

The Effect of Stress on Cognitive Load Measurement

Dan Conway^{*1,2}, Ian Dick^{*1,3}, Zhidong Li^{1,4}, Yang Wang¹, Fang Chen^{1,4}

¹NICTA, NSW 2015, Australia

²University of Newcastle, NSW 2308, Australia

³University of Sydney, NSW 2006, Australia

⁴University of New South Wales, NSW 2052, Australia

ian.dick01@gmail.com, jujuspacejazz@hotmail.com
{zhidong.li, yang.wang, fang.chen}@nicta.com.au

Abstract. Human physiological signals have been widely used to non-invasively measure cognitive load (CL) during task execution. A major challenge for CL detection is the presence of stress, which may affect physiological measurements in ways that confound reliable detection of CL. In this experiment we investigated the effect of stress on cognitive load measurement using galvanic skin response (GSR) as a physiological index of CL. The experiment utilized feelings of lack of control, task failure and social-evaluation to induce stress. Mean GSR values were shown to be significantly different between CL levels in the ‘no-stress’ condition, but not when including the ‘stress’ condition. On the other hand, features extracted from GSR signals based on peak detection exhibited consistent behaviour under both conditions, demonstrating the usefulness of the features as cognitive load index even when a person’s stress level is fluctuating.

Keywords: Cognitive load, galvanic skin response (GSR), stress.

1 Introduction

The term cognitive load (CL) refers to the amount of cognitive resources required for a person to complete a certain task. CL has been shown to have important implications for learning [13], safety in driving [4], aviation [16], and user interface design [10]. Being able to accurately monitor CL in real world environment would have important and far-reaching implications. This capability could enable the real-time detection of cognitive overload, when a person’s cognitive resources are overloaded with information, and this could help to avoid dangerous situations, such as an air traffic controller with too many concurrent responsibilities. CL monitoring could also enable the design of better computer interfaces, since interface designs could be rated based on the CL levels they impose on users, and designs with poor performance could be disregarded.

* First authors

Physiological signals have previously been proposed as a method of quantifying CL. Some notable successes in CL evaluation have been achieved via signals such as respiration, brain wave, and pupillary response [6] [14] [15] [17] [18]. Another physiological signal that has been used successfully to quantify CL, and which is the focus of this research effort, is galvanic skin response (GSR), a measure of the electrical conductance of the skin shown to be sensitive to both CL and stress [5] [9].

Any given physiological signal used as an index for CL is likely to be affected by various additional inputs from the human body. Thus, a major task of cognitive workload measurement via physiological means is demonstrating the diagnosticity and construct-validity of the CL index. One of the major contributors to change in human physiological systems is stress [8]. Stress has been shown to effect both the sympathetic and parasympathetic nervous systems and, in its more extreme states, results in large changes to physiological function that may well obscure the relationship between a physiological indicator and CL. Furthermore, stress may, in some circumstances, be a confounding factor for CL in that changes in CL may correlate with changes in stress levels. Construct-validity must be established before we can safely assert that changes in physiological indicators are the result of CL and not stress or other confounding factors. This experiment studies the effect of stress on CL measurement using GSR.

Although definitions of stress vary, there is good consensus in the literature regarding conditions where it is likely to arise [3] [7] [11]. Failure at a task, together with feelings of lack of control, in situations where participants are evaluated by others is a widely used paradigm for stress induction. These conditions could be operationalised in an experimental paradigm developed by Dedovic et al. called the ‘Montreal Imaging Stress Task’ (MIST) [2]. The experiment presented here closely follows the MIST protocol with minor operational adjustments.

2 Related Work

GSR has been used successfully in the past to index CL. In an experiment involving traffic control management it was illustrated that the mean GSR of test subjects increases as the difficulty of cognitive tasks increases [12]. In addition, [9] analysed the time and frequency domains of recorded GSR signals and showed that CL imposed through arithmetic and reading tasks can be indexed by GSR of test subjects.

