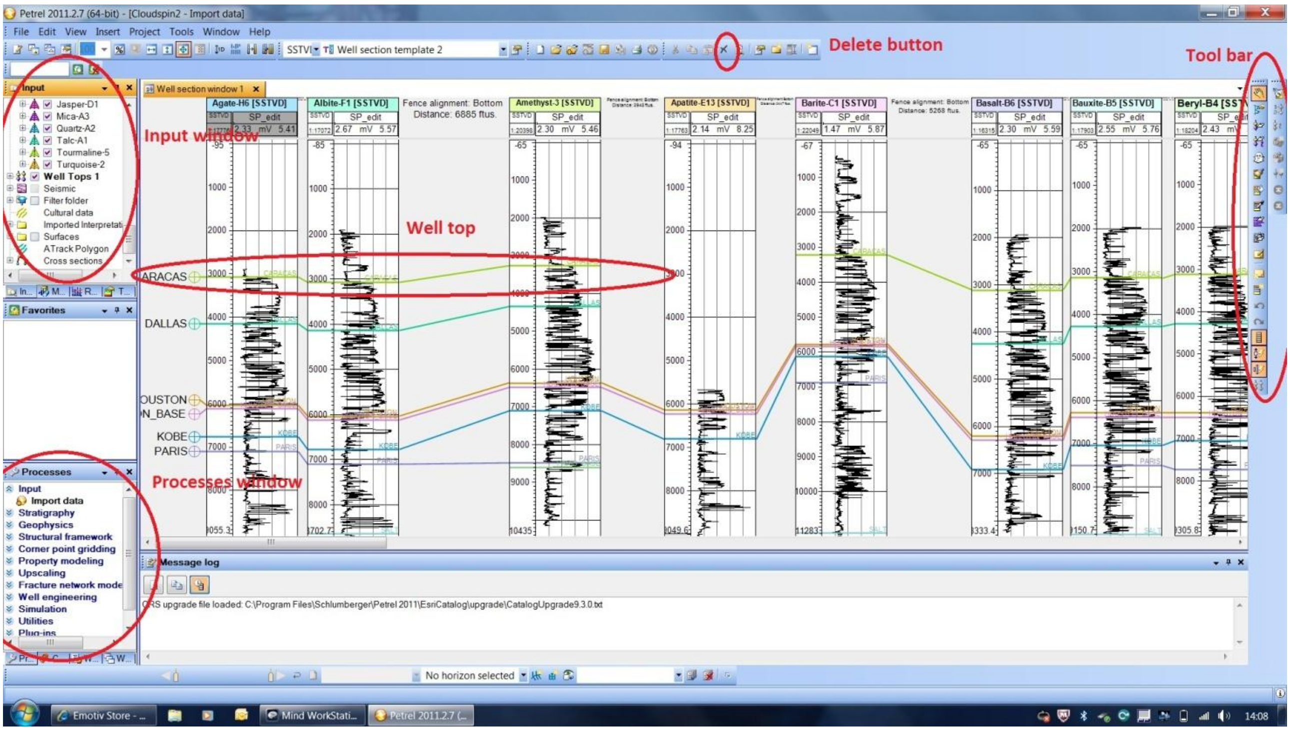
Analysis of data and observations

The tests we designed were not useful with participants with no prior experience with Petrel or geology. When we tested the user from Mission Control group we had to interrupt after a while. We would not get any information on how the test participant felt about the workflow in Petrel when we did all the work for them. It is not possible for a user to comment on a workflow when they know nothing of the software or the meaning of the figures on the computer screen.

When the participants from our group performed the tests, we were all able to complete all tasks because we had seen it done before and played a bit with the software beforehand. The Mouse Tracking software we used showed how many clicks were used to complete all tasks. The heatmap shows where the user clicked the most on the screen.



Heatmap from one of the tests



Petrel workflow

After the tests, we asked the participants to rate the software on first impression, GUI, how intuitive it was and its learning curve. 1 is the lowest score and 10 is the highest. We also asked them to comment on functions they found particularly confusing during the test.

The test participants would like to see the following changes in the software:

1. An undo button
2. Larger Input window
3. Select well top by double clicking
4. Flatten well top by right clicking
5. Delete well top by right clicking
6. Make active well top more visible
7. Created well tops should be instantly visible
8. The possibility to choose a well top from the flatten-menu

9. Searching for a well top

1. All participants would like to have an undo button for when they made a mistake. This would save time in the workflow.
2. Some users found the Input Window too small. This is where the user can get an overview of the project and see all the well tops. A lot of the operations use this window and the user needs to scroll a lot up and down to find the file/place he is looking for. As it is used a lot in the workflow, this window should be bigger.
3. The participants missed the possibility to select a well top by double clicking it with the mouse. The user had to select a well top in the Input window instead of clicking the well top directly.
4. The way to flatten a well top in the current version of Petrel takes 5 operations. This should be done simply by a right click menu.
5. The user should also be allowed to delete a well top from a right click menu. Now the user have to choose a well top from the input menu and click a somewhat hard to find delete button on top of the screen.
6. When the user select a well top, the line of the active well top is a bit thicker than the other well tops, but it is still hard to distinguish. It should be more visible, perhaps with the use of color.
7. When you create a well top, you have to go to the Input window and check it before it comes up on the screen. Some of the test participants would like to see it instantly on the screen.
8. When you are in the Flatten-menu, the only way to choose which well top to flatten is by going out of the Flatten-menu and click on a well top in the input window. This is cumbersome and it should be possible to choose a well top within the Flatten-menu.
9. If you are working with a big project in Petrel, you can end up with a huge amount of well tops. In this case it would be useful with a search function to find the one you are looking for in the Input window instead of scrolling up and down.

