Real-time Driver Cognitive Strain Evaluation System Based on Multivariate Biosignal Analysis

Takeshi Shimizu (shimizu-takeshi-gx@ynu.jp)  
Yokohama National University

Keisuke Shima  
Yokohama National University

Takayuki Mukaeda  
Yokohama National University

Shu Muraji  
Mazda Motor Corporation (Japan)

Juntaro Matsuo  
Mazda Motor Corporation (Japan)

Masayoshi Horiue  
Mazda Motor Corporation (Japan)

Article

Keywords:

Posted Date: November 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2183799/v1

License: ☺ ☁ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Additional Declarations: No competing interests reported.
Real-time Driver Cognitive Strain Evaluation System Based on Multivariate Biosignal Analysis

Takeshi Shimizu¹*, Keisuke Shima¹*, Takayuki Mukaeda¹*, Shu Muraji²*, Juntaro Matsuo²* and Masayoshi Horiue²*

¹Graduate School of Engineering Science, Yokohama National University, Japan
²Mazda Motor Corporation, Japan

* Corresponding author: Takeshi Shimizu (shimizu-takeshi-gx@ynu.jp).

Abstract

Traffic accidents are often attributed to reduced attentional capacity, as vehicle operations involve complex tasks that depend on the driver’s mental workload (MWL). Cognitive strain in MWL affects focus on other vehicles and pedestrians, which can result in serious accidents. Therefore, a method for effective cognitive-strain evaluation is required.

In this study, the driver cognitive strain was determined using the a posteriori probability method. The posteriori probability of the driver strain state is computed using multidimensional biological signals mapped into a probabilistic space based on a Gaussian mixture model. Ultimately, the driver’s cognitive strain was estimated using a regression function based on the posteriori probability. In the experiments, tasks tracked the preceding vehicle, and N-back evaluation helped elucidate the cognitive strain. Other experiments have shown an increase in the evaluation value during the satnav operation. Accordingly, the proposed method is considered suitable for real-time driver cognitive strain evaluation.

The annual incidence of traffic accidents in Japan remains relatively high (approximately 380,000 in 2020) despite a gradual decline. Approximately 56% of accidents are attributed to a lack of attention to safe driving and other forms of neglect, with satnav systems and other types of in-vehicle technology contributing to drivers’ cognitive strain.

A driver requires appropriate handling while taking attentional resources to various objects such as traffic lights and other cars [1], [2]. In today’s driving environment, driver cognitive strain associated with in-car technology use can be caused by serious traffic accidents. The variety of high-performance user interfaces (UIs) provided for convenience also increases this strain, creating a need for designs that consider the timing of information provision and UI placement. Against this background, the technique for allowing driver cognitive strain evaluation is anticipated to assist in the design of UI and reduce accidents.

In previous research, driver attention has been defined in terms of physical awareness (perception, assimilation, and processing of information), cognition, and appropriate responses (movement). Perception resources can be evaluated based on fluctuations in both eyes and time spent looking aside. Movement resources can be evaluated from the three-dimensional motion analysis and muscle potentials. However, the mental workload (MWL) used to evaluate cognitive resources was a subjective evaluation index. Quantitative evaluation of cognitive resources is difficult because superficial results are not always fully clarified.

Extensive research conducted to evaluate driver MWL based on attentional resources has generally involved questionnaire-based subjective ratings [3]–[6] or physiological evaluations [7]–[24]. However, studies based on physiological assessment are often limited to the classification of the participant’s MWL condition, and no method has been proposed that allows the real-time evaluation of MWL.

The authors previously reported a real-time MWL evaluation using multiple biological signals and probability neural network usage [25]. However, the analysis and discussion of cognitive loads has been insufficient because of the limited number of contributors involved.
This study outlines a method for real-time driver cognitive strain evaluation that considers the characteristics observed in various bio-signals and operating information (such as sightline information and steering angle) based on a probabilistic evaluation model to determine the impact on operation. The cognitive strain evaluation approach was examined using a larger number of participants and improved experimental protocols, as previously presented by IEEE SMC2020[25].

