Evaluating cognitive task result through heart rate pattern analysis

Juan Yu[1], *Guang Yuan Liu*[1,2,3] ✉, *Wan Hui Wen*[1,2], *Chuan Wu Chen*[1]

[1]*School of Electronics and Information Engineering, Southwest University, Chongqing 400700, People's Republic of China*
[2]*Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Chongqing 400700, People's Republic of China*
[3]*Chongqing Collaborative Innovation Center for Brain Science, Chongqing, 400700, People's Republic of China*
✉ *E-mail: liugy@swu.edu.cn*

Published in Healthcare Technology Letters; Received on 11th August 2018; Revised on 31st August 2019; Accepted on 18th November 2019

The measurement of the right and wrong results of cognitive tasks has important applications in many commercial and educational areas such as the drivers' training system, the simulation training and online learning system. This Letter aims to distinguish the heartbeat pattern of cognitively wrong responses to that of cognitively right responses based on the electrocardiogram (ECG) through 36 subjects with different professional backgrounds. The experimental design methods were double-digit and five-digit addition/subtraction, which were blindly selected by subjects from a black box. Through the R–R interval (RRI) series obtained from the ECG data, some linear, nonlinear and moment features were extracted to evaluate the cognitive task results by using pattern recognition methods. The binary classification of RRI datasets indicated that autonomic nerve patterns of the right and wrong cognitive heartbeat responses were distinguishable.

1. Introduction: The measurement of cognitive task results not only has important commercial and educational applications in evaluating the users' performance in cognitive tests [1], but also can enable real-time personalised content generation for distant learning or usability testing of applications related to human–machine interactions [2]. For example, the mental and physical states of a driver have been widely recognised as the crucial point in every issue concerning the development of models headed to improve vehicle safety. As a result, almost all the new in-vehicle technologies, currently being developed at a high rate, fertilise devices to monitor the driver's psychophysiological state in time [3]. In the literature [4], a driving support system was designed to detect the lack of driver's awareness of the real-time traffic environment by tracking the driver's physiological information such as gaze, head rotation directions and the intervals between R waves [hereafter, R–R interval (RRI)] in an electrocardiogram (ECG) waveform. The literature [5] presented a method for automatic classification of drivers' mental states using facial action units as input features, which was similar to the method of cognitive task result in this Letter. Simulation training has been an effective way to complement the clinical training of medical students. The literature [6] was conducted to explore the relationships between emotion and cognitive load, and diagnostic performance during simulation training. The distinguish of autonomic nerve patterns of the right and wrong cognitive heartbeat responses was also helpful for the evaluation of simulation training. In [7], a multimedia teaching system for English courses arranged learners' time and learning materials, monitored their learning state in time, summarised knowledge, deepened learners' memory and improved learning efficiency. Assessing cognitive task result is essential, as it contributes to understanding the complexity of the learning task. New learning paradigms in [8, 9] have hypothesised that a system, which monitors working memory capacity in real time and accordingly adjusts training difficulty can improve learning efficiency.

Previous research have shown that high and low cognitive loads can be distinguished by using physiological features, e.g. galvanic skin response [10], electroencephalography [11] and blood volume pulse [12]. Cognitive tasks targeting different levels of cognitive difficulty include the Stroop test, math test and event recall test [12]. Conventionally, the cognitive task performances of many tests are usually evaluated by human teachers. However, it is not convenient for the training or teaching system to involve the participation of human teachers. Therefore, this Letter tried ECG signal for automatic cognitive task result evaluation by the machines.

In terms of classification, Huang *et al.* [13, 14] put forward the extreme learning machine (ELM), a learning algorithm for single-hidden layer feedforward neural network, which not only averts falling into local optima, but also achieves faster learning speed and better generalisation performance than traditional feedforward neural network learning algorithm.

The literature [15] analysed the same everyday activities of five people working together based on the ECG signal. Therefore, in this Letter, we tried to apply math test as the cognitive task, which is equivalent to simulate a working environment, acquired the ECG signal of the subjects [16] while they were doing the cognitive tasks, extracted RRI features and distinguished the RRI data corresponding to the wrong cognitive task result from those corresponding to the right cognitive task result. The ELM was applied for pattern recognition of the RRI data.

The rest of this Letter is organised as follows: Section 2 describes the experiment design; Section 3 presents the data analysis; and Sections 4 and 5 are the discussion and conclusion.

