2017

Towards a Continuous Assessment of Cognitive Workload for Smartphone Multitasking Users

Angel Jimenez-Molina
University of Chile, ajimenez@dii.uchile.cl

Hernan Lira
University of Chile, hlira@ing.uchile.cl

Recommended Citation

Jimenez-Molina, A. & Lira, H. (2017). Towards a continuous assessment of cognitive workload for smartphone multitasking users. *H-Workload 2017: The first international symposium on human mental workload*, Dublin Institute of Technology, Dublin, Ireland, June 28-30. doi:10.21427/D7290X Isbn: 9781900454637(vol)

Towards a Continuous Assessment of Cognitive Workload for Smartphone Multitasking Users

Angel Jimenez-Molina and Hernan Lira

University of Chile, Faculty of Physical and Mathematical Sciences, Department of Industrial Engineering, Beauchef 851, Of. 603, Santiago, Chile
ajimenez@dii.uchile.cl, hlira@ing.uchile.cl

Abstract. The intermeshing of Smartphone interactions and daily activities depletes the availability of cognitive resources. This excessive demand may lead to several undesirable cognitive states, which can be avoided by continuously assessing the user cognitive workload. Recently, many attempts have emerged to assess this workload by using psychophysiological signals. This paper provides evidence that it is possible to train models that accurately identify in short time windows such cognitive workload by processing heart rate and blood oxygen saturation signals. This assessment could be applied in Smartphone notifications delivery, interface adaptations or cognitive capabilities evaluation.

1 Introduction

Multitasking is a usual behavior in people these days. Nowadays it is common to see people performing different tasks in the Smartphone while daily activities are being conducted, such as walking, talking or waiting the bus. Activities and Smartphone tasks demand the simultaneous use of several cognitive resources, such as perception, attention, short and long term memory, motor control, among others [11]. Excessive demand of cognitive resources can lead to distractions, increase errors, provoke stress and frustration, and reduce ability for mental planning, problem solving, and decision making [1, 6, 9, 10, 11].

In order to avoid these states, it is necessary to continuously assess the cognitive workload – the perceived level of mental effort that a task or multiple tasks induce in the person [15]. Typically, it has been evaluated using subjective rating methods applied after the person has already finished the task, making the evaluation dependent of her/his final perception [14]. The static nature of these methods makes them unfit for real-time evaluation [3]. In addition, they are also constrained by a reporting bias and lack ecological validity.

Recently, psychophysiological signals like cardiovascular or brain measures are being used to assess cognitive workload. This is based on the empirical evidence of the correlation between the physiological responses triggered by the nervous system and psychological stimuli [2]. The cardiovascular system is especially interesting for this purpose since it is highly sensitive to neurological and psychological processes [4].

In this paper, we leverage the capabilities of heart rate (HR) and blood oxygen saturation (SpO_2) to research the possibility of assessing mental workload in real time during Smartphone interactions. In this regard, this paper attempts to answer the following research questions:

- RQ_1. Are HR and SpO_2 capable of discriminating between tasks with different levels of cognitive demand?
- RQ_2. Is it possible to accurately classify on real time the users cognitive workload according to different levels of cognitive demand by combining HR and SpO_2?

To answer these research questions, an experiment was conducted in which 50 users performed several tasks in different scenarios of cognitive resource demand.

This paper is organized as follows. Section 2 presents the related work. Methods are introduced in Section 3. Section 4 shows the experimental results. Discussion and conclusion are presented in Section 5.

2 Related work

The most comprehensive, quantitative existing method to assess cognitive workload is the multiple-resource model (MRM) of Wickens [13]. As shown in Table 1, it offers empirical evidence of three cognitive dimensions that can lead to competition and interference among cognitive resources if they are simultaneously used. The first dimension consists of three stages for processing task information, known as perceptual, central processing and response stages. The second dimension groups the cognitive resources demanded for each stage, e.g. selective attention, perception, working memory or motor control. The third dimension consists of a set of attributes that characterize cognitive resources. For instance, as shown in Table 1, the input modality - auditory, visual or tactile - and the processing code - spatial or verbal jointly describe the selective attention.

The interference among attributes of simultaneously demanded resources increases the task difficulty, which generates different levels of performance. For instance, Oulasvirta et al. [11] proposes a framework based on Wickens's MRM, in which presents evidence of competition due to the use of attention by both Smartphone interaction tasks and daily activities in outdoor environments. There are three types of interferences: due to the input modality, the processing stage and the processing code. The first one is complete and occurs when two or more tasks use the same input modality. The second interference can be complete or partial and occurs when tasks use resources from the same stage. Finally, the third stage is complete and is verified when there exists a conflict in the processing code.

