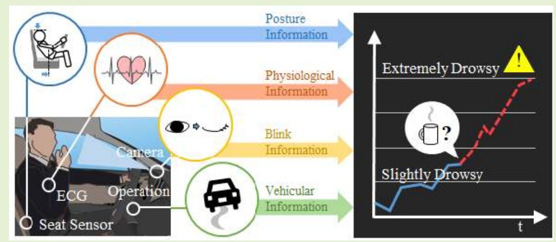
Comprehensive Drowsiness Level Detection Model Combining Multimodal Information

Mika Sunagawa, Shin-ichi Shikii, Wataru Nakai, Makoto Mochizuki, Koichi Kusukame, and Hiroki Kitajima

***Abstract*—This paper presents a drowsiness detection model that is capable of sensing the entire range of stages of drowsiness, from weak to strong. The key assumption underlying our approach is that the sitting posture-related index can indicate weak drowsiness that drivers themselves do not notice. We first determined the sensitivity of the posture index and conventional indices for the stages of drowsiness. Then, we designed a drowsiness detection model combining several indices sensitive to weak drowsiness and to strong drowsiness, to cover all drowsiness stages. Subsequently, the model was trained and evaluated on a dataset comprised of data collected from approximately 50 drivers in simulated driving experiments. The results indicated that posture information improved the accuracy of weak drowsiness detection, and our proposed model using the driver's blink and posture information covered all stages of drowsiness (F1-score 53.6%, root mean square error 0.620). Future applications of this model include not only warning systems for dangerously drowsy drivers but also systems which can take action before their drivers become drowsy. Since measuring the information requires no restrictive equipment such as on-body electrodes, the model presented here based on blink and posture information can be used in several practical applications.**



***Index Terms*— Driver fatigue, driving performance, drowsiness detection, multi-modal sensing, slight drowsiness.**

I. Introduction

DRIVER drowsiness can cause serious accidents. According to the National Highway Traffic Safety Administration, between 2011 and 2015 approximately 30,000 accidents per year have been due to drowsy driving resulting in approximately 800 fatalities [1]. As indicated by the European New Car Assessment Programme's 2020 assessments of automotive safety, which considers driver-state monitoring systems as a factor in assigning safety ratings, driving safety support systems require further improvement to prevent accidents caused by this type of human error [2].

To date, numerous driver drowsiness detection systems have been proposed [3]–[5]. These systems are comprised of three main components: a drowsiness evaluation scale, direct measurement indices, and a classification method. To estimate a driver's state, the classification method is applied to the directly measured indices and the driver's rating on the evaluation scale is established.

Drowsiness evaluation scales are defined in several ways [6]–[8]. The Karolinska Sleepiness Scale (KSS) [6] is one of the most widely used scales [9], [10]. It is a subjective evaluation, in which drivers report their perceived arousal on a scale from one to nine. Wierwille and Ellsworth [7] and Kitajima *et al.* [8] have also proposed alternative evaluation scales. Kitajima et al. defined drowsiness as an interval scale consisting of five levels, with "1" for not drowsy, "5" for extremely drowsy, and evenly spaced levels of "2," "3," and "4" between "1" and "5" (hereinafter termed "the drowsiness level"). Table I shows the corresponding features of these levels. The drowsiness level is also widely used, because the level determined by trained observers based on the driver's facial expressions shows a high correlation with the driver's self-reported drowsiness level and can thus be used as a substitute for subjective reports [11], [12].