**Related research**

**THE NASA-TLX SUBJECTIVE RATING SCALE.** The NASA-TLX is widely used to evaluate MWL. This method calculates MWL ratings by multiplying the evaluation values of six indices (MD: mental demand, PD: physical demand, TD: temporal demand, PE: performance, EF: effort, and FL: frustration level) and their weights [5]. The weighting values were determined using the pairwise comparison method, and the numerical values of the six indices were measured using a visual analog scale. Finally, the evaluation value is calculated by multiplying the results obtained using each method [3].

\[
WWL = \sum \left( x_i \times w_i \right) / \sum w_i
\]

Here, \( x_i \) and \( w_i \) are the numerical values and weights for each scale, respectively.

While NASA-TLX supports the overall evaluation of MWL experimental conditions, completion of the related questionnaire places a significant burden of time and other stress variables on subjects. MWL determination during the tests is impractical, and the results may vary temporally.

**PHYSIOLOGICAL EVALUATION INDEX.** The main indices in the physiological evaluation method for MWL are the cardiovascular system (such as wrist pulse wave and electrocardiogram) and electroencephalogram, which are controlled by the autonomic nervous system. These indices are appropriate for this purpose because they are intimately related to fluctuations in sympathetic and parasympathetic nervous system activities. Increased heart rate (HR) index and decreased heart rate variability (HRV) index with increasing MWL have previously been reported. Although HR and HRV exhibit high sensitivity to low MWL levels, such fluctuations are insensitive to high MWL values [8][9]. The MWL condition of a driver can be classified with high accuracy using these biosignals [10]–[23]. However, these studies cannot consider real-time estimation and evaluation of the driver using simple sensor methods to evaluate the MWL. In MWL evaluation, these features are normalized and input directly as explanatory variables in multiple regression analysis and mathematical modelling. Previous research has shown high accuracy using this approach, with a correlation coefficient close to \( R = 0.8 \) [24]. However, real-time evaluation tends to be less accurate because of the difficulty of using MWL values for teacher data monitoring during experiments.

**Evaluation system [25]**

Mathematical modeling often requires MWL-based signal analysis of biologically obtained data. The application of various types of data results in a significant increase in the dimensions of related variables, potentially causing dimensional impairment and lower evaluation accuracy [26], [27]. As a result, there is a need for a technique that allows regression and evaluation to accurately determine the characteristics in a multidimensional context. Outliers, noise, and other stochastic data in driver features are ignored in regression-related modeling with factors, such as direct variables, indicating that content can significantly impair evaluation accuracy.

The authors’ approach is presented in Fig. 1 along with the monitoring and analysis of a range of driver-related biological factors. This approach involves the determination and dimensional reduction of related properties, along with the stochastic establishment of driver strain states via driver utilization. Such discrimination supports the expression of operational conditions from an a posteriori probability, thereby supporting analytical accuracy. The application of the resulting data to the explanatory variable usage promotes resistance to outliers/noise and cognitive strain evaluation in real-time.
Analysis of biological data and determination of related characteristics. The variables $x_m(t) \in \mathbb{R}^m (m = 1, \ldots, M, L \, t \, m \in \mathbb{R} x_0 (m = 1, \ldots, M)$, $x_{mL} \, t$ representing biological signal dimensions, and $t$ representing the temporal variable) were used to analyze the biological signals from the driver, and the $n$-th-order differential value $x_{mL} \, t$ was determined via differential filtering $(1, \ldots, n \in \mathbb{N})$. Standard deviation $SD_x$ and other statistical variables are established in this manner. The element $x_{mL} \, t$ associated with the average vector of $x_{mL} \, t$ at time $t$ is expressed as

$$x_{mL}^{AVG}(t) = \frac{1}{T} \int_{t-T}^{t} x_{mL}(t) \, dt$$

Here, $T$ is the length of the mean vector based on the calculation, and $x_{mL}(t - T) = 0$ for $t < T$. With $[t - T', t]$, a frequency analysis was conducted for $x_{mL}(t)$, and the power spectrum density $P_{mL}(f)$ was determined to establish integrals in $k$ frequency sections $f_s$, the ratio of total energy, section-specific ratios, and other variables. The characteristic vector term, $x(t) \in \mathbb{R}^m$, was applied as a collective reference to these terms.

