Indexing Cognitive Load using Blood Volume Pulse Features

Jianlong Zhou
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Jianlong.zhou@data61.csiro.au

Syed Z. Arshad
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Syed.arshad@data61.csiro.au

Simon Luo
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Simon.luo@data61.csiro.au

Kun Yu
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Kun.yu@data61.csiro.au

Shlomo Berkovsky
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Shlomo.berkovsky@data61.csiro.au

Fang Chen
Data61, CSIRO
13 Garden St. Eveleigh
NSW 2015 Australia
Fang.chen@data61.csiro.au

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053140

Abstract
Physiological responses contain rich affective information even when humans are not expressing any external signs. In this paper, we investigate the use of the Blood Volume Pulse (BVP) signals for indexing cognitive load. An experiment, which introduced cognitive load as a secondary task in a decision making context was conducted in the study. BVP signals were analyzed in order to establish relationships between BVP and cognitive load levels. A set of features (e.g. peak and max features) was found to be significantly distinctive across different cognitive load levels. The identified BVP features can be used to set up machine learning models for the automatic classification of CL levels in intelligent systems.

Author Keywords
BVP; cognitive load; peak and max features.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Cognitive load (CL, also known as mental workload) has important implications for various application areas such as adaptive automation and training, driving safety, and military command and control [5]. It refers

to the amount of mental demand imposed on a user by a particular task, and is associated with the limited capacity of the user's working memory and the ability to process novel information [3]. The concept of working memory has set the foundation for the cognitive load analysis, and cognitive load research is often considered as the examination of how and to what extent the working memory is deployed and utilized during a specific cognitive task.

The cognitive load experienced by a user in a task can strongly influence productivity and performance, as high levels of cognitive load is known to cause stress, decrease performance and experience of users, and hinder their ability to learn [2]. Cognitive load measurement (CLM), therefore, plays an important role in applications involving human-machine interactions.

In recent years, a number of methods have been developed to estimate the CL level. These methods can be categorized into four types, ranging from subjective ratings, performance, behavioral to physiological measures [3,20]. Among these, physiological approaches allow CL to be measured at a high rate and with a high degree of sensitivity because of the continuous availability of human physiological signals. Much work has focused on finding physiological features indexing CL, especially emphasizing galvanic skin response (GSR) and electroencephalogram (EEG) features.

Sympathetic activation has been found to cause changes in heart rate, stroke volume and peripheral cardiovascular resistance [19]. These effects can be sensed by monitoring the amount of blood perfusion in a peripheral region of the body, such as the tip of a finger, which is proportional to the opposition presented by blood to the infrared light. This technology is called Blood Volume Pulse (BVP), and it measures the blood volume in the skin capillary bed in the finger with photoplethysmography (PPG). BVP is often used as an indicator of affective processes and emotional arousal, which play an essential role in rational decision making, learning and cognitive tasks [19]. Because of its wide usage (even available on some smart phones or watches), BVP is considered a user-friendly method for obtaining an individual's physiological signals.

This paper aims to investigate the use of BVP features for indexing CL levels in cognitive tasks. Various BVP features are extracted and analyzed, in order to find their correlations with CL. It is found that peak features and max features of the BVP signal showed significant differences at various CL levels, especially for short-time BVP signals where Heart Rate Variability (HRV) is difficult to assess.

Related Work

BVP signal is an ideal means for CL investigation as it is robust, relatively cheap to collect, and unobtrusive to the user. It yields a continuous measure related to arterial activity changes correlated with the sympathetic branch of the neural system. Zhai et al. [19] used BVP to detect stress in HCI applications. BVP signal features such as BVP period (also called inter-beat interval), amplitude, and HRV related frequency features were extracted to index stress. It was observed that the Low Frequency (LF) bands of heart beat signals are consistently related to CL, while high frequency (HF) bands were also shown to be sensitive to mental effort [13]. In general, the heart rate increases and overall HRV decreases when mental effort increases [14]. Kennedy and Scholey [9] found that cognitive processing is associated with higher heart rate values. The 0.1 Hz component of HRV is often considered as an effective measure of mental strain and is likely to indicate emotional strain (stress reactions) or general activation [15].

There has been a large body of work that uses Respiratory Sinus Arrhythmia (RSA) to classify an individual's health or mental state. For example, Healey and Picard [6] found the physiological links between the RSA and a driver's stress level. Although these studies have shown the links between the mental state and RSA changes, these correlations have largely been analyzed manually.

Therefore, BVP, which is used to measure the heart rate and related features, can serve as an objective indicator and an automatic physiological measure, relatively free from demand characteristics and report biases in CLM. However, little work has looked into the use of BVP for indexing user's CL levels.