The relationship between GSR and stress has also been examined. In an experiment involving driving tasks, [5] were able to successfully classify different driving periods based on the stress levels of the driver. They extracted useful features from GSR signals recorded during the experiment based on peak detection and input these features into machine learning classification algorithms with positive results. [19] combined GSR with several other physiological signals to classify the stress states of test participants through the use of machine learning tools. Stress levels were induced by having subjects complete a ‘Paced Stroop Test’, where the colour of a word that spells a different colour must be identified.

An interesting study on discriminating stress from cognitive load was carried out by Setz et al. [11]. However, they did not experimentally manipulate CL and only demonstrate the ability to differentiate between ‘stress’ and ‘no-stress’ conditions, where CL level was consistent between the two conditions. Nonetheless, the feature detection processes they outlined appear promising, and have been further studied in the experiment presented here.

3 Experiment

11 male students and employees (24-49 years’ old, ten right handed and one left-handed) took part in the experiment. Participants were offered one movie ticket and biscuits as recompense for their participation.

All participants had the voluntary nature of the experiment explained to them and then filled out a paper version of the Kessler K-10 Psychological Distress Scale [1] to ascertain that they were unlikely to be vulnerable to ongoing negative effects from the stress condition. Only participants who scored less than 19 (thus fell into the category ‘likely to be well’) were permitted to continue the experiment. Three potential candidates were rejected via this means.

3.1 Apparatus

All experimental stimuli were presented on a computer screen using custom software whilst participants were sitting comfortably at a desk. GSR signals were collected using a Thought Technology ‘ProComp Infiniti’ interface and its ‘SCFlex/Pro’ skin conductance sensor. The sensors were attached to the non-dominant hand for all participants. GSR signals were sampled at a rate of 10Hz. The brain wave signals were also recorded with a low-cost EEG device ‘Emotiv EPOC’. However the analysis of the EEG data is out of the scope of this paper. Participants were asked to remain still and only move their dominant hand for mouse control during the experiment.

3.2 Procedure

The experiment consisted of a within-subjects, six-way factorial design. There were math questions of three difficulty levels (low, medium and high) administered under two different stress conditions: ‘no-stress’ and ‘stress’. For level 1 problems (low difficulty), three terms were added together. For level 2 (medium difficulty), each problem consisted of four terms, with both addition and subtraction required. Level 3 problems (high difficulty) consisted of five terms, with addition, subtraction and multiplication required. The multiplication terms were in a random position within the problem.

All participants undertook the ‘no-stress’ condition first. Participants were told that they would be completing math tasks but it was emphasized that their performance/accuracy was not important. After submitting some basic demographic information, a two minute baseline period was carried out where the participants were told,

via an on screen prompt, that they should just relax and let their mind wander. Then three two-minute blocks of math tasks were presented, each with 4 multiple choice answers. Tasks were not time limited and feedback was not provided. The three blocks in the ‘no-stress’ condition were of level 1, 2 and 3 difficulty in sequential order. The participant was given a two-minute resting period in between each block.



Fig. 1. Experiment setup.

After block three, the participants were asked, via on screen prompts, to nominate a ‘target score’ for further tasks based on their estimation of their performance so far. Once submitted, the stress condition ensued. They were told that their performance would now be monitored. They were also informed of time limits for further trials, and video screens were switched on so that the test subject could see a video of themselves and also of other people observing them (see Figure 1). Now in the stress condition, three more blocks of level 1, 2 and 3 math tasks were carried out, again with two-minute pauses in between each block but with time limits now imposed on each trial. Feedback (‘correct’, ‘wrong’ or ‘out of time’) was provided for one second after each trial.

4 Analysis

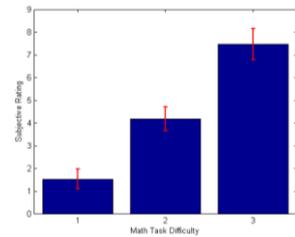


Fig. 2. Subjective rating of task difficulty.

In order to test the validity of the methods used in the experiment for inducing different levels of cognitive load, one-way ANOVA test of pooled subjective ratings was conducted. Results showed that the difference between the different difficulty levels

was significant ($F = 82.32$, $p < 0.05$). It can be seen in Figure 2 that the means of each group were increasing with the task difficulty level.