Analysis of operating status. Multidimensional characteristic vectors are used to determine the operating situation of the driver based on stochastic neural network considerations. As the curse of dimensionality may impair the precision of evaluation with high-dimensional vector input, $x(t)$ a characteristic vector mapping is applied to the low-dimensional characteristic space $V$ involved in driver evaluation prior to defining the characteristic vector $x'(t) \in \mathbb{R}^D$ with $D' < D$. A Bayesian discrimination probability neural network was applied to determine the $x'(t)$ as a post-dimensional compression characteristic vector. Prior learning of load-operation characteristic vectors supports the stochastic evaluation of the driver’s strain state from an a posteriori probability $P(c \mid x'(t))$, with $c$ being the driver operation status.

The cognitive strain evaluation model. The relevant operational status was referenced to evaluate and establish cognitive strain values (e.g., useful field of view, NASA-TLX). For the a posteriori probability $P(c \mid x'(t))$ of the operation status determined from the stochastic neural network, the regression function was referenced, and the function parameter setting was performed based on the least-squares approach.

$$y = \phi(P(1 \mid x'(t)), \ldots, P(C \mid x'(t)))^T \mid \theta$$

where $y$ represents the cognitive strain evaluation, $\phi(\cdot)$ represents the regression function, and $\theta$ represents the parameters of the regression function. Thus, multivariate biosignals can be used for the real-time evaluation of cognitive strain.

Ethics statement and consent. This research was performed in accordance with relevant guidelines and regulations and approved by the local ethics committee of Mazda Motor Corporation (letter no: VDD-154-04). Informed consent was obtained from all participants.
Experiments

The proposed method, published in IEEE SMC2020 [25] had difficulty estimating the cognitive load of a task and required a more detailed discussion with a larger number of participants. Therefore, the method proposed in the previous study was improved by focusing on driver cognitive strain. In addition, these experiments were improved by changing the N-back task from images to sound and increasing the number of participants. The driving simulation system in the experiment consisted of pedals and a steering wheel (LPRC-15000, Logicool), driving seat (Playseat), and Carsim (mechanical simulation), which was used as a driving simulator. The ECG was monitored using a telemeter picker (ZB-151H), discharge electrode (L-150), and multichannel telemeter (WEB-7000) of the NIHON KOHDEN CORPORATION. ViewTracker II (Ditect) was used to obtain the angle of gaze and pupil diameter of both the eyes.

Method. In the experiment, six participants aged 21-24 perform six conditions over two days. To induce changes in cognitive strain while keeping perception and movement resources relatively constant, a car-tracking scenario was chosen for the primary task. Because the experiment assumed a highway, this scenario followed a car at 80 km/h on an endless straight road. The N-back task is a cognitive task for memorizing numbers and is widely used as a subtask during driving. In the experiment, the participants memorized a number heard randomly from a speaker and answered whether the number was the same as the number n ago. To equalize the cognitive strain effects, the number of correct answers for the N-back task was the same in all experiments. The N-back task selected for the cognitive task involved numbers from zero to nine. In addition, the experiment involved driving without a subtask (normal) and driving with subtasks (zero–three back). After the experiment, the participants answered the NASA-TLX questionnaire. Here, numbers were displayed at two-second intervals in the N-back task. A break of approximately 10 min was allowed between each experiment because mental fatigue from the previous experiment was assumed to affect the next experiment.

Before the experiment, all participants learned how to operate the driving simulator and were given 30 min of training on each task. The experimental setup is shown in Fig. 2. Participants completed the N-back task while striving to maintain a constant distance from the vehicle preceding the driving simulator.

Formulation of the cognitive stain evaluation system. 1) Biosignal monitoring and characteristic extraction. In the cognitive strain evaluation system, types of various information such as ECG signals, line-of-sight information, and steering/pedal operation information are obtained by measuring equipment and simulators. The characteristics were extracted from the ECG RR interval, horizontal gaze angle (Deg $X$), vertical gaze angle (Deg $Y$), right eye pupil diameter (Dia $r$), left-eye pupil diameter (Dia $l$), steering angle (SA), pedal pushing depth (DDA), steering angular velocity (SAV), and temporal variations in pedal pushing depth (DDA$_v$). The extracted characteristics were mean ($^{m}x_{n}^{AVG}$) and standard deviation ($^{s}x_{n}^{STD}$) and integrated value of the power spectrum density ($^{s}F_{n}(f)$) was determined to establish integrals in the $k$ frequency sections $f_{k}$ for each frequency band. The power spectrum was calculated by performing Fast Fourier Transform (FFT) on the resampled data after applying a cubic interpolation of each data point. Moreover, the root mean square ($^{n}x_{n}^{RMS}$) was extracted as a characteristic of the operation information and line of sight.

