

GSR and Blink Features for Cognitive Load Classification

Nargess Nourbakhsh^{1,2}, Yang Wang¹, and Fang Chen¹

¹ National ICT Australia (NICTA), Sydney, Australia

² The University of Sydney, Sydney, Australia

{nargess.nourbakhsh, yang.wang, fang.chen}@nicta.com.au

Abstract. A system capable of monitoring its user's mental workload can evaluate the suitability of its interface and interactions for user's current cognitive status and properly change them when necessary. Galvanic skin response (GSR) and eye blinks are cognitive load measures which can be captured conveniently and at low cost. The present study has assessed multiple features of these two signals in classification of cognitive workload level. The experiment included arithmetic tasks with four difficulty levels and two types of machine learning algorithms have been applied for classification. Obtained results show that the studied features of blink and GSR can reasonably discriminate workload levels and combining features of the two modalities improves the accuracy of cognitive load classification. We have achieved around 75% for binary classification and more than 50% for four-class classification.

Keywords: Cognitive load, galvanic skin response, eye blink, machine learning.

1 Introduction

Being continuously aware of user's mental status is an important step in making intelligent systems interacting with people. Such systems can properly change their interface and interactions to match the imposing workload with the current working memory of their user. In this way, the optimum performance will be obtained and many human errors will be avoided. Therefore, it is necessary to measure mental load accurately and in real-time.

Cognitive load is commonly used to refer to the load that performing a particular task imposes on the person's cognitive system [19]. Different methods have been applied for quantifying cognitive workload; however, not all of them are useful for developing adaptable systems. Subjective (self-reporting) [18] and performance-based measurement [3] techniques have been widely used and, regarding implementation, are usually the most convenient methods. However, asking subjects to rate the experienced mental workload means several interruptions and distractions from performing the principal tasks. Moreover, both methods are post-task processing and can be done when the task is finished, thus are not useful for real-time cognitive load assessment. In contrast, human behaviors and physiological responses can continuously and non-intrusively demonstrate user's cognitive states while performing the intended task.

Several physiological signals have been used for cognitive load measurement: signals from heart [20], eye [27], brain [2, 12] and skin [17]. Galvanic skin response (GSR), which is electrical conductance of skin, is a low-cost, easily-captured, robust physiological signal. Previous studies have used skin conductance in detecting emotions [15] or differentiating between stress and cognitive load conditions [16], and a few ones have found relations between GSR features and mental workload [17, 25]. Some others have tried but did not obtain satisfactory results for detecting cognitive load from GSR [7, 11].

Speech [10], pen input [21] and eye movements [5] are instances of behavioral signals used in cognitive load measurement. Eye activity can reveal valuable information about mental workload. In contrast with some eye based features (such as pupil dilation) which can only be gathered through an expensive eye tracker, eye blink can be obtained with an acceptable accuracy through a conventional camera. Therefore it is a low-cost and easily-obtained signal which can be used for cognitive load measurement. Some previous works have studied blink variations in regard to modality (visual versus acoustic) [13] or location (central versus peripheral) [6] of presenting stimuli. Another research has measured the blink rate in resting, reading and talking conditions [1]. A few studies have examined blink features in two cognitive load levels and found them related with mental load level [5, 8].

There are various application domains for cognitive load measurement, from brain-computer interactions to air traffic control. However, in some domains such as driving and education it is essential to be able to measure this load at a low cost, with short preparation time and minor restriction of users' movements. Considering that GSR and eye blink are suitable measures in such situations, in this paper we have explored features of these two signals captured during arithmetic experiments consisting of four cognitive load levels. We have assessed how useful every single feature is for classifying mental workload level and how combining features from the two signals affects the accuracy of cognitive load classification.

Support vector machines (SVM) and Naïve Bayesian classifiers are two popular machine learning algorithms in human-computer interaction studies [14]. Some previous works have used SVM for recognizing drowsiness [28] or different emotions [9, 22, 24] from physiological features. Naïve Bayesian classifiers have been used for detecting human emotions from facial expressions [23] or physiological signals [4, 9]. In this study, we have used these two types of classifiers for cognitive load classification of GSR and blink features.

2 Experiment

The data was collected from thirteen healthy 24 to 35-year-old volunteers who signed consent forms before the experiment and were awarded with movie vouchers for their participation. The experiment included 8 arithmetic tasks with 4 difficulty levels. Each subject performed two trials of each task level and the whole eight trials were performed in a randomized order. In each task four numbers were shown one by one, each for three seconds. Subjects were supposed to add-up these four numbers and

select (by clicking the mouse using their right hand) the correct answer from three numbers which were next presented on the screen. First to fourth difficulty levels respectively included binary numbers (0 and 1), one-digit numbers, two-digit numbers and three-digit numbers. Before appearing the first number of each task, a slide containing one, two or three ‘x’ symbols (according to the number of digits in the task) was presented for three seconds. There was no time limit for answering and the background was always black. There was a 6-second rest time between consecutive tasks. After finishing the experiment, subjects rated task difficulty levels in a questionnaire (ranging from 1 to 9).

