Title

Guido Zuidhof
Supervisors:
Radboud Universiteit Nijmegen
guido.zuidhof@student.ru.nl

December 17, 2014

Abstract

1 Introduction

1.1 Cognitive load

1.2 Physiological response to cognitive load

How a human body responds to cognitive load/stress.

1.3 Determining cognitive load

How this cognitive load could be determined, i.e. by measurement.

1.3.1 Measurement of physiological response

Measuring heart rate, breathing, eeg, gsr, etc.

With an increasing variety in consumer

Galvanic Skin Response

1.3.2 Classification

How sense can be made from this data

1.3.3 Applications

How this can be used.

Learning Most applications and research regarding cognitive

Usability testing Cognitive load measures, and more generally stress and arousal measures, are used by human computer interaction (HCI) developers and researchers to evaluate the usability of software or systems [1] [4].

Schmutz et al. [4] investigated cognitive load measurement in web applications. Concluded was that cognitive load measures can be used as a ẗvaluable additional measure of efficiency", which strongly correlated to general satisfaction of the user.

1.4 Research aim

Using features extracted from GSR data to determine cognitive load, data gathered from experiment.

1.5 State of the art

Galvanic skin response readings have been studied in regard to cognitive load.

Shi et al. [5] investigated the possibility of using GSR as an index of cognitive load during the use of unimodal and multimodal versions of the same interface. It has been shown that multimodal interfaces "support the user in managing cognitive load" [3], and thus multimodal interfaces have a lower corresponding cognitive load.

Shi et al. found that GSR values increase when cognitive load levels increase across all subjects, and conclude that GSR can be used to "serve as an objective indicator of [a] user's cognitive load".

2 Methods

2.1 Experiment

The experiment is mostly based on the arithmetic task experiment conducted by Nourbakhsh et al. [2]. The experiment involves adding up four numbers, displayed one by one, and selecting the right from three possible answers.

There are four difficulty levels of this task. Difficulty one involves adding up binary numbers (1 and 0). Difficulty level two, three and four consists of numbers with a length of one, two and three digits respectively.

The subject is first shown an amount of stars corresponding to the difficulty level for 8 seconds, after which the numbers are shown one by one for 4 seconds. Three answers are then presented on the screen, of which one is correct. The subject then clicks the correct answer using a standard computer mouse. There was no time limit for selecting an answer.

2.1.1 Apparatus

The task was displayed in a 15.6" laptop screen, input was given using a standard computer mouse. Two galvanic skin response sensors were used in this experiment. The participants wore both at the same time on the left arm.

2.1.2 Participants

Six participants took part in this experiment, five male and one female. Their ages ranged between 21 and 25 at the time of the experiment. They were not compensated for their participation.

2.1.3 Affectiva Q Sensor

This sensor is embedded in a wrist band, it uses a dry electrode. The sensor side of the wrist was placed on the bottom of the wrist. The sampling rate of this sensor was set to 8 samples per second.

2.1.4 BIOPAC MP30

The BIOPAC MP30 was used in conjunction with finger GSR sensors. This sensor wraps around two fingers, in this experiment the index and ring finger. It uses gel electrodes, Grass EC33 Electrode Paste was used.

The sampling rate was set to 1000 samples per second.

2.1.5 Task

2.1.6 Experimental design

The task consisted of 12 arithmetic tasks, with 4 difficulty levels. Three of every difficulty level were completed by the participants, in a random order.

The task consisted of adding up three numbers, and selecting the right answer from three options. The numbers to add were showed one by one, for four seconds. Before a task starts, a number of stars is shown equal to the difficulty level, for eight seconds. The tasks followed eachother without breaks. There was no time limit for selecting the answer, there was no feedback whether the selected answer was correct.

2.1.7 Procedure

The participants were first explained the task and seated in front of the laptop. Then the sensors were applied to the left hand and wrist. The participant is then instructed to place the left hand on the table in front of them and not to move it, in order to prevent movement or motion artifacts.