Each characteristic was calculated using $T$ seconds of information and the output of $s$ seconds interval. These characteristics were used to determine the characteristic vector, $x(t) \in \mathbb{R}^{D}$. The experiment yielded seven sets of data extracted from ECG signals, 24 from line-of-sight information, and 25 from operational information. A total of 56 dimensional ($D = 56$) characteristic vectors $x(t) \in \mathbb{R}^{56}$ were generated by arranging these data.

The values for the time width for obtaining the characteristic value and output interval were 60 s ($T = 60$) and one second ($s = 1$), respectively. Therefore, the number of characteristic vectors per experiment was 540 ($N = 540$).

In addition, the integrated power spectrum density as an ECG characteristic was calculated from three frequency intervals ($k = 3$) for low frequency (0.04-0.15 Hz), middle frequency (0.078-0.137 Hz), and high frequency (0.15-0.4 Hz) bands. In terms of visual and operational information, the three frequency sections were low frequency (0-10 Hz), middle frequency (10-20 Hz), and high frequency (20-50 Hz) bands.

2) Dimensionally compressed probabilistic neural network classification. The hidden Markov model (HMM) is a probabilistic application that is widely used to classify time-series data. This model consists of the internal probabilities of transition based on the Markov process, and the output probability for each state. In particular, a continuous density HMM, which supposes a continuous function in the output probability density function of the HMM, is mainly used. However, it is difficult to identify high-dimensional time-series data using CD-HMM because of the increased number of parameters in Gaussian mixture model (GMM). The estimation of cognitive strain requires a high-dimensional
characteristic vector \( \mathbf{x}(t) \in \mathbb{R}^D \), which is extracted from biological signals and operational information. Therefore, classification using CD-HMM requires a dimensionality reduction.

Hence, time-series discriminant component analysis (TSDCA) [28] was adopted for the identification model. Owing to the assumption of GMM in a characteristic space after dimensional reduction, TSDCA permits the acquisition of the GMM parameter by learning. Therefore, this model allows for the estimation of the strain states of the driver.

3) The cognitive strain evaluation. Cognitive strain evaluation values were calculated using a multiple regression model. In this model, the MWL values were adopted as objective variables, and the a posteriori probability of each driver's strain state was used as an explanatory variable. The characteristic vector of the state and the power function shown in the following equation are used in the model formation:

\[
y = \mathbf{a} \left[ P(1 \mid \mathbf{x}'(t))^\theta, \ldots, P(C \mid \mathbf{x}'(t))^\theta \right]^T
\]  

In the experiment, the characteristic vector \( \mathbf{x}(t) \in \mathbb{R}^{26} \) was produced from multivariate biosignals and operation information to determine the driver's strain state. The driver cognitive strain was evaluated using the NASAT-TLX and a regression model after learning in a posteriori probability. Here, the parameters \( \theta \) and \( \mathbf{a} = \mathbf{b} \) of the regression model were designed using the evaluation values of NASA-TLX and a posteriori probability of the driver's cognitive strain state.

In the evaluation, the training data \( c = 2 \) were adapted to the experimental results from the zero and three-back conditions. The test data were adapted for new data that were not used for learning in the regression model, and the error with the NASA-TLX ratings was computed.

**Results and discussion.** Fig. 3 shows the average error rate determined from the NASA-TLX and each subtask for all participants. It can be observed that the NASA-TLX ratings (Fig. 3 (a)) and the average incorrect answer rate (Fig. 3 (b)) increased with N-back difficulty. In the experiment, an increase in NASA-TLX ratings indicated an increase in the cognitive strain of participants because variations other than cognitive resources were quite small. Therefore, an increase in MWL indicates an increase in the cognitive strain of participants, and the proposed method can estimate cognitive strain.