Directly measured indices are typically classified into three groups: driving information, driver behavioral information and physiological information. Driving information includes lane departures and steering wheel movement [13]–[15]. Behavioral information is one or more visible signals that can be measured using a camera or other unobtrusive equipment, and includes information such as eye state, facial features and head movement [16]–[18]. Physiological information, such as heart rate variability and electroencephalography signals, is usually

Manuscript received November 19, 2019; accepted December 12, 2019. Date of publication December 16, 2019; date of current version March 5, 2020. The associate editor coordinating the review of this article and approving it for publication was Dr. Cheng-Ta Chiang. *(Corresponding author: Mika Sunagawa.)*

M. Sunagawa, S.-i. Shikii, W. Nakai, M. Mochizuki, and K. Kusukame are with Panasonic Corporation, Kadoma 571-8501, Japan (e-mail: sunagawa.mika@jp.panasonic.com; shikii.shin-ichi@jp.panasonic.com; nakai.wataru@jp.panasonic.com; mochizuki.makoto001@jp.panasonic.com; kusukame.koichi@jp.panasonic.com).

H. Kitajima is with the Ohara Memorial Institute for Science of Labour, Tokyo 151-0051, Japan (e-mail: h.kitajima@isl.or.jp).

Digital Object Identifier 10.1109/JSEN.2019.2960158

TABLE I
DROWSINESS LEVELS AND FEATURES [8]

Drowsiness level	Features
1: Not drowsy	Line of sight moves fast and frequently. Facial movements are active, accompanied by body movements.
2: Slightly drowsy	Line of sight moves slowly. Lips are open.
3: Moderately drowsy	Blinks are slow and frequent. There are mouth movements.
4: Significantly drowsy	There are blinks that seem conscious. Frequent yawning.
5: Extremely drowsy	Eyelids close. Head tilts forward or falls backward.

measured by direct contact with the subject [19]–[21]. Notably, especially in recent years, active development has focused on practical-use oriented systems with non-intrusive measures [22]–[25].

Machine learning methods, such as neural networks and support vector machines, have been used to classify drowsiness state based on measured indices [25]–[28]. There are two ways to train machine learning models: personalized, in which the training dataset includes the target user's data, or otherwise. Due to large individual differences in measured indices, personalized models can detect drowsiness more accurately [29], [30]. However, personalization raises the question of how to obtain training data in real-life scenarios. In order to avoid this problem and train a model to be adaptable to various drivers based only on datasets collected and labeled in advance, the model must be adjusted using subject-wise cross-validation, which evaluates the model training and evaluation data on a subject by subject basis [28].

Past studies have targeted the detection of severe drowsiness that noticeably degrades driving performance, indicated by features such as beginning to stray out of one's lane, greater variation in steering wheel movements, and erratic driving speed [12], [15], [31]. One reason for this focus on severe drowsiness is because a good index to detect weak drowsiness in various drivers has not yet been found. However, effective accident prevention requires the detection of slight drowsiness that occurs before driving performance degrades, to avoid progression to severe drowsiness.

This study aimed to detect both severe and slight drowsiness without personalization. To detect slight drowsiness, we introduced the posture-related index. As posture is maintained by the central nervous system [32], the effects of drowsiness on the nervous system are expected to appear in postural change. However, body posture has only rarely been evaluated in slight drowsiness detection. Therefore, we first investigated the sensitivity of directly measured indices, including a sitting posture-related index, for drowsiness. Then, we proposed a drowsiness detection model that combines multiple complementary information sources—specifically, indices sensitive to weak drowsiness and others to strong drowsiness—to capture all drowsiness levels. We evaluated our proposed model using a driving simulator dataset without personalization to assess the adaptability of the model to various drivers. Achieving detection of all drowsiness stages, especially slight drowsiness, enables proactive assistance that keeps drivers from falling into a dangerous state by encouraging rest or providing an alertness-inducing stimulus.