Experiment

A user experiment was designed and it aimed at determining CL levels from the measured BVP signals.

Task Design

The experiment was designed using water pipe failure prediction task [1,11] replicated in the lab environment. Each subject was asked to make a budget plan, i.e. a budget in terms of pipe length to be inspected, using the failure prediction models learned from the historical pipe failure records, in order to minimize water pipe failure. Two ML models were provided for each estimation task. Participants were required to make decisions by selecting one of two presented ML models and then making a budget estimate based on the selected ML model. CL was induced using the dual-task design: the primary task was predicting the pipe length to check using the data displayed via graphs; and the secondary task was retaining a random sequence of digits for the duration of the primary task [4]. Four increasing levels of cognitive load (labeled as CL1, CL2, CL3 and CL4) were induced using three, five, seven and nine digits to be retained during the primary task and then to be recalled after the primary task. Every subject undertook 36 prediction tasks – 9 under each CL level. The order of tasks was randomized.

Participants

A total of 42 participants were recruited for the study. Ages of participants ranged from 20 to 57 years. The participants were with different background of researchers and administrative staff.

Experimental Conditions & Data Collection

In this study, participants were to perform similar tasks under four increasing cognitive load level conditions (independent variable). The CL levels administered were labeled from CL1 to CL4. BVP devices from ProComp Infiniti of Thought Technology Ltd were used and worn on the proximal part of the middle finger of the left hand to collect blood volume pulse of subjects.

Analysis

BVP Features

An example BVP signal collected in the experiment is shown in Figure 1. As a periodical signal, BVP is associated with three frequency bands: Very Low Frequency (VLF) (0.00-0.04Hz), Low Frequency (LF) (0.05-0.15Hz), and High Frequency (HF) (0.16-0.40Hz). The duration of the BVP signal in this study is under 5 minutes and this is too short to consider VLF activity. On the other hand, LF band reflects sympathetic activity and HF band is related to parasympathetic activity [19]. Kristal-Boneh et al. [10]

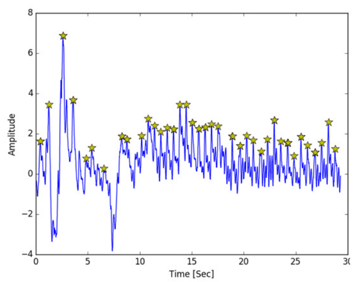

Figure 1: An example of BVP signal.

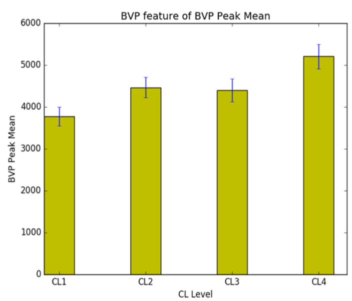

Figure 2: Peak mean BVP among four CL levels of tasks.

showed that when experiencing mental stress, the sympathetic activity increases whereas parasympathetic activity decreases.

The feature extraction process of BVP signals involves the following steps: 1) signal smoothing, 2) signal normalization, 3) extrema detection, and 4) feature encoding.

At the signal smoothing step, the observed signal is convolved with a Hanning window [16]. The window size, which behaves as a cut off frequency, is dictated by the maximal admissible heart rate in a normal situation, namely 200 beats/minute [8][18]. This filtering removes the dicrotic notches from the original signal, because they are not part of the stress reaction [17]. We observed that BVP was highly subjective, and it differed across participants. Therefore, BVP signals are normalized using Z-Normalization, to compensate for subjective differences between various signals before the extrema detection:

$$S_N = \frac{S - \mu}{\sigma} \qquad (1)$$

where μ and σ are the mean and variance of the BVP signal of all tasks conducted by a subject respectively, and S and S_N are the original and normalized BVP signals respectively.

As emotional responses create variations in HRV, the mean and variance of the BVP are extracted as features to index CL. The RSA features such as LF, HF and the LF/HF ratio are extracted to measure the activity of the parasympathetic nervous system. In this work, besides statistical features and frequency features, peak features that reflect the strength of the BVP signals are also extracted. In order to extract peak features (illustrated with stars in Figure 1), extremum detection is performed on the smoothed signal.

In summary, BVP features extracted in this study include:

- Statistical features: mean of the BVP values in a task, variance of the BVP values in a task;
- Peak features: mean of peaks of the BVP signal in a task, variance of peaks of the BVP signal in a task;
- Max features: amplitude of the maximum of the BVP signal in a task (or the largest peak of the signal in a task), time of the maximum of the BVP signal in a task, relative (to the entire task duration) time of the maximum of the BVP signal in a task.
- Frequency features: LF, HF, LF/HF ratio.