Fig. 4 shows the correlation and errors between the NASA-TLX and the evaluation results for each experimental condition. The correlation coefficient obtained was 0.705 (\( R = 0.705 \)), indicating a strong positive correlation between the NASA-TLX ratings and evaluation values. The error rate (ER) for the evaluation and monitoring values was computed as:

![Figure 2. Experimental set-up](image-url)
The error rates were 9.37% for the learning data and 22.29% for the test data. Based on these results, the evaluation values for the proposed method were allowed to reproduce the monitoring values to a certain extent. In the results of the two-back high-load test, the average error rate was as low as 10.10%. The high-load condition was more accurate for estimation in the proposed method.

The discussion of the experimental results is affirmed. From the results of the experiment (Fig. 3(a)), the driver’s cognitive strain increased with N-back difficulty. This implies that the NASA-TLX ratings highly reflect the mental state of the driver, and the driver’s cognitive strain varies according to the experiments. The results of the evaluation of the driver cognitive strain using the proposed method were highly correlated with the NASA-TLX ratings, and the evaluation accuracy was particularly high when the cognitive task load was heavy. This trend was obtained because the participants concentrated on the high load task and the characteristic biological signals were measured stably.

Most conventional methods attempt to classify the MWL condition and do not consider estimating the MWL with simple sensors in real time. In contrast, the proposed method enables the estimation and calculation of MWL values using multiple regression analysis based on posterior probability from probabilistic neural networks. Therefore, it is difficult to determine the effectiveness of the proposed method by simply comparing it with the classification accuracy of previous studies. However, a similar trend was demonstrated in previous studies [18], [19]. Therefore, the proposed method can accurately detect attentional resource exhaustion and is expected to prevent accidents caused by inattention.

However, the estimation accuracy decreased in the low and no-load tasks. In the high-load cognitive task cases, the variations in each resource were considered small because the drivers were required to focus on the subtasks. On the other hand, in the execution of relatively simple cognitive tasks and normal operations, resources are more likely to fluctuate because of the surplus of resources. The NASA-TLX rating was one for each experiment, and validation during
the experiment could not be evaluated. In addition, noise, such as gaze deviation, easily occur when a driver becomes distracted [8], [19]. Therefore, the estimation accuracy of the proposed model decreases. Therefore, measures, such as monitoring the MWL during the experiment and noise suppression, are necessary.

**Experiments using actual equipment**

In these experiments, driving performance and the driver cognitive strain during the operation of a satnav system in an actual car were evaluated.

**Method.** In the experiment, six participants aged 21-24 perform six conditions over two days. The participants operated voice recognition, dial-type input device (commander) operation and satnav touch panel operation under the assumption of the vehicle in motion. The operational tasks were set search tasks for a complex artist-search task involving scrolling the display and several commands and a comparatively simple peripheral facility search involving two or three commands from the top screen. The display above the satnav system shows search targets.

Fig. 5 outlines the experiment in which the subjects drove behind another vehicle (Section IV) while performing in-car tasks. The MWL values of the participants were estimated for use in a related evaluation system (Section IV). Participants completed the NASA-TLX questionnaire, and their cognitive strain was analyzed by considering resource fluctuations.

**Results and discussion.** Fig. 6 shows examples of the results from the real-time driver cognitive strain evaluation, with shaded areas expressing periods of equipment operation. The higher evaluated value during the surgery confirmed the effectiveness of this technique. Fig. 7 shows that the average distance from the other vehicle increased overall during the more demanding artist search. Most data show shorter distances for the peripheral facility search than for the artist search, with the touch operation being the highest. Operational task comparison highlights the difficulty of the artist search, which results in a higher MWL for the driver and poorer driving performance. Results relating to equipment operation also underlined that the three resources were the most consumed by touch-panel operation. Previous studies have shown results of increase in MWL and decreases in driving performance due to equipment operation [15] [20] – [21]. In Fig. 6(a), the estimated value increased before device operation, indicating that the participant had time between checking the display and operating the equipment. However, the results of driving performance alone cannot be used to evaluate driver cognitive strain because all three resources exert an influence. Accordingly, cognitive strain needs to be discussed further in the evaluation of the proposed method.