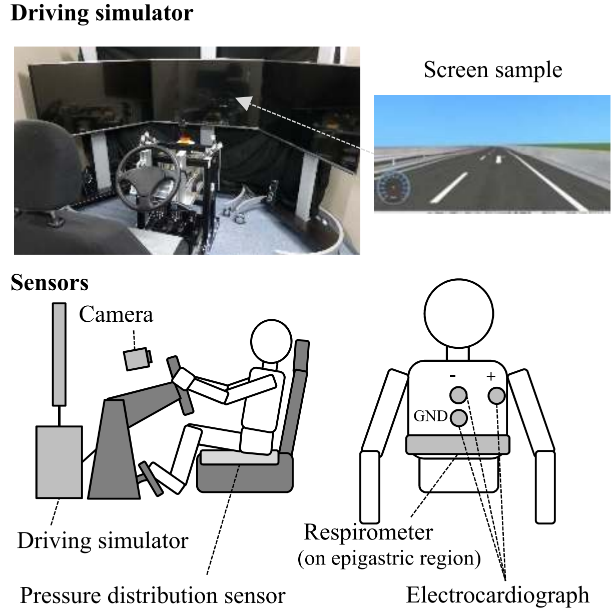


Fig. 1. Experimental environment of the driving simulator.

The remainder of this paper is organized as follows. Section II describes our data collection procedure. Section III presents our assumptions on the sensitivity of indices for the drowsiness stages and a model that is capable of detecting all drowsiness levels based on these sensitivities. Section IV details the results of the model evaluation and Section V analyzes the evaluation results. Finally, Section VI states our conclusions.

II. DROWSINESS DATA COLLECTION

The dataset was comprised of data collected from N = 49 participants (average age: 38.8 ± 5.7; 26 males and 23 females), all possessing regular driver's licenses. The participants were asked to get adequate sleep to feel rested the night before the experiment. They were not informed of the purpose of the experiment in advance, and they participated without any preconceptions of the expected results. The experiment was conducted with the approval of the in-house Ethics Review Committee and with the informed consent of the participants.

Fig. 1 shows the virtual road environment established for the driving simulator experiment, which was presented on a display in front of the subject. The course was a monotonous highway circuit that included curves. Before the main trial, the subject practiced driving for approximately 10 minutes to familiarize themselves with the experimental environment and to become relaxed. Subsequently, the subject drove along the circuit for approximately one hour between 14:00-16:00 in an air-conditioned room with temperature control set to 24 °C. The subject was instructed to follow a vehicle that was shown to be travelling at a constant speed. The subject was also instructed to not fall asleep as would be expected during regular driving.

While the subject was driving, the subject's face was recorded with a camera operating at 60 fps. The driving simulator recorded the movements of the pedals and steering wheel at 10 Hz. The subject's electrocardiogram and respiration were measured at 200 Hz (BIOPAC Systems, Inc., MP150 with BN-RSPEC). The seat pressure distribution was measured at 5 Hz (Sumitomo Riko Co., Ltd., SR Soft Vision SVZB4545L).

The drowsiness level is the ground truth in this paper. Subjective evaluations such as KSS are not suitable to monitor slight drowsiness because obtaining the KSS requires the driver to self-report, and consequently has an arousing effect. The drowsiness level, which is an external evaluation, can monitor fine variations in drowsiness without disturbing drivers. To obtain the drowsiness levels, two trained raters evaluated the facial expressions of drivers. In advance, the raters had been lectured on drowsiness level by a psychologist and trained to evaluate the drowsiness level based on facial expressions with a rating-training dataset. It was confirmed that the raters' evaluations showed high concordance (weighted Cohen's kappa [33] was over 0.9). After the driving experiment, these trained raters determined the drowsiness level of each subject every five seconds using the captured video of each subject's face. The values assigned by raters were averaged over three-minute intervals, which is the time window for signal processing to calculate drowsiness indices using acquired data, and the value thus obtained was specified as the true value of the drowsiness level.

Here, we show the correspondence between the drowsiness level and the KSS, to enable the findings in this paper based on drowsiness level to be applied to studies based on the KSS. Before and after the driving experiment, each subject was asked to self-report the KSS and the drowsiness level; the KSS report was based on the Japanese version [10]. The correlation of self-reported drowsiness levels and evaluated drowsiness levels, rated at the nearest timepoint, was 0.74. Fig. 2 shows the correspondence between the KSS values and the self-reported drowsiness levels, indicating the correspondence frequencies before and after driving by blue circles and red circles, respectively. The high correspondence between KSS 7 and drowsiness level 3 is consistent with previous results on driving performance that show steep degradation starting at KSS 7 [34], [35] and drowsiness level 3 [11], [12].

