# Estimating Cognitive Load Complexity Using Performance and Physiological Data in a Driving Simulator


Joonwoo Son
Daegu Gyeongbuk Institute of Science & Technology
HumanLAB
Deagu, South Korea
+82 53 785 4740
json@dgist.ac.kr

Myoungouk Park
Daegu Gyeongbuk Institute of Science & Technology
HumanLAB
Deagu, South Korea
+82 53 785 4751
violet1211@dgist.ac.kr



## ABSTRACT
This paper suggests an algorithm for estimating driver's cognitive workload complexity using driving performance and physiological data. The algorithm adopts radial basis probabilistic neural networks (RBPNN) to construct estimation models. In this study, combinations of two driving performance data including standard deviation of lane position (SDLP) and steering wheel reversal rate (SRR), and two physiological signals including heart rate (HR) and skin conductance level (SCL) were considered as measures of cognitive workload. Data for training and testing the RBPNN models were collected in a driving simulator in which fifteen participants drove through a highway and were asked to complete auditory recall tasks which consist of three levels of difficulty. The best performing model, which uses SDLP and SCL data over a 20s-window, could identify four graded levels of cognitive workload with an average accuracy of 85.6%. The results demonstrated that the model using SDLP and SCL was outperforming than the other combinations among performance and physiological measures.



## Categories and Subject Descriptors
J.4 [Social and Behavioral Sciences]

## General Terms
Algorithms, Human Factors

## Keywords
Cognitive Workload, Cognitive Workload Estimation, Driving Performance, Physiology, Neural Network


## 1. INTRODUCTION
Identification of a driver's workload and spare capacity is crucial in the design of adaptive automotive user interface [1]. By monitoring driver's workload, the adaptive interface system can provide timely and affordable information when the driver has the spare capacity to understand and respond to it.

Workload refers to the amount of resources that is required to perform a particular task. Two major types of driving workload are visual and cognitive workload [2]. Visual workload is straightforward, but cognitive workload is difficult to measure directly because it is essentially internal to the driver [3].



*AutomotiveUI'11, November 29-December 2, 2011, Salzburg, Austria Adjunct Proceedings*


Nevertheless, there have been efforts to measure cognitive workload using subjective measures, physiological measures [4], eye movement measures [5], and driving performance measures [1]. Among those measures, driving performance measures can detect the cognitive workload using easy and less expensive methods through readily available in-vehicle information [6]. However, driving performance measures are known to have limitations compared to others due to small changes according to the cognitive workload. That is, the performance measures are able to detect a high cognitive workload condition, but their classification accuracy is not enough to distinguish three levels of cognitive difficulty, which were applied in this study as secondary tasks to add cognitive workload [7].

In the meantime, physiological measures have been proposed as useful metrics for assessing workload in the user interface design and optimization process [3]. Mehler et al. found that when the three levels of task difficulty were randomly ordered, and a recovery interval was provided between tasks, a near linear increase in heart rate and skin conductance appeared across the demand levels [3]. For both heart rate and skin conductance, each task level was statistically differentiated from single task driving and from each other. These findings demonstrate that both heart rate and skin conductance provide the sensitivity to discriminate incremental changes in cognitive workload.

Thus, this paper suggested an algorithm for estimating driver's cognitive workload using driving performance and physiological data, especially the standard deviation of lane position (SDLP) and steering wheel reversal rate (SRR) among performance measures, and heart rate (HR) and skin conductance level (SCL) in physiological data. The results demonstrated that the combination of performance and physiological data, especially SDLP and SCL, can be effectively used as inputs of cognitive estimation models which can distinguish driving only and three levels of dual task conditions at the high accuracy rate.

## 2. MEASURES AND MODELS FOR COGNITIVE LOAD ESTIMATION

### 2.1 Driving Performance Measures
Some studies have shown that cognitive distraction undermines driving performance by disrupting the allocation of visual attention to the driving scene and the processing of attended information. Consequently, cognitive workload leads to significantly reduced lane keeping variation and increased response times to sudden obstacles. In this paper, therefore, lateral controllability was used as driving performance measures under

cognitive workload. The lateral position variation and steering wheel activity were selected to assess lateral controllability.