4.1 Analysis using mean GSR

The mean GSR values were inspected to study the effect of stress on cognitive load measurement. The distribution of normalised mean GSR values corresponding to the sub-sections of math task difficulty 1, 2 and 3 under both ‘no-stress’ and ‘stress’ conditions can be seen in Figure 3a. To investigate the relationship between mean GSR and cognitive load when no stressful stimuli are present, we conducted ANOVA analysis on these GSR values under ‘no-stress’ conditions and found that there are statistically significant differences between the 3 different levels ($F = 10.5$, $p < 0.05$), and there is a noticeable upward trend in mean GSR that corresponds to an increase in task difficulty.

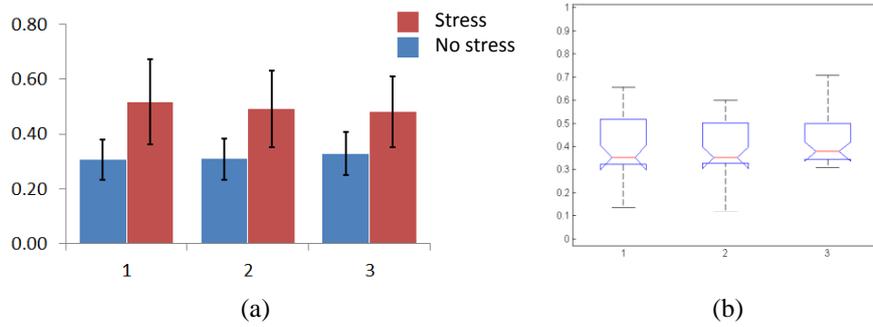


Fig. 3. Distribution of normalized mean GSR values for task difficulty levels 1,2,3 under the ‘no-stress’ and ‘stress’ conditions.

The positive correlation between cognitive load and GSR could no longer be observed, however, once the data from the ‘stress’ part of the experiment are included for consideration. Figure 3b shows the distribution of the normalised mean GSR values for math task difficulty levels 1, 2 and 3, with both the ‘no-stress’ and ‘stress’ data included. ANOVA analysis of these values does not produce significant results ($F = 0.05$, $p = 0.95$). The results indicate that mean GSR cannot effectively index cognitive load when stress levels are fluctuating, since mean GSR is sensitive to stress and the correlation between cognitive load and mean GSR becomes obfuscated when stress is a confounding factor. To overcome this problem, feature extraction technique has been employed for workload evaluation in the following section.

4.2 Analysis using ‘peak’ features

Similar to [5], several features corresponding to the peaks in the signals were extracted from the smoothed GSR signals. The following definitions were made: S_D is the distance along the x-axis from the local min preceding a peak to the local max of the

peak (i.e. peak duration); S_M is the distance along the y-axis from the local min preceding a peak to the local max of the peak (i.e. peak magnitude); S_F is the number of peaks divided by the task period (i.e. peak frequency). Figure 4 illustrates these concepts. In this work we used these three peak based GSR features to study the effect of stress on cognitive load measure. Except S_M , the other two features, S_F and S_D , demonstrated their usefulness for indexing CL even when stress is a confounding factor.

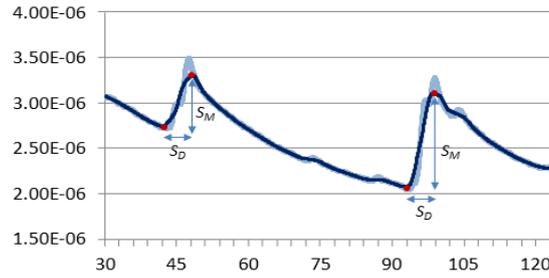


Fig. 4. Example of a smoothed GSR signal adorned with S_D and S_M features.