Fig. 2. Frequencies for all pairs (Drivers' self-reported drowsiness levels, KSS value). Blue circles show the frequency of drowsiness level and KSS before driving the circuit; red circles show the frequency after driving.

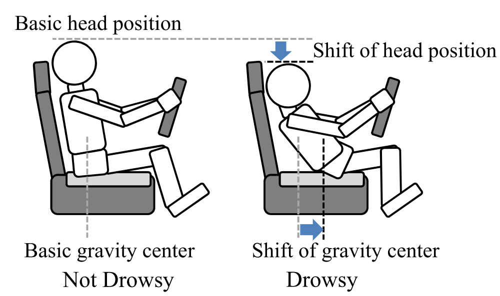


Fig. 3. An example of presumed changes in posture with drowsiness progression. The right figure highlights the changes from "Not Drowsy" to "Drowsy."

III. Model Construction

This section describes the drowsiness indices and our proposed drowsiness detection model. First, we introduce indices used in the drowsiness detection model and describe their characteristics, including our assumptions on postural features. We next present the proposed drowsiness detection model based on these characteristics. Finally, we describe our model training and evaluation method.

A. Drowsiness Indices

Adding to commonly used drowsiness indices, we introduced a posture index. Posture control involves the central nervous system [32] and thus posture change can reflect changes in the activity of nervous system. For instance, when people get drowsy, the body loses muscle engagement and posture is loosened little by little without individuals noticing, as illustrated in Fig. 3. Conversely, posture also has influence on sleepiness [36]. Although we do not discuss here which comes first, we assumed that the drivers' posture change—related to the change of nervous activity—has correlation to weak drowsiness of which they are not aware. Therefore, to detect weak drowsiness, we used a sitting posture-related index that represents such gradual changes in posture, adding to conventional drowsiness indices. Table II shows the categories of drowsiness indices and the primary measurements for each one. Notably, the overall drowsiness index includes the following four sub-indices: 1) vehicular, 2) blink, 3) posture, and 4) physiological.

1) Vehicular Index: As drowsiness deepens, driving accuracy declines, resulting in abrupt movements such as sudden course corrections due to loss of concentration and brain processing power. These changes related to driving operation, such as frequency in speed changes from acceleration to deceleration or vice versa and steering wheel movement—which is the sum total of the absolute values of angle changes and steering

TABLE II
MULTI-MODAL INDEX

Category		Measurements
Driving Information	1) Vehicular Index	Speed change frequency, standard deviation of steering angle, lateral acceleration, etc.
Behavioral Information	2) Blink Index	Blink duration, blink speed, blink frequency, PERCLOS [16], etc.
	3) Posture Index	Shift of center of gravity, lateral dispersion of center of gravity, head position, etc.
Physiological Information	4) Physiological Index	Breathing intervals, Heart rate, Heart rate variability (LF/HF, SDNN, pNN50) [37], etc.

wheel reversal frequency—are defined as the vehicular index and evaluated in each three-minute time window.

2) Blink Index: This study used conventional blink measures, including: PERCLOS [16], which indicates the proportion of time that the driver closes his/her eyes per unit time; blink frequency, which is the total number of blinks within the three-minute time window divided by three; and blink duration, which is the average time required to blink once within the three-minute time window. To calculate these blink indices, upper and lower eyelid positions were first identified from the face video, and blinks were detected from brief changes of the eyelid positions.

3) Posture Index: To capture changes in sitting posture, the average shift of the driver's center of gravity on the seating surface from the initial position, the lateral displacement of the position of the center of gravity, and the average head position coordinates on the facial image were selected as posture measures, and all evaluated per three-minute time window.

4) Physiological Index: Physiological features reflect nervous system activity [37]. Heart rate-related and respiration-related features, which can be captured without disturbing driving operation, were used as the physiological index, including: the low frequency/high frequency (LF/HF) ratio which is the ratio of the low frequency power to the high frequency power of heart rate variability; heart rate per minute, which is the total number of beats within the three-minute time window divided by three; and average breathing intervals within the three-minute time window.