Regarding subjective methods to assess cognitive workload, the most widespread example is the NASA Task Load Index (TLX), which measures the mental and physical performance, as well as the effort and frustration of the user [8].

Haapalainen et al. [7] assesses tasks of visual perception and cognitive speed using several sensors, including HR. Through a Naive Bayes classifier, the median of the heat flux and the average absolute median of the electrocardiogram, they achieve an average accuracy of 81.1%. Similarly, Fritz et al. [5] measure tasks performed by professional software developers using eye tracker, electrodermal activity and electroencephalogram. They obtain 84.38% precision with features extracted from the three signals training a Naive Bayes classifier. Ryu & Myung [12] use electroencephalogram, electrooculogram and electrocardiogram to evaluate the resolution of math sums with different levels of cognitive demand. The three studies mentioned above use the NASA TLX as the standard to evaluate their experimental designs [8].

3 Methods

Participants. An experiment was conducted with 50 participants, all engineering students of University of Chile (33 men and 17 women; *mean age* = 22.4 years, $SD = 2.8$ years), recruited through the institutional news Web application. None of them suffered from cardiovascular diseases or was taking medications that could have affected their normal behavior, HR or SpO_2 levels. All of them were familiar with the use of smartphones. They received CL$5,000 (about US$10) for their participation.

Task Design. Each participant was required to perform four treatments designed to induce low and high cognitive workload. Treatments were designed according to the Wickens's MRM. The first two treatments consist of two subtasks - reading a simple e-mail (ST_1) and replying a simple e-mail (ST_2) -, each of them performed in two scenarios: without verbal stimuli and with verbal stimuli. The presence or not of verbal stimuli defines if the treatment induces high or low cognitive workload respectively. For the third and fourth treatments participants must perform five subtasks under the same scenarios: reading an e-mail with search instructions (ST_3), opening an application suitable to perform the search instructions (ST_4), entering the search parameters in the application (ST_5), reading and analyzing the search results (ST_6), and replying the e-mail with the results (ST_7). The design of these treatments aims to manipulate the cognitive workload through the cognitive interferences phenomenon described in Wickens's theories. Thus, a task will be more cognitively demanding as there are more cognitive interferences in it.

At treatment one (T_1) the participant must answer simple e-mails free of interruptions. According to Wickens's MRM she/he only needs to control her/his personal space while performing the subtasks. Treatment two (T_2) requires the participant to perform ST_1 and ST_2, but at the same time to verbally answer questions from the experimenter while listening to music (in her/his native language) through a hearing aid. Thus, the participant should pay attention, examine, analyze, and make sense of the questions, while using auditory and visual input modalities and working memory. At treatment three (T_3) the participant is free of interruptions, but she/he must read and respond an e-mail requesting to search for hotel prices, travel, cars and restaurant names using a specific application. Finally, treatment four is similar to treatment three, but with the

Table 1. Cognitive Resources Demand and Interferences for each Treatment

Processing Stage	Perceptual Encoding				Central Processing	Responsive	
Cognitive Resource	Selective Attention			Perception	Working Memory	Motor Control	
Resource Attribute	Input Modality		Processing Code	Processing Code	Processing Code	Processing Code	
Treatment \ **Code**	Auditory	Visual	Verbal	Verbal	Verbal	Spatial	Verbal
T_1		Low				Low	
T_2	Low	Low	Low	Low	**High**	**High**	
T_3		Low				Low	
T_4	Low	Low	Low	Low	**High**	**High**	

cognitive workload induced by the experimenter questions and the music listening. One interference takes place in both, treatment one and three. This is caused by the simultaneous use of motor control resources and verbal code attributes. In contrast, at treatments two and four several partial interferences occur. This since both, the subtasks and the verbal stimulus require motor control and working memory resources simultaneously. We summarize in table 1 the level of cognitive interference associated to each treatment given the demanded cognitive resources, the input modality and processing code utilized for the task.

Apparatus. HR and SpO_2 were captured with a pulseoximeter that measure the pulse in beats per minute (bpm) and the blood oxygen saturation through the relation of the light irradiated from the blood with hemoglobin (Hb) and hemoglobin with oxygen ($Hb + O_2$). Both signals were captured at a rate of 50 *Hz*. The pulseoximeter is connected to an e-health shield from the Coocking Hacks company that sends data to a Linux server on the cloud.