#### *2.1.1 Lateral position variation*

Lateral position variation is one of the most commonly used driving behavior metrics. Reduced variation in lateral position when engaged with a cognitive task could be interpreted as a symptom of driver overload and increased risk of incorrect decisions. Lateral position variation can be calculated as the standard deviation of lateral position (SDLP). In this paper, a high pass filter with 0.1 Hz cut off frequency is applied on lane position data to reduce dependency of data length.

#### *2.1.2 Steering wheel activity*

Cognitive secondary tasks yield increased steering activity. The increase is mainly in smaller steering wheel movements, the majority of which are smaller than 1 degree. The steering wheel reversal rate can be used for measuring the increase of smaller steering wheel movements. It is defined as the number, per minute, of steering wheel reversals larger than a certain minimum angular value, i.e. 0.1 degree.

### 2.2 Physiological Measures

In some situations physiological indices may be more sensitive than performance-based measures for detecting initial changes in mental workload [3-4]. That is, physiological measures may show increased activation before the appearance of significant performance decrements. Mehler et al. selected heart rate and skin conductance (sweat gland activity) as primary measures of interest, because those measures can indicate changes or differences in relative workload before, or in the absence of, significant performance-level effects [3].

#### *2.2.1 Cardiovascular activity*

Basic cardiovascular measures (heart rate and blood pressure) have been shown to increase with escalating cognitive demand or workload in a range of environments [8-9]. Brookhuis and De Waard [9] reported that heart rate increased with heightened task demand, such as entering a traffic circle, and dropped as task demands decreased, for instance, driving on a two-lane highway. Thus, heart rate, the number of heart beats per unit time, usually per minute, was selected as a physiological measure to estimate cognitive workload complexity.

#### *2.2.2 Electrodermal activity*

Electrodermal activity (EDA) refers to the electrical changes in the skin and can be distinguished in tonic and phasic activity. Tonic EDA, the Electrodermal Level (EDL) or Skin Conduction Level (SCL), is the average level of EDA or baseline activity. Phasic EDA includes the Electrodermal Response (EDR), which is most similar to the formerly common measure Galvanic Skin Resistance (GSR). Mehler et al. suggested that skin conductance clearly documented a change in physiological arousal associated with the increasing complexity of auditory n-back tasks [3], although these findings are contrasted with the HASTE project findings, which found skin conductance is sensitive to increases in visual but not auditory secondary tasks during simulation [8].

### 2.3 Radial-basis Probabilistic Neural Network Models

Radial basis probabilistic neural networks (RBPNN) are applied for estimating driver's cognitive workload using driving performance and physiological measures. It is known that RBPNNs are suitable for classification problems such as cognitive workload complexity estimation. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a complete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes.

## 3. MODEL CONSTRUCTION

### 3.1 Data Source

#### *3.1.1 Experimental setup*

The experiment was conducted in the DGIST fixed-based driving simulator, which incorporated STISIM Drive™ software and a fixed car cab. The virtual roadway was displayed on a 2.5m by 2.5m wall-mounted screen at a resolution of 1024 x 768. Sensory feedback to the driver was also provided through auditory and kinetic channels. Distance, speed, steering, throttle, and braking inputs were captured at a nominal sampling rate of 30 Hz. Physiological data were collected using a MEDAC System/3 unit and NeuGraph™ software (NeuroDyne Medical Corp., Cambridge, MA). A display was installed on the screen beside the rear-view mirror to provide information about the elapsed time and the distance remaining in the drive.

#### *3.1.2 Subject*

Subjects were required to meet the following criteria: age between 25-35, drive on average more than twice a week, be in self-reported good health and free from major medical conditions, not take medications for psychiatric disorders, score 25 or greater on the mini mental status exam [11] to establish reasonable cognitive capacity and situational awareness, and have not previously participated in a simulated driving study. The sample consisted of 15 males, who are in the 25-35 age range (M=27.9, SD=3.13).

#### *3.1.3 Cognitive workload*

An auditory delayed digit recall task was used to create periods of cognitive demand at three distinct levels. This form of n-back task requires participants to say out loud the nth stimulus back in a sequence that is presented via audio recording [11]. The lowest level n-back task is the 0-back where the participant is to immediately repeat out loud the last item presented. At the moderate level (1-back), the next-to-last stimuli is to be repeated. At the most difficult level (2-back), the second-to-the-last stimulus is to be repeated. The n-back was administered as a series of 30 second trials consisting of 10 single digit numbers (0-9) presented in a randomized order at an inter-stimulus interval of 2.1 seconds. Each task period consisted of a set of four trials at a defined level of difficulty resulting in demand periods that were each two minutes long.