B. Predicted Characteristics of the Indices

Each category of index has different characteristics in relation to drowsiness level, as shown in Fig. 4. Of the conventional indices, vehicular measures change little with slight drowsiness but rapidly with severe drowsiness, as indicated in previous studies [12], [34]. This means vehicular measures are sensitive to very deep drowsiness, but not to weak drowsiness. Blink measures change in a relatively wide range of drowsiness level compared to vehicular measures, but still reflect moderate or deep drowsiness, as indicated in [34]. Physiological measures show very wide individual differences [38] and were therefore regarded as ineffective for the proposed method, which did not train on individual drivers.

In contrast to the above indices, the posture index likely changes with weak drowsiness, as described above. In the case of severe drowsiness, posture changes are considered to stop as drivers lose most of their body power and fully lean on

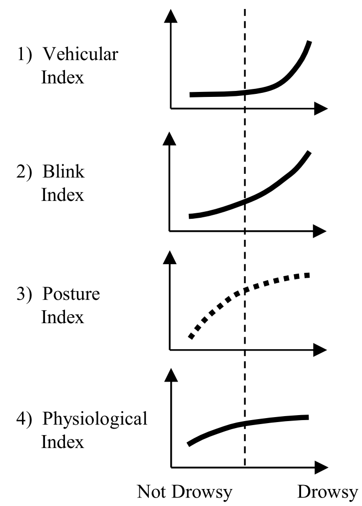


Fig. 4. General characteristics of drowsiness indices. The posture index's curve is a presumpti.

the backrest. As such, the posture index is sensitive to weak drowsiness and not to severe drowsiness.

C. Model Selection

An appropriate drowsiness detection model was established based on the index characteristics described in the previous subsection. Each index was treated as an output obtained when a function f was applied to the driver's drowsiness level, as follows:

$$\left[\, Index1_i \;\; Index2_i \;\; \ldots \;\; IndexN_i \,\right]^T = f\,(level) \tag{1}$$

Therefore, the model for obtaining the drowsiness level from the indices was regarded as the inverse function f^{-1}, as follows:

$$level = f^{-1}\left(\left[\, Index1_i \;\; Index2_i \;\; \ldots \;\; IndexN_i \,\right]^T\right) \tag{2}$$

Since the indices were nonlinear with respect to drowsiness level, as shown in Fig. 4, f and f^{-1} are both nonlinear functions. Furthermore, the drowsiness levels were defined as equidistant, and therefore a nonlinear regression model was used as f^{-1} to identify the drowsiness level distance for each measure. For the drowsiness detection models used conventionally, neural networks and support vector machines can be used as nonlinear regression models. However, neural networks need a large dataset to train. Therefore, Radial Basis Function (RBF)-kernel based-support vector regression [39],

[40], which is a modified support vector machine for nonlinear regression, was used in this study.

The estimated formula is defined as follows:

$$x_i = \left[Index1_i \ Index2_i \ \ldots \ IndexN_i \right]^T \tag{3}$$

$$f^{-1}(x_i) = \sum_{j=1}^{n} \left(a_j - a_j^*\right) \cdot K\left(x_i, x_j\right) + b \tag{4}$$

where x_i is the i-th example, x_j is a support vector, and $K\left(x_i, x_j\right)$ is the RBF kernel defined as

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right), \tag{5}$$

and a_j, b and support vectors are calculated using hyperparameters C, ε, and γ, as described in [39]. The optimization of the hyperparameters is referred to in Section III-D.

The indices used in the drowsiness detection model were combined to cover the entire drowsiness scale. As mentioned in Section III-B, blink index is sensitive to most drowsiness stages but not to weak drowsiness. The posture index, which is expected to be sensitive to weak drowsiness, can compensate for this shortcoming. Therefore, we propose a comprehensive drowsiness detection model combining blink and posture indices, which enables more accurate detection of all drowsiness levels. Adding other indices, such as vehicular index, might improve the accuracy for more severe drowsiness detection. Therefore, the accuracy of drowsiness detection based on the other index combinations is also provided for comparison.