The Subtasks were performed on a Samsung Galaxy S5 Smartphone. The interactions with the smartphone were video taped with an eye tracker in the shape of glasses. Recording the interaction is required to get the starting and ending timestamps of each subtask, which allows to measure duration of the subtasks in milliseconds. The eye tracker is connected to a Microsoft Surface Tablet that stores the video streams locally. Both the data in the cloud and the streams in the Tablet were latter synchronized using the Unix timestamp.

Experimental Procedure and Task Validation. The experiments were conducted individually through the following procedure: 1) The participant is asked to read and complete an informed consent (authorized by the ethical committee of the Faculty of Physical and Mathematical Sciences of University of Chile); 2) the participant is asked to close her/his eyes during 90 seconds in order to relax; 3) the experimenter randomizes the order in which the treatments will be performed and applies them; 4) after each treatment the participant must complete the NASA TLX test and is asked to close eyes for 30 seconds in order to relax again.

A pretest was conducted with five participants, who are not included in the experimental group, in order to validate each treatment design. A repeated mea-

Table 2. Repeated Measures ANOVA for Heart Rate and Oxygen Saturation

Signal	Measure	ST_1	ST_2	ST_3	ST_4	ST_5	ST_6	ST_7
HR	F	1044	3771	919	438.5	2622	2330	2345
	p-value	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
SpO_2	F	115.2	442.7	199.2	168.6	703.3	588.6	893.7
	p-value	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

sures analysis of variance (RM-ANOVA) is performed over the obtained data. As shown in Table 2, the difference between the means of the NASA TLX score for treatment 1 and 2, as well as treatment 3 and 4, are statistically significant ($p = 0.024 < 0.05$). Therefore, each pair of treatments, designed based on Wickens's MRM, induces different cognitive workloads on the participants.

Data Processing. Data records of each subtask were labeled with the cognitive workload (low or high) induced by each treatment (previously validated through the NASA TLX score). Each record consists of timestamp, subtask, HR value, SpO_2 value and label. These records are divided in a set of one-second time window. For each window, the mean, median and standard deviation (SD) are computed as features. In this way, new records are created following the structure $\{timestamp$, $subtask$, $mean\_HR$, $mean\_SpO_2$, $median\_HR$, $median\_SpO_2$, SD_HR, SD_SpO_2, $label\}$. By applying correlation and principal component analysis, it is determined that the *mean* is the best feature for both signals. Due to the high variance in the biological measures between humans, the measure within subjects is standardized by subtracting the *mean* from each value and dividing the difference by the SD.

4 Experimental Results

In average, the HR in treatments one and three is 76.90 bpm $\pm$ 3.7 bpm versus 82.70 bpm $\pm$ 4.8 bpm in treatments two and four. The values of SpO_2 are 96.06 bpm $\pm$ 1.8 bpm and 98.51 bpm $\pm$ 1.2 bpm according to the same treatments. Clearly, both HR and SpO_2 increase with the difficulty of the subtask.

Statistical Analysis. In order to answer the first research question it is necessary to apply a RM-ANOVA, since all the participants are exposed to the four treatments. The null hypothesis indicates that the *mean* for each group are equal. However, as shown in Table 2 it is possible to reject the null hypothesis both for HR and SpO_2. Therefore there is evidence that the signals used in this paper allow to discriminate different levels of cognitive workload (RQ_1).

Machine Learning. To perform the classification of cognitive workload level, a 10-fold cross validation is performed in which the classes are evenly distributed within each set and randomly selected in order to avoid bias. The independent variable is the cognitive level of the task. Table 3 shows the classification results

Table 3. Classification Results using Support Vector Machine and HR + $SpO2$

ST	Accuracy	Precision	Recall	ST	Accuracy	Precision	Recall
ST_1	82.78%	81.80%	76.68%	ST_5	81.23%	83.53%	78.03%
ST_2	81.94%	88.12%	84.01%	ST_6	84.40%	88.68%	82.89%
ST_3	80.19%	72.85%	77.19%	ST_7	85.71%	83.98%	87.65%

for each subtask by training a model through Support Vector Machine and features of HR and $SpO2$ combined (RQ_2). It is noticeable that the classification models have measures of accuracy and performance in range within previous studies in the field. The combination of both signals has the best predictive power (greater than 80 %). Separately, HR dominates $SpO2$, in terms of accuracy and precision, on each subtask. For instance, in ST_1, HR and $SpO2$ have an accuracy of 78.49% and 68.12% respectively. It is worth to remark that ST_6 and ST_7 have the best results, being ST_7 the subtask with the higher accuracy (85.71%).