#### *3.1.4 Procedure*

Following informed consent and completion of a pre-experimental questionnaire, participants received 10 minutes of driving experience and adaptation time in the simulator. The simulation was then stopped and participants were trained in the n-back task while remaining seated in the vehicle. N-back training continued until participants met minimum performance criteria. Performance on the n-back was subsequently assessed at each of the three demand levels with 2 minute breaks between each level. When the simulation was resumed, participants drove in good weather through 37km of straight highway. Minutes 5 through 7 were used as a single task driving reference (baseline). Thirty seconds later, 18 seconds of instructions introduced the task (0, 1 or 2-back). Each n-back period was 2 minutes in duration (four 30 second trials). Two minute rest/recovery periods were provided before presenting instructions for the next task. Presentation order of the three levels of task difficulty was randomized across participants.

## 3.2 Model Characteristics and Training

### *3.2.1 Definition of cognitive workload*

The cognitive workload was classified into four categories based on primary and secondary task complexity. The secondary tasks, so called n-back tasks, have three levels of difficulty. The 0-back task is a low-level cognitive challenge, but it is not particularly difficult and was not intended to be significantly stressful. The 1-back condition requires an additional step up in cognitive load in that the individual must both correctly recall from short-term memory the item presented previously as well as entering and holding the new item in memory. It was expected that the 1-back would have moderate impact on individuals. The 2-back form of the task requires highest cognitive load to recall from short-term memory within the n-back tasks.

### *3.2.2 Input features*

Two driving performance measures, standard deviation of lane position (SDLP) and steering wheel reversal rate (SRR), and two physiological data, Heart Rate (HR) and Skin Conductance Level (SCL), were considered as input features to estimate the levels of driver's cognitive workload in the RBPNN models.

SDLP was calculated from 0.1 Hz high pass filtered lateral position data with removing lane changes using the AIDE project guidelines. SRR was calculated by counting the number of steering wheel reversal from the 2Hz low pass filtered steering wheel angle data per minute. For cognitive workload, the reversal angles, which have more than 0.1 degree of the gap size, were counted.

HR was converted from Inter-beat Interval (IBI) which was calculated after removing irregular distance between peaks, irregular peak form, and presence of low-frequency component in ECG using the Librow`s R-peaks detection algorithm (Librow$^{TM}$, Ukraine). SCL was measured with a constant current configuration and non-polarizing, low-impedance gold-plated electrodes. Sensors were placed on the underside of the outer flange of the middle fingers of the non-dominant hand without gel.

### *3.2.3 Summarizing parameters*

In this paper, window size was considered as the summarizing parameter for the inputs. Window size denotes the period over which performance and physiological data were averaged. The comparisons of window size could identify the appropriate length of data that can be summarized to reduce the noise of the input data without losing useful information. This paper considered three window sizes: 10, 20 and 30 seconds.

### *3.2.4 Model training and testing*

Radial basis probabilistic neural networks (RBPNN) were used to construct the driver's cognitive workload estimation models. In this paper, the models were trained using the NEWPNN function in MATLAB. For training and testing RBPNN models, data of four task periods, which consist of a single task (driving only condition) and three dual tasks (n-back task condition), were used. A task was divided into multiple segments based on window size. For example, if the model uses 30s window, one task period divided into four segments as shown in Figure 1. In the same manner, 20s window set has six segments and 10s window set has twelve. In each task, half of the segments, i.e. two segments per subject in 30s window, were used for training and the other segments were used for testing. Thus, each neural net was trained and tested using different sets of measurements, i.e. 15x2, 15x3 and 15x6 examples for 30s, 20s and 10s window, respectively. Since the estimator is always evaluated on the data disjoint from the training data, the performance evaluated through the cross validation scheme correctly reflects the actual generalization capability of the derived estimator [6]. Model performance was evaluated with testing accuracy, which is the ratio of the number of instances correctly identified by the model to the total number of instances in the testing set.