2. Experiment settings: The cognitive tasks applied in the current work were double-digit and five-digit addition/subtraction math tests. The subjects and the experiment procedure will be introduced in this section.

2.1. Subjects: By advertising at various colleges in Southwest China University, a group of ordinary college students (17–24 years old) with different professional backgrounds voluntarily responded to the recruitment. The most basic inclusion criterion for the subject was without the history of the adverse disease. Before the experiment, the subjects signed the informed consent and were informed of the procedure of the experiment. They were also required to have a good sleep the night before the experiment. About 36 subjects took part in this experiment.

2.2. Experiment procedure: To avoid contingency, each subject needed to participate in six rounds of experiments, three times for double-digit addition/subtraction and another for five-digit addition/subtraction, and there were three sections in each time.

Six sets of test papers were numbered from one to six, and the numbers were written on six sheets of paper, which were placed in a black box. Subjects could only blindly select test papers to conduct experiments, achieving the purpose of experimental randomness. For the test papers, we selected arithmetic problems from a test library, a database for psychological research in the School of Psychology, Southwest China University, to form three sets of two-digit addition/subtraction test papers, in which the same number of operands and the same arrangement of the plus and minus signs were to ensure the principle of the same difficulty. Then, we combined with the principle of maximising the differences between the questions in the same test paper to avoid the influence of the law of the test paper itself on the correct rate, which can also reflect certain randomness. Furthermore, the five-digit addition/subtraction test papers were also composed according to this standard.

Before the start of the first section, a Shimmer3 ECG device was connected to the subject to acquire an ECG signal at a sampling frequency of 512 Hz. In the first section, the subjects were asked to sit and relax for 5 min, and the ECG data acquired in this section served as the baseline data. In the second stage, the subjects randomly selected arithmetic tasks (double-digit or five-digit addition/subtraction math test) and then performed tasks up to 20 min; meanwhile, the experimental time was precisely controlled through the alarm clock. Finally, the subjects were still required to sit and relax for 5 min. The ECG data were acquired throughout the whole experiment. After the experiment, the ECG data was checked to exclude the electrode failure data out of the normal ECG data. To unify the data analysis standards, for two types of arithmetic task groups (double-digit math task group and five-digit math task group), we only analysed the data of subjects with three complete experimental data. Therefore, according to the above screening principles, only the normal ECG data of 20 subjects in the double-digit math task group and those of 27 subjects in the five-digit math task group were obtained for further analysis.

To categorise the ECG data samples according to the right or wrong cognitive task results, the math test papers of the subjects were evaluated by the experimenter, and the test time corresponding to the right and wrong test results were manually marked.

3. Data analysis

3.1. ECG data preprocess: To get the RRI series from the ECG data, the peaks of the R waves in the ECG signal were located by an automatic R peak locating algorithm presented in [17]. Then, the RRI series were calculated from the intervals between two consecutive R peaks. The wrong RRIs due to the state renewal of the Shimmer3 ECG device were eliminated from the normal RRI series.

Since the subjects' baseline RRI level varied from person to person, the following formula was applied to eliminate the baseline difference:

$$\boldsymbol{Y}[i] = \boldsymbol{X}[i] - \boldsymbol{M}; \quad i = 1, 2, 3, \ldots, n \tag{1}$$

where $\boldsymbol{M}$ means the average of the baseline RRI series of a subject and $\boldsymbol{X}$ represents the RRI series during the cognitive task.

Through (1), a new series called RRI_t series could be obtained for further analysis. We have extracted several pieces of data of each subject with right and wrong cognitive heartbeat responses for comparative analysis.

3.2. Feature extraction: In this Letter, we extracted the common statistical features, LZ complexity, the energy features based on wavelet packet decomposition and the n-order ($n = 0$–19) Legendre moments of RRI_t series. In terms of statistical characteristics, the mean, the standard deviation, the first and second differences between consecutive RRI_t and the root mean square (RMS) were extracted.

3.3. Two sample t-test and sign test: To detect whether the two samples existed significant difference or not for five-digit or double-digit addition/subtraction math test, two sample t-test and sign test were done for the features. First, it was necessary to verify whether the feature samples conformed to the normal distribution, and then perform t-test on the samples subjected to the normal distribution and carry out sign test to the non-normal distribution samples. The significant level was set to 0.01. The test results were shown in Tables 1 and 2 for double-digit addition/subtraction and Tables 3 and 4 for five-digit addition/subtraction, from which we could know that there existed significant difference for some features.