5 Discussion and Conclusion

Psychophysiological signals are an objective mean to evaluate cognitive workload of Smartphone multitasking users. The current availability of non-invasive and inexpensive sensors, opens plenty of room to improve this approach. Particularly, in the area of Ubiquitous Computing, by providing a new type of information as an input to better understand user context.

This paper provides evidence to support that HR and SpO_2 signals are capable of discriminating between tasks of different difficulty (RQ_1), and that it is possible to classify between two levels of cognitive difficulty by training machine learning models (RQ_2). Given that the treatments were designed based on real scenarios, these findings could be applied to everyday activities, outside controlled environments. In this way, this work contributes towards a continuous, non-invasive detection of the users cognitive workload while she/he interacts with Smartphones in their daily activities. This line of work has great potential to impact areas such as design and dynamic adaptation of mobile applications, identification and assessment of individual cognitive differences, determination of appropriate moments to deliver content or interrupt the user, among others.

The extension of this work runs through different lines of action. Firstly, to gather more relevant information by increasing the number of acquired signals. Secondly, to apply this methodology to other user contexts such as the Web. Finally, to improve the machine learning models.

Acknowledgments. This work was financed by Conicyt Fondecyt 11130252 and had the continuous support of “Instituto Sistemas Complejos de Ingeniería” (Conicyt: Proyecto Basal FBO16). Also, the authors would like to thank to Jorge Gaete-Villegas and all the experiment participants.

References

1. Bailey, B. P., & Iqbal, S. T. Understanding Changes in Mental Workload During Execution of Goal-directed Tasks and Its Application for Interruption Management. *ACM Transactions on Computer-Human Interaction*, 14(4) (2008). doi:10.1145/1314683.1314689
2. Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). *Handbook of psychophysiology*. Cambridge University Press (2007).
3. Chen, S., Epps, J., & Chen, F. A comparison of four methods for cognitive load measurement. *In Proceedings of the 23rd Australian Computer-Human Interaction Conference*, 76-79. ACM (2011). doi:10.1145/2071536.2071547
4. Fredericks, T. K., Choi, S. D., Hart, J., Butt, S. E., & Mital, A. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics*, 35(12), 1097-1107 (2005). doi:10.1016/j.ergon.2005.06.002
5. Fritz, T., Begel, A., Mller, S. C., Yigit-Elliott, S., & Zger, M. Using psychophysiological measures to assess task difficulty in software development. *In Proceedings of the 36th International Conference on Software Engineering*, 402-413. ACM (2014). doi:10.1145/2568225.2568266
6. Gwizdka J. Distribution of Cognitive Load in Web Search. *Journal of the American Society for Information Science and Technology*, 61(11), 21672187 (2010). doi:10.1002/asi.21385
7. Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. Psycho-physiological measures for assessing cognitive load. *In Proceedings of the 12th ACM international conference on Ubiquitous computing*, 301-310. ACM (2010). doi:10.1145/1864349.1864395
8. Hart, S. G., & Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183 (1988). doi:10.1016/S0166-4115(08)62386-9
9. Longo, L., Rusconi, F., Noce, L. & Barrett, S. The importance of human mental workload in web design. *In Proceedings of the 8th International Conference on Web Information Systems and Technologies* (2010).
10. Miettinen, M., & Oulasvirta, A. Predicting time-sharing in mobile interaction. *User Modeling and User-Adapted Interaction*, 17(5), 475-510 (2007). doi:10.1007/s11257-007-9033-x
11. Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. *In Proceedings of the SIGCHI conference on Human factors in computing systems* 919-928. ACM (2005). doi:10.1145/1054972.1055101
12. Ryu, K., & Myung, R. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, 35(11), 991-1009 (2005). doi:10.1016/j.ergon.2005.04.005
13. Wickens, C. D. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2), 159-177 (2002). doi:10.1080/14639220210123806
14. Wilson, G. F., & Eggemeier, F. T. Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance*, 329360 (1991).
15. Yin, B., Chen, F., Ruiz, N., & Ambikairajah, E. Speech-based cognitive load monitoring system. *In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2041-2044. IEEE (2008). doi:10.1109/ICASSP.2008.4518041