## 4. RESULT AND DISCUSSION

The performance of the RBPNN models varies from the combined input features and window sizes. Among different combinations of inputs, i.e. SDLP, SRR, HR and SCL, the performance using SCL only and SCL and SDLP outperformed as shown in Table 1.

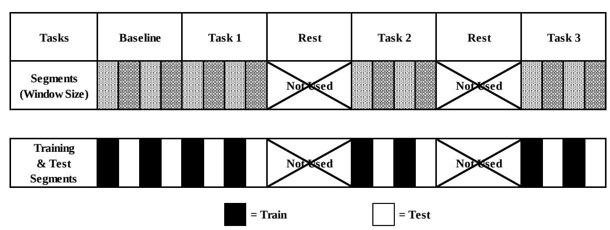


**Figure 1. Allocation of Segments to Training and Testing Sets**

**Table 1. Model performance with different window size**

| | | All | Driving Performance | Physiology | | Driving & Physiology (Combination) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | HR SCL | SCL | SDLP SCL | SDLP HR | SRR SCL | SRR HR |
| 10s | Baseline | 55.6 | 54.4 | 65.6 | **94.4** | **86.7** | 32.2 | 71.1 | 32.2 |
| | 0-Back | 44.4 | 17.8 | 52.2 | **74.4** | **70.0** | 32.2 | 60.0 | 32.2 |
| | 1-Back | 55.6 | 50.0 | 60.0 | **84.4** | **86.7** | 34.4 | 71.1 | 34.4 |
| | 2-Back | 43.3 | 8.9 | 57.8 | **82.2** | **90.0** | 31.1 | 64.4 | 31.1 |
| | **Average** | 49.7 | 32.8 | 58.9 | **83.9** | **83.3** | 32.5 | 66.7 | 32.5 |
| 20s | Baseline | 60.0 | 64.4 | 64.4 | **93.3** | **91.1** | 26.7 | 51.1 | 26.7 |
| | 0-Back | 42.2 | 33.3 | 33.3 | **80.0** | **73.3** | 28.9 | 57.8 | 28.9 |
| | 1-Back | 35.6 | 11.1 | 57.8 | **86.7** | **86.7** | 20.0 | 37.8 | 20.0 |
| | 2-Back | 37.8 | 24.4 | 53.3 | **82.2** | **91.1** | 33.3 | 46.7 | 33.3 |
| | **Average** | 43.9 | 33.3 | 52.2 | **85.6** | **85.6** | 27.2 | 48.3 | 27.2 |
| 30s | Baseline | 66.7 | 76.7 | 63.3 | **90.0** | **90.0** | 33.3 | 63.3 | 50.0 |
| | 0-Back | 30.0 | 20.0 | 50.0 | **80.0** | **70.0** | 33.3 | 50.0 | 30.0 |
| | 1-Back | 40.0 | 26.7 | 53.3 | **83.3** | **86.7** | 36.7 | 33.3 | 30.0 |
| | 2-Back | 20.0 | 36.7 | 46.7 | **86.7** | **90.0** | 33.3 | 56.7 | 33.3 |
| | **Average** | 39.2 | 40.0 | 53.3 | **85.0** | **84.2** | 34.2 | 50.8 | 35.8 |

Due to the fact that skin conductance clearly changed in physiological arousal associated with the levels of cognitive load complexity, the best performance appeared when the models have SCL as an input feature. Although SCL model and SCL-SDLP model have same performance on average, SCL-SDLP model outperforms classifying the highest cognitive workload which must be detected correctly. The best performing model, which uses SDLP and SCL data over a 20s-window, could identify four graded levels of cognitive workload with an average accuracy of 85.6%. With this model, the estimation accuracy rate of driving only criteria, i.e. no cognitive workload condition, was 91.1%, and under cognitive workload criteria the accuracy of the lowest, moderate, and the most difficult cognitive load estimation were 73.3%, 86.7%, and 91.1%, respectively.

The results demonstrated that the model using SDLP and SCL was outperforming than the other combinations among performance and physiological measures. The main contributor of the high accuracy rate in this model was skin conductance level, which provides clear changes associated with difficult level of cognitive workload, but relatively lower threshold to distinguish higher mental workload. According to Mehler et al., the additional increases in skin conductance between the 1-back and 2-back were minimal and not statistically significant. The near flattening of the response curve for all physiological measures during the 1-back and 2-back tasks may indicate that a threshold had been reached relative to the amount of additional effort that participants were willing or able to invest in the combined demands of driving and the secondary cognitive task [6]. Thus, SCL and SDLP based model provides better performance to identify higher levels of mental demand than SCL based model.