To collect galvanic skin response, the GSR device from ProComp Infiniti of Thought Technology Ltd was used and the sensors were attached to the subject’s left hand finger (all subjects were right-handed). The sampling frequency was 10Hz. Eye activity data was recorded with a remote eye tracker (faceLAB 4.5 of Seeing Machines Ltd) which operated at a sampling rate of 50Hz and continuously recorded eye data. A 21” LCD monitor and a usual computer mouse were used for presenting the tasks and obtaining user inputs.

3 Cognitive Load Measurement

Figure 1 shows the average subjective ratings of the task difficulty levels. One-way analysis of variance (ANOVA) on the self-reporting scores showed a highly significant difference between task levels ($F_{3,48}=108.63$, $p<0.05$).

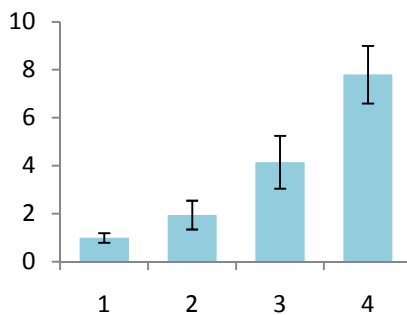


Fig. 1. Subjective rating of task difficulty levels

We also examined the response time for each task. By response time we mean the time between *disappearing the last (fourth) number* of the task and *selecting the answer*. Average response time of all subjects for each task difficulty level are shown in Figure 2. It can be seen that response time has a direct relation with the task difficulty level: harder tasks take longer response time. Results of ANOVA test on response time of different task levels are significant ($F_{3,48}=62.59$, $p<0.05$). These observations about subjective rating and response time show that the designed tasks have effectively manipulated the cognitive load.

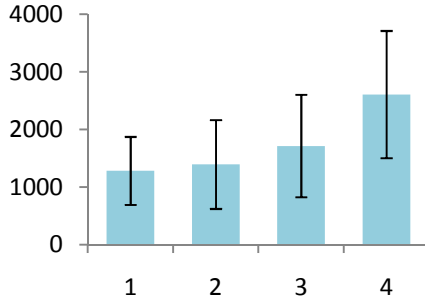


Fig. 2. Response time for each task difficulty level (milliseconds)

Two GSR and two blink features were calculated for each task:

- accumulative GSR (summation of GSR values over task time)
- GSR power spectrum (frequency power)
- blink number (number of blinks in the task)
- blink rate (number of blinks in the task divided by task time)

The time between *appearing the first number* and *inputting the answer* was considered as the task time in which every feature was computed. We observed that GSR and blink values are highly subjective, that is they differ from person to person. In order to omit the subjective differences, we calibrated each feature of task j of participant i by dividing it by the average of all similar features of all tasks of that subject:

$$\text{calibrated_feature}(i,j) = \frac{\text{feature}(i,j)}{\frac{1}{m} \sum_{j=1}^m \text{feature}(i,j)} \quad (1)$$

where m is the number of tasks ($m=8$). Furthermore, we averaged each feature between tasks with same difficulty levels for each subject. Figures 3 and 4 show the average values of the studied (calibrated) features in the four task levels.

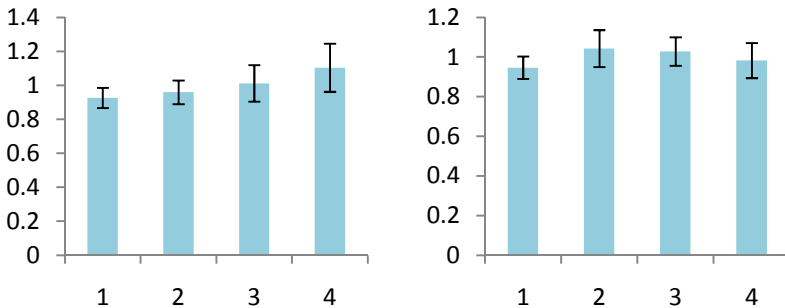


Fig. 3. Average GSR features of all subjects for the four task levels: accumulative GSR (left), GSR frequency power (right)