3.4. Feature selection and data classification: In the previous section, we have extracted 30 features, where they probably contained features that were either redundant or irrelevant

Table 1 Results of t-test for double-digit addition/subtraction math test

Feature	H	P	Tstat	Df	Sd
mean	1	4.6×10^{-6}	4.8368	118	0.0697
Std	0	0.8496	4.1774	118	0.0169
2diff	0	0.8874	4.5053	118	0.0205
3diff	0	0.7869	3.8229	118	0.0224
E20	1	0.0069	−1.5099	118	0.1472
E30	0	0.1248	−1.1570	118	0.1596
Lp0	1	2.01×10^{-6}	4.8368	118	0.0704
Lp2	1	0.0014	3.0539	118	0.0417
Lp3	0	0.8521	−1.0502	118	0.0276
Lp4	1	0.0022	2.2799	118	0.0323
Lp6	0	0.6177	−0.3001	118	0.0303
Lp7	0	0.7753	−0.7590	118	0.0287
Lp8	1	0.0099	1.7669	118	0.0359
Lp9	0	0.0593	1.5720	118	0.0311
Lp10	1	0.0027	2.8400	118	0.0373
Lp12	0	0.5742	−0.1875	118	0.0286
Lp13	0	0.9274	−1.4666	118	0.0389
Lp14	1	0.0024	2.8758	118	0.0434
Lp16	0	0.0476	1.6818	118	0.0515
Lp18	0	0.0798	1.4156	118	0.0529
Lp19	0	0.6681	−0.4358	118	0.0448

$H = 0$ is the null hypothesis, $H = 1$ is the alternative hypothesis, P is the probability that the null hypothesis is adopted, Tstat is the value of the t-test statistic, Df is the degrees of freedom and Sd is the standard deviation.

Table 2 Results of sign test for double-digit addition/subtraction math test

Feature	H	P	Sign
RMSSD	1	5.21×10^{-9}	52
LZ	0	0.7838	28
E10	1	0.0073	21
E50	0	0.6989	28
Lp1	1	0.0073	21
Lp5	0	0.5190	27
Lp11	0	0.2451	25
Lp15	0	0.8974	29
Lp17	0	0.0519	22

$H = 0$ is the null hypothesis, $H = 1$ is the alternative hypothesis, P is the probability that the null hypothesis is adopted and Sign is sign statistic value.

Healthcare Technology Letters, 2020, Vol. 7, Iss. 2, pp. 41–44
doi: 10.1049/htl.2018.5068

Table 3 Results of t-test for five-digit addition/subtraction math test

Feature	H	P	Tstat	Df	Sd
mean	1	0.0024	2.8563	160	0.0469
Std	0	0.7856	5.0883	160	0.0172
E10	1	0.0096	2.0911	160	0.0865
E20	1	5.54×10^{-5}	−3.9642	160	0.1290
E30	1	1.65×10^{-5}	−4.2724	160	0.1293
Lp0	1	0.0024	2.8563	160	0.0474
Lp1	0	0.8512	−4.3362	160	0.0297
Lp2	0	0.7184	−0.5792	160	0.0277
Lp3	0	0.4834	0.0417	160	0.0269
Lp4	0	0.3473	0.3934	160	0.0301
Lp5	0	0.4529	0.1185	160	0.0335
Lp6	1	0.0090	2.0924	160	0.0337
Lp8	0	0.3504	0.3849	160	0.0325
Lp9	0	0.9380	−1.5460	160	0.0353
Lp10	0	0.3582	0.3639	160	0.0359
Lp11	0	0.6982	−0.5203	160	0.0379
Lp12	1	0.0094	0.9203	160	0.0402
Lp13	0	0.9084	−1.3366	160	0.0360
Lp14	0	0.1909	0.8768	160	0.0382
Lp15	0	0.8097	−0.8793	160	0.0420
Lp16	1	0.0059	1.2002	160	0.0497
Lp17	0	0.9915	−2.4100	160	0.0453
Lp18	1	0.0081	2.4291	160	0.0445
Lp19	0	0.5357	−0.0897	160	0.0478

Table 4 Results of sign test for five-digit addition/subtraction math test

Feature	H	P	Sign
2diff	1	1.18×10^{-11}	70
3diff	1	8.62×10^{-10}	66
RMSSD	1	3.81×10^{-10}	68
LZ	0	0.0817	40
E50	1	0.0073	28
Lp7	0	1	41

and should be removed from the set of original features. It meant that feature selection was inevitable, which could eliminate irrelevant or redundant features, to decrease the number of features, improve the precision of the model and lessen running time. Sequential backward selection (SBS) [18] was the choice in this Letter. Its basic idea was to remove one of the remaining features at a time from the complete set of features so that the post-evaluation function value was optimal.