## 5. CONCLUSION

In this paper, we proposed an algorithm for estimating driver's cognitive workload using driving performance and physiological data. Especially, SDLP and SRR, and HR and SCL were considered as cognitive load indices for the driving performance and physiological, respectively. In order to collect driving data, participants drove through highway in a driving simulator and were asked to complete three different levels of auditory recall tasks. The driver's cognitive workload estimation algorithm was developed using RBPNN models that were implemented by MATLAB NEWPNN function.

The results show that the proposed SCL-based or SCL and SDLP-based RBPNN models were able to identify driver's cognitive workload complexity with high accuracy. The model performance was assessed with the cross-validation scheme, which is widely adopted by the machine learning community. As a result, the highest workload estimation accuracy rate in overall model performance was 85.6%. And it is also expected that the accuracy can be improved by applying more sophisticated algorithms.

## 6. ACKNOWLEDGMENTS

This research was supported in part by Daegu Gyeongbuk Institute of Science and Technology (DGIST) Research Program of the Ministry of Education, Science, and Technology (MEST), and Establishment Program of Industrial Original Technological Base of the Ministry of Knowledge Economy (MKE).

## 7. REFERENCES

[1] Son, J., Lee, Y., and Kim, M. 2011. Impact of traffic environment and cognitive workload on older drivers' behavior in simulated driving. *International Journal of Precision Engineering and Manufacturing.* 12, 1 (Feb. 2011), 135-141.

[2] Angell, L., Auflick, J., Auustria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., and Kiger, S. 2006. *Driver Workload Metrics Task2 Final Report*. DOT HS 810 635. National Highway Traffic Safety Administration.

[3] Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. 2009. The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record*. 2138, (Mar. 2010), 6-12. DOI= http://dx.doi.org/10.3141/2138-02

[4] Son, J., Mehler, B., Lee, T., Park, Y., Coughlim, J., and Reimer, B. 2011. Impact of cognitive workload on physiological arousal and performance in younger and older drivers. In *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment* (Tahoe Lake, CA, USA, June 27-30, 2011). DA2011, Univ. of Iowa, Iowa City, IA, 87-94.

[5] Reimer, B. 2009. Impact of Cognitive Task Complexity on Drivers' Visual Tunneling. *Journal of the Transportation Research Board of the National Academies*. 2138, (Mar. 2010), 13-19. DOI= http://dx.doi.org//10.3141/2138-03

[6] Zhang, Y., Owechko, Y., and Zhang, J. 2004. Driver cognitive workload estimation: A data-driven perspective. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems* (Washington D.C., USA, October 03-06, 2004). ITSC2004. IEEE, New York, NY, 642-647.

[7] Son, J. and Park, S.W. 2011. Cognitive Workload Estimation through Lateral Driving Performance. In *Proceedings of the 16th Asia Pacific Automotive Engineering Conference* (Chennai, India, October 06-08, 2011). APAC16. SAEINDIA, India, SAE2011-28-0039.

[8] Engström, J., Johansson, E., and Östlund, J. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour.* 8, 2. (Mar. 2005), 97-120.

[9] Brookhuis, K. A. and D. De Waard. 2001. Assessment of Drivers' Workload: Performance and Subjective and Physiological Indexes. In *Stress, Workload, and Fatigue*, P. A. Hancock and P. A. Desmond, Ed. Lawrence Erlbaum Associates, Mahwah, NJ, 321–333.

[10] Demuth, H. and Beale, M. 2000, *Neural Network Toolbox User's Guide*, MathWorks, Inc., Natick, MA.

[11] Folstein, M.F., Folstein, S.E., and McHugh, P.R. 1975. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 3 (Nov. 1975) 189-198.

[12] Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K.M., Orszulak, J., Long, J., Kim, M.H., Lee, Y.T. and Coughlin, J. F. 2010. Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving. *International Journal of Automotive Technology*, 11, 4 (Aug. 2010) 533-539.