Test results:

	Number of left mouse clicks	Number of right mouse clicks	Time spent (seconds)
User 1	467	17	1180
User 2	285	12	1264
User 3	389	20	1325
User 4	189	5	
Mean	332,5	13,5	1256,333333
Median	337	14,5	1264
Mode			
Range	278	15	145
Standard dev.	121,286163	6,557438524	72,8033882
Variance	14710,33333	43	5300,333333

There was a big range in the number of mouse clicks by the different users, this is also reflected in the high number of the standard deviance; this was attributed to the different strategies employed by each user. User 1 systematically went through all the different menu options in order to find the functionalities needed to complete the given tasks, while user 2 and 3 spent more time thinking about where it would be logical for the functionalities to be located and then clicked on these menus instead of searching through every option.

After calculating Pearson's r for the relationship between the total number of mouse clicks and the time spent solving the tasks, Pearson's r came out to – 0.4804622849, we saw that the correlation between mouse clicks and time spent solving the tasks was not significant since the correlation only becomes significant above 0.5 or below -0.5, this means that an increase in the number of mouse clicks will cause a small decrease in the time spent solving the tasks.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	First impressions	How intuitive	GUI	Learning curve
User 1	5	3	3	4	2	3	2	1	2	3
User 2	5	2	4	4	3	3	2	2	1	2
User 3	4	3	3	4	3	2	2	1	3	2
User 4	1	1	1	1	1	1	1	1	3	1
Mean	3,75	2,25	2,75	3,25	2,25	2,25	1,75	1,25	2,25	2
Median	4,5	2,5	3	4	2,5	2,5	2	1	2,5	2
Mode	5	3	3	4	3	3	2	1	3	2
Range	4	2	3	3	2	2	1	1	2	2
Std dev.	1,89297	0,95743	1,25831	1,5	0,95743	0,95743	0,5	0,5	0,95743	0,816496581
Variance	3,58333	0,91667	1,58333	2,25	0,91667	0,91667	0,25	0,25	0,91667	0,666666667

As seen in the chart above Petrel received very low scores in all categories, this is most likely caused by the fact that all our test participants were novice users and Petrel is a very complicated software that takes time to learn and get used to. If we tested experienced Petrel users the result might have been very different.

Error sources

A large source of error in this data set is the data from user 4, since user 4 did not complete the usability test his data numbers are very low compared to the other users; this drags the measures of central tendency down compared to what they would have been if he had completed the test.
Also four users is not enough to get statistically significant data, especially not when only three of them complete the usability test

Conclusion

After analyzing both the notes, graphs and numerical data collected during the course of the project, we could find no coherent readings supporting our hypothesis. Each individual reading produced radically different results with the same set of tests, showing that no generic definition of emotions can be used in evaluating such readings. This is further supported by the fact that Emotiv admits that the curves not necessarily represent what they promise, but could mix up or react to something

different entirely.
We sent all our numerical data to our representative at Schlumberger for a second opinion, but he came to the same conclusion; the data is unreliable and incoherent.

Seeing as the interpreted results merely reflects reactions to simple stimuli, the task of measuring more complex reactions to an abstract situation like operating complex software seems impossible with Emotiv EPOC. It should, however, be pointed out that we have not allowed sufficient time for the signals to stabilize during our tests – only allowing an average of 20 minutes instead of the recommended 2 hours. We have no data to show the potential gain in accuracy when allowing more stabilization time, and can therefore not bring any insight regarding this, though some signs indicate that this probably would not matter in the long run if we take into account the quality of the materials and the limitations in the number of sensors and lack of amplifying means.

We also contacted a researcher in the field of neuro science[6] and inquired about the basic use of EEG in clinical settings, and learned that the equipment used in such situations differ greatly from the EPOC in several aspects:

- They use 32-64 sensors fitted in a tight cap
- The signals are amplified with an external amplifier
- The signals are ready to read immediately, requiring no “stabilization time”

Furthermore, the Emotiv headset was not built with heavy use in mind. It expanded after a while, making it difficult to stay in place on the users head. This could be improved by making the device in a more endurable material than plastic, maybe aluminium. The sensors should have been made in a material more susceptible to water (gold would be a good option), as they quickly started to rust. The feltpads were not very durable either. After a moderate amount of use, they often fell out of the headset, making fitting difficult.

The usability tests we performed with Petrel would give us better results if we had tested experienced users of the software instead of ourselves. We did, however, find several flaws in the software and possible ways to improve them.

---

6 Lasse Bang, Researcher/ PhD Candidate, Regional Eating Disorder Service (RASP), Oslo University Hospital, Ullevål

The Emotiv headset might be a fun and exciting device for someone who is going to use it for games – with the other functionalities it offers (cognitive, expressive and the gyroscope). For someone who needs an EEG device for research, we would not recommend Emotiv. It is too unstable, it takes too much time to conduct a single test, it is lacking in features and power, it's not very duarble and it might not even interpret the brainwaves correctly. Perhaps in the near future, the company will improve the functionality of the Emotiv headset sufficiently for this kind of usage, making it a good, affordable option for researchers who want to work with EEG. If so, it is probably advisable to either consult an expert in interpretation of EEG-signals and/or design the tests as a combination of standard usability testing and EEG.

Until such an eventuality occurs, the Emotiv EPOC probably can't be used as much more than a toy as it stands today.