As the SBS is a wrapper-type algorithm, the accuracy of classification was applied as a criterion to judge if the model of feature subsets was good or bad. ELM was chosen as the classifier, and leave-one–subject-out cross-validation [19] was employed for the model's training and testing. Since the recognition was to distinguish autonomic nerve patterns of right and wrong cognitive heartbeat responses, thus the sum of false-negative rate and false-positive rate could be regarded as the standard to judge the feature subsets' performance for the recognition. Moreover, the error rate of each iteration's optimal solution could be calculated.

The results shown that the error rate of the 26th iteration's optimal solution was the lowest, the corresponding feature subset contained the first and second differences between consecutive RRI_t, the RMS and n-order ($n = 0$, 1) Legendre moments for double-digit addition/subtraction math test. Besides, the first and second differences between consecutive RRI_t and n-order ($n = 1$, 17) Legendre moments of the RRI_t series were the optimal feature subset for five-digit addition/subtraction.

Table 5 Confusion matrix of cognitive task results for double-digit addition/subtraction

	Right task result, %	Wrong task result, %
classified as the right task result	84.75	15.25
classified as the wrong task result	15.25	84.75

Table 6 Confusion matrix of cognitive task results for five-digit addition/subtraction

	Right task result, %	Wrong task result, %
classified as right task result	80	20
classified as wrong task result	20	80

According to the feature subset, the classification rate of autonomic nerve patterns of right and wrong cognitive heartbeat responses could be obtained, which were shown in Tables 5 and 6 for double-digit and five-digit addition/subtraction. The results indicated that autonomic nerve patterns of right and wrong cognitive heartbeat responses were distinguishable.

4. Discussion: Autonomic control of the heart is achieved through sympathetic and parasympathetic effects on the cardiac muscle, adjusting the duration between sequential heartbeats. Many existing studies have found that tension could cause excess sympathetic activation and parasympathetic withdrawal [20]. Therefore, to explore whether the right and wrong cognitive heartbeat responses were distinguishable, we evaluated cognitive task results through heart rate pattern analysis in this Letter. Moreover, the related RR interval was applied to investigate the influence of different cognitive task results on the heart. This was based on the fact that the RR interval is highly credible and assures the randomness property [21]. The advantage of this method over previous methods was that previous methods did not conduct research on cognitive task result through ECG signal. For example, in the literature [22], initiatory experimental consequences, which performed on only four channels of EEG, manifested that the proposed system was capable of precisely detecting the cognitive workload of the driver with tremendous potential for improvement by using deep learning on EEG signal. Addition to the methods on EEG, the methods on ECG developed in the present Letter are valuable because of their easy accessibility to the accurate and convenient measurement of physiological information.

From the test results of all features, we could know that the heart rate patterns of autonomic activity were distinguishable through some features between right and wrong cognitive task result states. However, due to the imbalance of the gender distribution in the school, there were fewer male participants in this experiment, which can be improved in future research.

5. Conclusion: This Letter suggested using RR intervals to determine whether autonomic nerve patterns of right and wrong cognitive heartbeat responses were distinguishable. A total of 30 features were extracted such as the mean and Legendre moment. SBS was the choice for feature selection. ELM and leave-one–subject-out cross-validation were applied for recognition and classification. The result indicated that autonomic nerve patterns of right and wrong cognitive heartbeat responses were distinguishable, and have been demonstrated to be of far-reaching significance in plenty of areas such as driver monitoring system and learning system.

6. Acknowledgments: This work was supported by the National Science Foundation of China (Grant nos. 61872301 and 61472330). We sincerely thank the students of Southwest University for their help during signal collections.