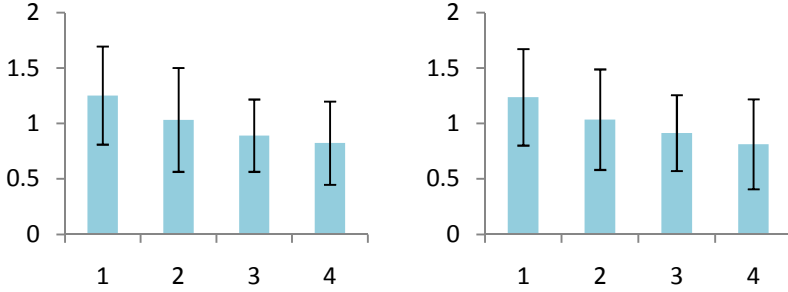


Fig. 4. Average blink features of all subjects for the four task levels: blink number (left), blink rate (right)

One-way ANOVA test was applied to statistically evaluate cognitive load level discrimination of each feature. Table 1 represents the results of statistical analysis (ANOVA test) of the studied features for four task difficulty levels. As can be seen, the results are significant for all four features.

Table 1. ANOVA results of features for four task difficulty levels

Feature	Results
Accumulative GSR	$F_{3,48} = 7.22, p < 0.05$
GSR Frequency Power	$F_{3,48} = 4.07, p < 0.05$
Blink Number	$F_{3,48} = 3.37, p < 0.05$
Blink Rate	$F_{3,48} = 3.22, p < 0.05$

4 Cognitive Load Classification

In this study, support vector machines (SVM) and Naïve Bayes classifiers were applied for cognitive load classification. For every feature, we have examined two- and four-class classification while the former means considering levels one and two as low load and levels three and four as high load. The cross validation method was leave-one-subject-out. In other words, in each round the classifier was trained by the data of the all subjects except one and data of the remaining subject was used for testing. The classification accuracies of all rounds were averaged.

Tables 2 to 5 show the cognitive load classification accuracies of the single features for two and four classes. Results of all features are reasonable, GSR features outperform blink features in two-class classification and results of blink number are better than those of blink rate. It is also worth mentioning that in most cases the two types of classifiers have very near or even similar (Table 4) performances on the single features. The largest difference is in classifying by use of accumulative GSR (Table 2) where classification accuracies of Naïve Bayes learners are about 5% higher than those of SVM in both 2-class and 4-class classifications.

Table 2. Classification accuracies of accumulative GSR

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	66.4%	34.6%
Naïve Bayes	71.2%	40.4%

Table 3. Classification accuracies of GSR frequency power

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	66.4%	37.5%
Naïve Bayes	65.4%	35.6%

Table 4. Classification accuracies of blink number

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	62.5%	40.0%
Naïve Bayes	62.5%	40.0%

Table 5. Classification accuracies of blink rate

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	57.5%	31.3%
Naïve Bayes	55.0%	32.5%

In the next step, we examined cognitive load classification using combinations of GSR and blink features. The combination of blink number and GSR frequency power resulted in the highest classification accuracies which can be seen in Table 6. Comparison with tables 3 and 4 reveals that combining blink number and GSR frequency power improves the classification accuracy in both two- and four-class classifications up to about 10% in the former and 16% in the latter. It can be observed that for combination of the two modalities (Table 6), similar to single feature cognitive load classifications, the classification accuracies of SVM and Naïve Bayes classifiers are close.

Table 6. Classification accuracies of blink number + GSR frequency power

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	71.5%	53.6%
Naïve Bayes	75.0%	50.0%

5 Conclusion

We have applied classification algorithms on blink and GSR features and combinations of them. Accumulative GSR, power spectrum of GSR, blink number and blink rate were significantly distinctive and had reasonable accuracies in both two- and four-class classification of cognitive load using support vector machines and naïve Bayes classifiers. Combining GSR and blink features improved the classification accuracy. As our next step towards fully automated and more precise cognitive load detection, we will apply automatic feature selection (e.g. [26]) of physiological cognitive measures.