BVP Feature Analysis

One-way ANOVA tests with post-hoc analysis using *t*-test were performed to evaluate CL level discrimination with BVP features. The ANOVA tests found that features of peak mean BVP ($F_{3,1248}$=19.685, p<.000), peak variance BVP ($F_{3,1248}$=19.157, p<.000), amplitude of max ($F_{3,1248}$=3.359, p<.018), and time of max ($F_{3,1248}$=7.714, p<.000) showed statistically significant differences among the four CL levels. Figure 2 shows peak mean BVP among four CL levels of tasks. The other three significant features (peak variance, amplitude of max, time of max) demonstrate similar trends.

Post-hoc analyses with *t*-tests were conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0125 (.05/4) for all pairwise differences. For the feature of peak mean BVP, the post-hoc tests found that (see Figure 2), CL1 showed a significantly lower peak mean BVP than CL2 (t =-4.099, p<.000), CL3 (t =-3.487, p<.000), and CL4 (t =-7.619, p<.000). It was also found that CL2 (t =-3.825, p<.000) and CL3 (t =-3.947, p<.000) had a lower peak mean BVP than CL4. The post-hoc t-tests also found similar significant differences among the CL levels for the peak variance BVP feature.

The post-hoc t-tests also found that CL3 (t =-2.6, p<.001) had a significantly lower amplitude of max BVP than CL4. It was found that CL1 had a significantly lower time of max BVP than CL3 (t =-3.727, p<.000) and CL4 (t =-4.662, p<.000).

Discussions and Ongoing Work

Discussions

Our analyses have shown that both peak features and max features of the BVP signal achieved significant differences among tasks at different CL levels. However, frequency features such as LF, HF, or LF/HF did not show significant differences among the CL levels.

This study found that the features of peak mean BVP and peak variance BVP showed significant differences. When the peaks of the BVP signal are analyzed, only the features with the frequency lower than the heart rate, i.e., VLF, LF, and HF are considered. These features are all associated with RSA. From this perspective, by analyzing peaks of the BVP signal, we approximate the features of RSA, which is a measure of the sympathetic and parasympathetic nervous system. Therefore, the significance of the peak features of BVP for CL levels is in line with previous findings [7].

The LF and HF RSA features were not significant, possibly because the tasks were not long enough to capture the RSA cycle. The lower bound of the RSA feature is 0.04Hz, i.e., 25 seconds are required to capture the full period of RSA. To get an accurate measurement of the activity of parasympathetic nervous system, more cycles of the signal are required. In our experiment, the tasks ranged from 20 to 120 seconds, which may not suffice to capture the activity of parasympathetic nervous system.

The mean and variance of the peaks can be seen as a proxy for HF and LF, as the changes in the peaks represent the changes in RSA. A larger variance may indicate an increased sympathetic and parasympathetic activity. If the value of both activities is higher than normal, this will result in a higher mean and variance of the signal.

The max features were also found significantly different across the CL levels. A maximum in the BVP signals could occur when the change in the sympathetic and parasympathetic activity is largest. When a person experiences an extraordinary emotional state, the activation of the sympathetic division of the nervous system helps the body to better cope with this state [12]. These changes can be reflected in the BVP signals with the variation of stimulations. The maximum of the BVP signal could also be due to human body movement, which could create artifacts in the signals that are detected as a maximum.

We acknowledge that physiological signals such as BVP may be affected by emotion besides cognition. In this study, we assume that subject's emotion keeps constant during the short task period, and therefore, the BVP changes mainly reflect cognition variations during task time.

Ongoing Work

Our ongoing work focuses on the development of machine learning models to classify CL levels based on the identified BVP features. Different widely used classifiers such as Random Forest (RF), Naïve Bayes (NB), AdaBoost (AB), and Support Vector Machine (SVM) will be firstly applied to evaluate the discrimination of BVP features in indexing CL levels. More advanced customized classifiers will then be developed to accurately classify CL levels.

Besides, we plan to fuse multiple modalities such as GSR, pupillary responses and BVP for improving the CL classification performance. Our ultimate goal is to develop automatic and real time CL measurement approaches during task time.

Implications

The findings in this paper can be used in various HCI applications, where non-intrusive devices, e.g., smart phones or watches, are used to record the BVP signals. By analyzing the signals, the user's CL level can be discovered in real time. The BVP based CLM has advantages of easily accessible and low cost and it is easier to implement than GSR or eye-tracker based CLM approaches.