2.2 Analysis

2.2.1 Affectiva Q Sensor Data

The data gathered from the Affectiva Q Sensor proved unusable. In most participants no proper contact seemed to have been made with the skin during the full experiment. Dry GSR sensors need some sweat to ensure a good connection with both electrodes. Given the short experiment duration (a few of minutes) this connection was likely never made.

In hindsight the participants could have worn the device longer before the first task start, or they could have been asked to perform some light exercise.

Any data discussed from here on is solely that collected using the BIOPAC MP30 and it's finger electrodes.

2.2.2 Preprocessing

The collected data contained a DC current, which was smoothed by filtering and resampling to 100 samples per second using Savitzky-Golay filtering. The data per subject was then cut into seperate frames, the tasks.

The data was then normalized per subject over all datapoints in all tasks. This was done by dividing every data point of the subject by the average of the subject over all tasks.

Let the average over all tasks:

$$AverageGSR(s) = \frac{\sum_{q \in Q_s} \sum_{t \in T_q} GSR(s,q,t)}{\sum_{q \in Q_s} |T_q|}$$

With s a given subject, Q_s a set of all tasks performed by s, and T_q the set of all measurements in task q.

The normalised GSR is then given by:

$$NormGSR(s,q,t) = \frac{GSR(s,q,t)}{AverageGSR(s)}$$

Note that the normalization is done by dividing by the average of the subject over all tasks, and not by simply transforming all measurements to be between 0 and 1. What (original) GSR value would become 1 and 0 would then be quite arbitrary per subject, and not all subjects would be on the same scale, as outliers can totally influence this range.

$$AvgTaskGSR(s,q) = \frac{\sum_{t \in T_q} NormGSR(s,q,t)}{|T_q|}$$

$$StdDevTaskGSR(s,q) = \sqrt{\frac{1}{|T_q|} \sum_{t \in T_q} (NormGSR(s,q,t) - AvgTaskGSR(s,q))^2}$$

$$DiffTaskGSR(s,q) = NormGSR(s,q,t_0) - NormGSR(s,q,t_{|T_q|})$$

This normalised GSR

These tasks were then smoothed one last time, to get rid of a small remaining amount of DC noise. This was again done using a Savitzky-Golay filter.

2.2.3 Transformation

From the measurements from every task features were extracted.

The extracted time domain features:

- Average over all points in the task. - Accumulative of all points in the task. - Standard deviation of all points in the task. - Difference between first and last point of the task. - Peaks

Average

Accumulative

2.2.4 Data mining

2.2.5 Interpretation

3 Results

3.1 ..?

3.2 ..?

4 Discussion

4.1 Viability

The results show that from the results

Unfortunately I can not make any conclusions regarding

4.1.1 Task bleeding

Another explanation may be the bleeding of tasks into other tasks. If for instance a subject first gets two hard tasks, and then an easy one, it is imaginable that the subjects GSR value for that easy task is higher than it would have been without first having done harder tasks.

This is mitigated by the random order in which the tasks are presented and the time between tasks, however with the relatively small subject pool it is not unimaginable this has had its effect on the classification accuracy.

4.2 Future research

Future research could include investigating how viable this method is given more portable (consumer) sensors in a real life setting.

5 Conclusion

References

[1] Robert JK Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

[2] Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A. Calvo. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, OzCHI '12, pages 420–423, New York, NY, USA, 2012. ACM.

[3] Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. When do we interact multimodally?: Cognitive load and multimodal communication patterns. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, ICMI '04, pages 129–136, New York, NY, USA, 2004. ACM.

[4] Peter Schmutz, Silvia Heinz, Yolanda Métrailler, and Klaus Opwis. Cognitive load in ecommerce applications: Measurement and effects on user satisfaction. *Adv. in Hum.-Comp. Int.*, 2009:3:1–3:9, January 2009.

[5] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. Galvanic skin response (gsr) as an index of cognitive load. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2651–2656, New York, NY, USA, 2007. ACM.