The S_F feature represents the frequency of peaks per sub-section. The ‘no-stress’ and ‘stress’ distributions for this feature are shown below in Figures 5a. For mean GSR, there was no common trend between the ‘no-stress’ and ‘stress’ conditions, and so there is no obvious way of using mean GSR to index CL when stress levels are fluctuating. In contrast to this, the S_F feature displayed in Figure 5 behaves similarly in both the ‘no-stress’ and ‘stress’ conditions. We can see that S_F is negatively correlated to task difficulty regardless of whether stress is present, although the relationship is stronger in the ‘stress’ condition. Figure 5b shows the distribution of the S_F feature for difficulty levels 1, 2 and 3, with both the ‘no-stress’ and ‘stress’ data included. ANOVA analysis was performed on this data to test the significance of the negative correlation. The result exhibited significant difference among the three difficulty levels ($F = 3.96$, $p < 0.05$).

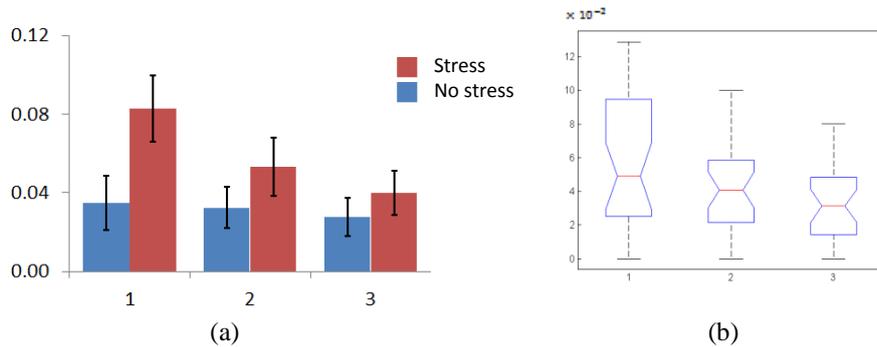


Fig. 5. Distribution of S_F feature for task difficulty levels 1, 2, 3 under the ‘no-stress’ and ‘stress’ conditions.

The S_D feature corresponds to the peak duration per sub-section. The distribution of normalized S_D feature (sum of peak durations divided by the sub-section period) corresponding to math task difficulty 1, 2 and 3 under both ‘no-stress’ and ‘stress’ conditions can be seen in Figure 6a. It turns out that this feature behaves quite similarly to the S_F feature, and is negatively correlated to task difficulty under both ‘no-stress’ and ‘stress’ condition. Figure 6b shows the distribution of the S_D feature for difficulty levels 1, 2 and 3, with both the ‘no-stress’ and ‘stress’ data included. The downward trend with increasing task difficulty could be observed. ANOVA analysis also generate significant difference among the three difficulty levels ($F = 5.14$, $p < 0.05$), indicating the feature could be useful as an index of cognitive load even under the influence of stress conditions.

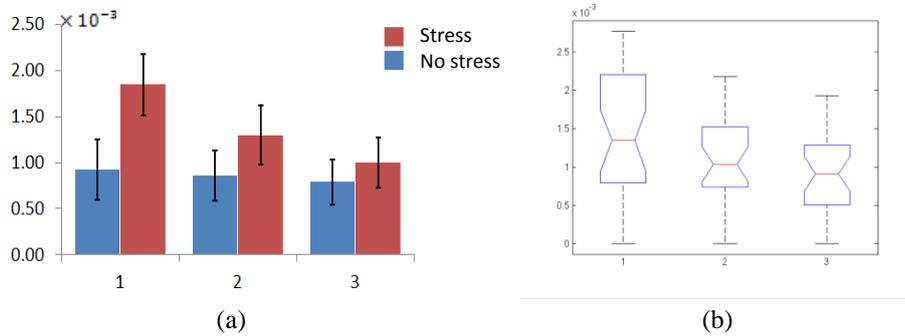


Fig. 6. Distribution of normalized S_D feature for task difficulty levels 1, 2, 3 under the ‘no-stress’ and ‘stress’ conditions.