Fig. 8 shows the average monitoring values from NASA-TLX and the evaluation values for the proposed method. The mean error of all evaluations was 33.03 % for peripheral facility retrieval. The figure without subject A was 33.7 % for artist retrieval. This value is greater than the error rate of the test data in the experiment described in Section IV. Here, data for Subject C in the peripheral facility search tasks were excluded owing to equipment malfunction.

The results for Subject A in the artist search task exhibited significant differences between monitoring and evaluation owing to the subjective evaluation method of the NASA-TLX. As the experiment lasted for two days, the subject evaluation criteria differed each day. Accordingly, in contrast to the driving performance results, the monitoring values for the peripheral facility search task were higher than those of the artificial search task. Most evaluation values showed a trend similar to that of driving performance, with touch panel operation data being the highest, and peripheral facility search data being the lowest. Thus, the proposed method can be considered appropriate for accurate MWL evaluation. An evaluation system formulated for other tasks using this approach could be considered suitable for evaluating different tasks.

In relation to cognitive strain, based on the above, the experiment described in Section IV was designed for minimal variation in non-cognitive resources. Thus, increased evaluation values represent higher cognitive strain, which was also observed in this experiment. Most of the results showed a similar trend for peripheral facility and artist searches, indicating that the proposed method allows for the evaluation of cognitive resources alone, with no significant influence from other resources. However, the estimation accuracy of this approach is low when the level of cognitive strain differs significantly between the training data and evaluation targets. The estimated values for Subject E showed no difference between the operating and task conditions. As satnav operation was more challenging than N-back tasks for this subject, estimation values were at the upper limit for each condition. This was also observed for Subject F. Thus, the evaluation system required the standardization of difficult tasks and the absence of an upper limit.
Figure 5. Flow of the experiment

Figure 6. Examples of results for real-time evaluation (Participant B)

Figure 7. Driving performance results
Conclusion

In this study, a customized method based on a previously proposed method in IEEE SMC2020 [25] was investigated to enable the real-time evaluation of driver cognitive strain. First, an experiment was conducted using a driving simulator with six participants to demonstrate the effectiveness of the proposed technique. In the driver strain cognitive evaluation experiment, cognitive resource fluctuations were monitored and evaluated using the operational task performance and four cognitive tasks. The cognitive strain of the participants was successfully evaluated using the monitored operational information and multivariate biosignals. A high correlation coefficient was achieved between the evaluation values and the NASA-TLX ratings, indicating the validity of the proposed method for driver cognitive strain evaluation.

Second, the participants performed driving and retrieval tasks using actual car equipment, and variability in cognitive resources was monitored and evaluated. The results demonstrate the validity of the proposed method for driver cognitive strain evaluation to enable real-time and real-world environments.

In future work, the accuracy of the evaluation system should be enhanced by analyzing the contribution ratio between each characteristic and the driver’s cognitive strain, along with the development of an evaluation model that integrates the regression function and a probabilistic neural net.

Reference

2. Wickens, C. D., Sandry, D. L., & Vidulich, M., Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, 25, pp. 227-248

Figure 8. Comparison of monitoring and estimated values
7. Soichiro, A., Haruhiko, S., Toshio, I., Kenta, Y., Teruo, Y., Yusuke, T., & Tadao, K., A study on controlling method of providing useful information to driver based on estimating driving workload considering individual differences. Transactions of Society of Automotive Engineers of Japan, pp. 261-266, 2017
23. Le, A. S., Aoki, H., Murase, F. & Ishida, K., A Novel Method for Classifying Driver Mental Workload Under Naturalistic Conditions With Information From Near-Infrared Spectroscopy. Frontiers in Human Neuroscience, Brain Imaging and Stimulation, 26 October 2018

Data availability
The datasets used and analyzed during the current study available from the corresponding author on reasonable request.

Acknowledgements
The authors received no specific funding for this work.

Author contributions
T.S. devised the protocol and performed the data acquisition, analysis, interpretation, drafting, and critical revision of the manuscript. K.S. and T.M. contributed to improve of the protocol and critical revision of the manuscript. S.M., J.M. and M.H. provided advising and the experiments revision. All authors reviewed the manuscript.