7 References

[1] Pega Z., Epps J., Chen F.: 'Spectral EEG features for evaluating cognitive load'. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, 2011, pp. 3841–3844

[2] Das D., Chatterjee D., Sinha A.: 'Unsupervised approach for measurement of cognitive load using EEG signals'. IEEE Int. Conf. Bioinformatics and Bioengineering, Institute of Electrical and Electronics Engineers, 2013, pp. 1–6

[3] Sena P., Pappalardo M., Pellegrino A., ET AL.: 'Studying the influence of cognitive load on driver's performances by a fuzzy analysis of lane keeping in a drive simulation'. Seventh IFAC Symp. Advances in Automotive Control, 2013, pp. 151–156

[4] Masahiro M., Haruki K., Koji O., ET AL.: 'Study on driver distraction detection using pattern recognition for the reduction of traffic accidents'. 16th World Congress on Intelligent Transport Systems and Services, 2009, vol. **17**, no. 1, pp. 57–59

[5] Ma Z.Y., Mahmoud M., Robinson P., ET AL.: 'Automatic detection of a driver's complex mental states'. 17th Int. Conf., 2017, pp. 678–691

[6] Fraser K., Ma I., Teteris E., ET AL.: 'Emotion, cognitive load and learning outcomes during simulation training', *Med. Educ.*, 2013, **47**, (8), pp. 1055–1062

[7] Li H.M.: 'Multimedia teaching design for English course based on cognitive load theory', *Agro Food Ind. Hi Tech*, 2017, **28**, (1), pp. 3138–3140

[8] Coyne J.T., Carryl B., Anna C., ET AL.: 'Applying real-time physiological measures of cognitive load to improve training'. Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience – Fifth Int. Conf., 2009, pp. 469–478

[9] Mazher M., Aziz A.A., Malik A.S., ET AL.: 'An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence', *Dep. Electr. Electron. Eng.*, 2017, **5**, (99), pp. 14819–14829

[10] Nargess N., Wang Y., Chen F., ET AL.: 'Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks'. 24th Australian Computer–Human Interaction Conf., 2012, pp. 420–423

[11] Hussain M.S., Calvo R.A., Chen F.: 'Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference', *Interact. Comput.*, 2014, **26**, (3), pp. 256–268

[12] Gokay R., Masazade E., Aydin C., ET AL.: 'Emotional state and cognitive load analysis using features from BVP and SC sensors'. 2015 IEEE Int. Conf. Multisensor Fusion and Integration for Intelligent Systems, 2015, pp. 178–183

[13] Huang G.B., Zhu Q.Y., Siew C.K.: 'Extreme learning machine: a new learning scheme of feed-forward neural networks'. Proc. 2004 IEEE Int. Joint Conf. Neural Networks, 2004, pp. 985–990

[14] Huang G.B., Zhu Q.Y., Siew C.K.: 'Extreme learning machine: theory and applications', *Neurocomputing*, 2006, **70**, (1), pp. 489–501

[15] Szczepanski A., Saeed K., Riener A., ET AL.: 'Wearable computers: a study on ECG signal changes during everyday activities'. 2011 Int. Conf. Biometrics and Kansei Engineering, 2011, vol. **1**, pp. 169–174

[16] Wen W.H., Qiu Y.H., Liu G.Y.: 'Electrocardiography recording, feature extraction and classification for emotion recognition'. World Congress on Computer Science & Information Engineering, 2009, vol. **4**, pp. 168–172

[17] Huang H.B., Xie Y.: 'The adaptive detection method of R wave in ECG signal', *Autom. Inf. Eng.*, 2013, **34**, (6), pp. 35–39

[18] Singh P., Verma A., Chaudhari N.S.: 'An experimental evaluation of feature selection based classifier ensemble for handwritten numeral recognition'. 2014 Int. Conf. Electronics and Communication Systems, 2014, pp. 1–8

[19] Jing F., Joshua W.W., Alexandra P.K., ET AL.: 'EEG-based affect and workload recognition in a virtual driving environment for ASD intervention', *IEEE Trans. Biomed. Eng.*, 2018, **65**, (1), pp. 43–51

[20] Wen W.H., Liu G.Y., Mao Z.H., ET AL.: 'Toward constructing a real-time social anxiety evaluation system: exploring effective heart rate features', *IEEE Trans. Affective Comput.*, 2018, **9**, pp. 1–10

[21] Moosavi S.R., Nigussie E., Levorato M., ET AL.: 'Low-latency approach for secure ECG feature-based cryptographic key generation', *IEEE Access*, 2018, **6**, pp. 428–442

[22] Almogbel M.A., Dang A.H., Kameyama W.: 'EEG-signals based cognitive workload detection of vehicle driver using deep learning'. IEEE 20th Int. Conf. Advanced Communication Technology, 2018, pp. 256–259

doi: 10.1049/htl.2018.5068