Conclusions and Future Work

This paper investigated the use of BVP features for indexing CL levels. It was found that peak and max BVP features were significantly distinctive among cognitive tasks. This research suggested new directions and potentials of using everyday human hand-held/worn devices such as smart phones, watches for robust, relatively cheap, and unobtrusive human cognitive load monitoring.

In the future, advanced CL load classification models will be developed to index CL levels based on BVP features. We also plan to fuse multiple modalities such as BVP, GSR, and pupil information for improving the CL classification performance.

Acknowledgements

This work was supported in part by the Asian Office of Aerospace Research & Development (AOARD) under grant No. FA2386-14-1-0022 AOARD 134131.

References

1. Syed Z. Arshad, Jianlong Zhou, Constant Bridon, Fang Chen, and Yang Wang. 2015. Investigating User Confidence for Uncertainty Presentation in Predictive Decision Making. *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, 352–360.
2. Paul Chandler and John Sweller. 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and instruction* 8, 4: 293–332.
3. Fang Chen, Jianlong Zhou, Yang Wang, et al. 2016. *Robust Multimodal Cognitive Load Measurement*. Springer.
4. Cary Deck and Salar Jahedi. 2015. The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review* 78: 97–119.
5. Marc Grootjen, Mark A. Neerincx, Jochum C. M. van Weert, and Khiet P. Truong. 2007. Measuring Cognitive Task Load on a Naval Ship: Implications of a Real World Environment. In D.D. Schmorrow and L.M. Reeves, eds., *Foundations of Augmented Cognition*. Springer, 147–156.
6. J. Healey and R. Picard. 2000. SmartCar: detecting driver stress. *Proceedings of 15th International Conference on Pattern Recognition 2000*, 218–221.
7. Timm Hörmann, Marc Hesse, Peter Christ, Michael Adams, Christian Menßen, and Ulrich Rückert. 2016. Fine-Grained Prediction of Cognitive

Workload in a Modern Working Environment by Utilizing Short-Term Physiological Parameters: *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*, 42–51.

8. M. J. Karvonen, E. Kentala, and O. Mustala. 1957. The effects of training on heart rate: a longitudinal study. *Annales Medicinae Experimentalis Et Biologiae Fenniae* 35, 3: 307–315.
9. D. Kennedy and A. Scholey. 2000. Glucose administration, heart rate and cognitive performance: effects of increasing mental effort. *Psychopharmacology* 149, 1: 63–71.
10. E. Kristal-Boneh, M. Raifel, P. Froom, and J. Ribak. 1995. Heart rate variability in health and disease. *Scandinavian Journal of Work, Environment & Health* 21, 2: 85–95.
11. Zhidong Li, Bang Zhang, Yang Wang, et al. 2014. Water Pipe Condition Assessment: A Hierarchical Beta Process Approach for Sparse Incident Data. *Machine Learning* 95, 1: 11–26.
12. Frederic H. Martini, Judi L. Nath, and Edwin F. Bartholomew. 2011. *Fundamentals of Anatomy & Physiology*. Benjamin Cummings, San Francisco.
13. G. Mulder. 1985. Attention, effort and sinus arrhythmia: how far are we? In J.F. Orlebeke, G. Mulder, and L.P.J. van Doornen, eds., *The psychophysiology of cardiovascular control*. Plenum Press, New York, 407–424.
14. L.J.M. Mulder. 1992. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology* 34, 2–3: 205–236.
15. P. Nickel and F. Nachreiner. 2000. Psychometric Properties of the 0.1Hz Component of HRV as an Indicator of Mental Strain. *Proceedings of the IEA 2000/HFES 2000: the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*.
16. Alan V Oppenheim and Ronald W Schafer. 2010. *Discrete-time signal processing*. Pearson, Upper Saddle River.
17. Erik Peper, Rick Harvey, I-Mein Lin, Hana Tylova, and Donald Moss. 2007. Is there more to blood volume pulse than heart rate variability respiratory sinus arrhythmia, and cardiorespiratory synchrony? *Biofeedback* 35, 2: 54–61.
18. H. Tanaka, K. D. Monahan, and D. R. Seals. 2001. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology* 37, 1: 153–156.
19. J. Zhai, A.B. Barreto, C. Chin, and Chao Li. 2005. Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. *Proceedings of IEEE SoutheastCon 2005*, 415–420.
20. Jianlong Zhou, Ju Young Jung, and Fang Chen. 2015. Dynamic Workload Adjustments in Human-Machine Systems Based on GSR Features. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2015*. Springer International Publishing, 550–558.