5 Conclusion

The experiment in this work helped to reinforce GSR as an index of cognitive load during task execution. Without the impact of stress, it appears that an increase in CL (induced by increasing the difficulty of tasks given to test subjects) results in an increase in mean GSR value. This relationship is, however, obfuscated when test subjects experience fluctuating levels of stress. Stress was introduced into the experiment using an adaptation of the MIST protocol, and this blurred the connection between GSR and CL.

GSR may still be useful as an index for CL even when stress is a confounding factor, if we consider peak based features extracted from the GSR signal other than the mean value. Both peak frequency in the signal and peak durations are negatively correlated to task difficulty and hence CL. These features could possibly be used to dissociate CL from stress and develop a stress-agnostic method of CL classification. Our future work in this area include more feature extraction and also machine learning to investigate whether these features may be used to classify a person’s CL under the influence of various confounding factors.

References

1. Andrews, G., Slade, T.: Interpreting scores on the Kessler psychological distress scale (K10). *Australian and New Zealand journal of public health*, vol. 25, pp. 494-497 (2007)
2. Dedovic, K., Renwick, R., Mahani, N. K., Engert, C., Lupien, S. J., Pruessner, J.: The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry & Neuroscience*, vol. 30, issue 5, pp. 319-325 (2005)
3. Dickerson, S. S., Kemeny, M. E.: Acute Stressors and Cortisol Responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, vol. 130, issue 3, pp. 355-391 (2004)
4. Engstrom, J., Johansson, E., Ostlund, J.: Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, issue 2, pp. 97-120 (2005)
5. Healey, J., Picard, R.: SmartCar: detecting driver stress. *International Conference on Pattern Recognition*, vol. 4, pp. 218 - 221. (2000)
6. Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., Zarjam, P.: Measuring cognitive workload with low-cost electroencephalograph, *IFIP Conf. Human-Computer Interaction (INTERACT)*, pp. 568-571, (2011)
7. Markus, H.: The effect of mere presence on social facilitation: an unobtrusive test. *Journal of Experimental Social Psychology*, vol. 14, issue 4, pp. 389-397 (1978)
8. Martin, G. N., Carlson, N.R., Biskit, W.: *Psychology*. Pearson Education, Harlow (2007)
9. Nourbakhsh, N., Wang, Y., Chen, F., Calvo, R. A.: Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. *Proceedings of the 24th Australian Computer-Human Interaction Conference. ACM*, pp. 420-423 (2012)
10. Saadé, R. G., Otrakji, C. A. First impressions last a lifetime: effect of interface type on disorientation and cognitive load. *Computers in human behaviour*, vol. 23, issue 1, pp. 525-535 (2007)
11. Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., Ehlert, U.: Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, issue 2, pp. 410-417 (2010)
12. Shi, Yu., Ruiz, N., Taib, R., Choi, E., Chen, F.: Galvanic skin response (GSR) as an index of cognitive load. *CHI extended abstracts on Human factors in computing systems*. (2007)
13. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. *learning and instruction*, vol. 4, issue 4, 295-312 (1994)
14. Veltman, J. A., Gaillard, A. W. K.: Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, vol. 41, issue 5, pp. 656-669 (1998)
15. Wang, W., Li, Z., Wang, Y., Chen, F.: Indexing cognitive workload based on pupillary response under luminance and emotional changes, *Int'l Conf. Intelligent User Interfaces (IUI)*, pp. 247-256, 2013.
16. Wilson, G. F.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation*, vol. 12, pp. 3-18 (2002)
17. Xu, J., Wang, Y., Chen, F., Choi, H., Li, G., Chen, S., Hussain, S.: Pupillary response based cognitive workload index under luminance and emotional changes. *CHI extended abstracts on Human factors in computing systems*. (2011)
18. Xu, J., Wang, Y., Chen, F., Choi, E.: Pupillary response based cognitive workload measurement under luminance changes, *IFIP Conf. Human-Computer Interaction (INTERACT)*, pp. 178-185, (2011)
19. Zhai, J., Barreto, A.: Stress detection in computer users through non-invasive monitoring of physiological signals. *Blood*, vol. 5 (2008)