References

1. Bentivoglio, A.R., Bressman, S.B., Cassetta, E., Carretta, D., Tonali, P., Albanese, A.: Analysis of Blink Rate Patterns in Normal Subjects. *Movement Disorders* 12, 1028–1034 (1997)
2. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, Space, and Environmental Medicine* 78, B231–B244 (2007)
3. Chandler, P., Sweller, J.: Cognitive Load While Learning to Use a Computer Program. *Applied Cognitive Psychology* 10, 151–170 (1996)
4. Chanel, G., Kronegg, J., Grandjean, D., Pun, T.: Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. In: Günsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) *MRCS 2006. LNCS*, vol. 4105, pp. 530–537. Springer, Heidelberg (2006)
5. Chen, S., Epps, J., Ruiz, N., Chen, F.: Eye Activity as a Measure of Human Effort in HCI. In: *International Conference on Intelligent User Interfaces*, pp. 315–318. ACM Press (2011)
6. Fogarty, C., Stern, J.A.: Eye movements and blinks: Their Relationship to Higher Cognitive Processes. *International Journal of Psychophysiology* 8, 35–42 (1989)
7. Haapalainen, E., Seungjun, K., Forlizzi, J.F., Dey, A.K.: Psycho-Physiological Measures for Assessing Cognitive Load. In: *International Conference on Ubiquitous Computing*. ACM Press (2010)
8. Holland, M.K., Tarlow, G.: Blinking and Mental Load. *Psychological Reports* 31, 119–127 (1972)
9. Horlings, R., Dacu, D., Rothkrantz, L.J.: Emotion Recognition Using Brain Activity. In: *International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, pp. II.1.1–II.1.6. ACM (2008)
10. Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., Leino, T.: Effect of Cognitive Load on Speech Prosody in Aviation: Evidence from Military Simulator Flights. *Applied Ergonomics* 42, 348–357 (2011)
11. Ikehara, C.S., Crosby, M.E.: Assessing Cognitive Load with Physiological Sensors. In: *Hawaii International Conference on System Sciences*, p. 295a (2005)
12. Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., Zarjam, P.: Measuring cognitive workload with low-cost electroencephalograph. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) *INTERACT 2011, Part IV. LNCS*, vol. 6949, pp. 568–571. Springer, Heidelberg (2011)

13. Lipp, O.V., Neumann, D.L.: Attentional Blink Reflex Modulation in a Continuous Performance Task is Modality Specific. *Psychophysiology* 41, 417–425 (2004)
14. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces. *Journal of Neural Engineering* 4, R1–R13 (2007)
15. Nakasone, A., Prendinger, H., Ishizuka, M.: Emotion Recognition from Electromyography and Skin Conductance. In: *International Workshop on Biosignal Interpretation*, pp. 219–222 (2005)
16. Niculescu, A., Cao, Y., Nijholt, A.: Stress and Cognitive Load in Multimodal Conversational Interactions. In: *International Conference on Human-Computer Interaction*, pp. 891–895. Springer (2009)
17. Nourbakhsh, N., Wang, Y., Chen, F., Calvo, R.A.: Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In: *Australian Computer-Human Interaction Conference*, pp. 420–423. ACM Press (2012)
18. Paas, F.G.W.C.: Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology* 84, 429–434 (1992)
19. Paas, F.G.W.C., Van Merriënboer, J.G.: Instructional Control of Cognitive Load in the Training of Complex Cognitive Tasks. *Educational Psychology Review* 6, 351–371 (1994)
20. Paas, F.G.W.C., Van Merriënboer, J.G.: Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology* 86, 122–133 (1994)
21. Ruiz, N., Taib, R., Shi, Y., Choi, E.H., Chen, F.: Using Pen Input Features as Indices of Cognitive Load. In: *International Conference on Multimodal Interfaces*, pp. 315–318. ACM Press (2007)
22. Sakr, G.E., Elhadj, I.H., Wejinya, U.C.: Multi Level SVM for Subject Independent Agitation Detection. In: *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 538–543 (2009)
23. Sebe, N., Lew, M.S., Cohen, I., Garg, A., Huang, T.S.: Emotion Recognition Using a Cauchy Naive Bayes Classifier. In: *2002 Proceedings of the 16th International Conference on Pattern Recognition*, vol. 1, pp. 17–20. IEEE (2002)
24. Setz, C., Arnrich, B., Schumm, J., Marca, R.L., Tr, G., Ehlert, U.: Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *Technology* 14, 410–417 (2010)
25. Shi, Y., Ruiz, N., Taib, R., Choi, E.H., Chen, F.: Galvanic Skin Response (GSR) as an Index of Cognitive Load. In: *Computer/Human Interaction Conference Work-in-Progress*, pp. 2651–2656. ACM Press (2007)
26. Wang, W., Li, Z., Wang, Y., Chen, F.: Indexing Cognitive Workload Based on Pupillary Response under Luminance and Emotional Changes. In: *International Conference on Intelligent User Interfaces*, pp. 247–256. ACM Press (2013)
27. Xu, J., Wang, Y., Chen, F., Choi, E.: Pupillary Response Based Cognitive Workload Measurement under Luminance Changes. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) *INTERACT 2011, Part II. LNCS*, vol. 6947, pp. 178–185. Springer, Heidelberg (2011)
28. Yeo, M., Li, X., Shen, K., Wildersmith, E.: Can SVM Be Used for Automatic EEG Detection of Drowsiness During Car Driving? *Safety Science* 47, 115–124 (2009)