D. Model Training and Evaluation

When evaluating the various models, the data for each drowsiness level was down-sampled to be equal in number to avoid the effect of drowsiness level imbalance, and thus 1000 data points were used to evaluate each model. Furthermore, to drop irrelevant and redundant measures, the indices used in each combination model were selected using L1 regularization before each model training. The accuracy of drowsiness detection models with different index combinations were compared using these data.

To train and evaluate a drowsiness detection model adaptable to drivers for general purposes without personalization, we used nested subject-wise cross-validation, which divides training and evaluation data subject by subject. The subjects were first divided into ten groups, and the data from only nine groups were further divided to select hyper-parameters of the model through cross-validation. The model was trained again with all the data from the nine groups using the selected parameters, and the data from the remaining one group was estimated using the trained model. This process was repeated, and the results estimated for each of the ten groups were designated as the final estimation results. As such, 1000 test samples were used.

The F1 score and root-mean-square error (RMSE) were used to evaluate estimation accuracy. Since this study focused on the estimation accuracy for each drowsiness level, the F1 score was calculated for each level. The precision for level i was the proportion of correctly predicted level i out of all predicted level i, and the recall for level i was the proportion of correctly predicted level i out of actual level i. Thus F1 for level i was calculated as

$$F1\,(leveli) = \frac{2 * precision\,(leveli) * recall(leveli)}{precision\,(leveli) + recall(leveli)} \tag{6}$$

The average F1 score for the drowsiness levels was designated as the estimation accuracy of the entire model. In addition, to compare with the existing method [11] which uses RMSE as the evaluation metric, the estimated results for the drowsiness rating values of each subject were also evaluated using RMSE as the accuracy evaluation of the entire model.

IV. Results

In this section, we first show the characteristics of acquired indices, on which our proposed model is based. We next present model performance.

Fig. 5 shows the index value distribution of typical measures from each of the index categories, and Table III shows the significant differences among the averages of index values by drowsiness level (Steel-Dwass test). The blue dashed lines in Fig. 5 represent the characteristics of each index and correspond to the general characteristics in Fig. 4. For example, the speed change frequency (VH1) has no significant differences in the range of drowsiness levels 1 through 3. On the other hand, the index increased rapidly from drowsiness level 4 to 5 as shown in Fig. 5, indicating correspondence with the general characteristics of the vehicular index, as shown in Fig. 4. The blink duration (BL1) showed statistically significant changes over the entire range of drowsiness levels, as shown in Table III—in particular, increasing rapidly at drowsiness level 5 in terms of the 25-75% range, as shown in Fig. 5. As predicted in Section III-A, the shift of center of gravity in the posture index (PS1) changed over almost all drowsiness levels, and was particularly sensitive at low drowsiness levels, showing its complementary relation with the blink index. Finally, the breathing interval of the physiological index changed over the entire drowsiness range only mildly, resulting in only minor differences in the index range among drowsiness levels.

Table IV shows the F1 scores and the average of RMSE by subject for each combination of indices. Among estimations based on one category of indices, the blink index resulted in the highest F1 score (0.429) and the smallest average subject RMSE (0.775). This error was smaller than the error of a previous published drowsiness detection method (0.91) that similarly used the driver's facial information [11], indicating that our method, which uses a regression model, achieved more accurate drowsiness detection than conventional methods that use non-regression models.

Fig. 6 shows the F1 score for each drowsiness level when only the blink index was used and when the blink index was combined with other indices. The detection accuracy of the blink only model is low in the range of drowsiness levels 1 through 3, compared with the range of levels 4 through 5. This result is consistent with the sensitivity of the blink index, which changes more in the higher drowsiness levels than the lower levels. Combining the blink index with the posture index

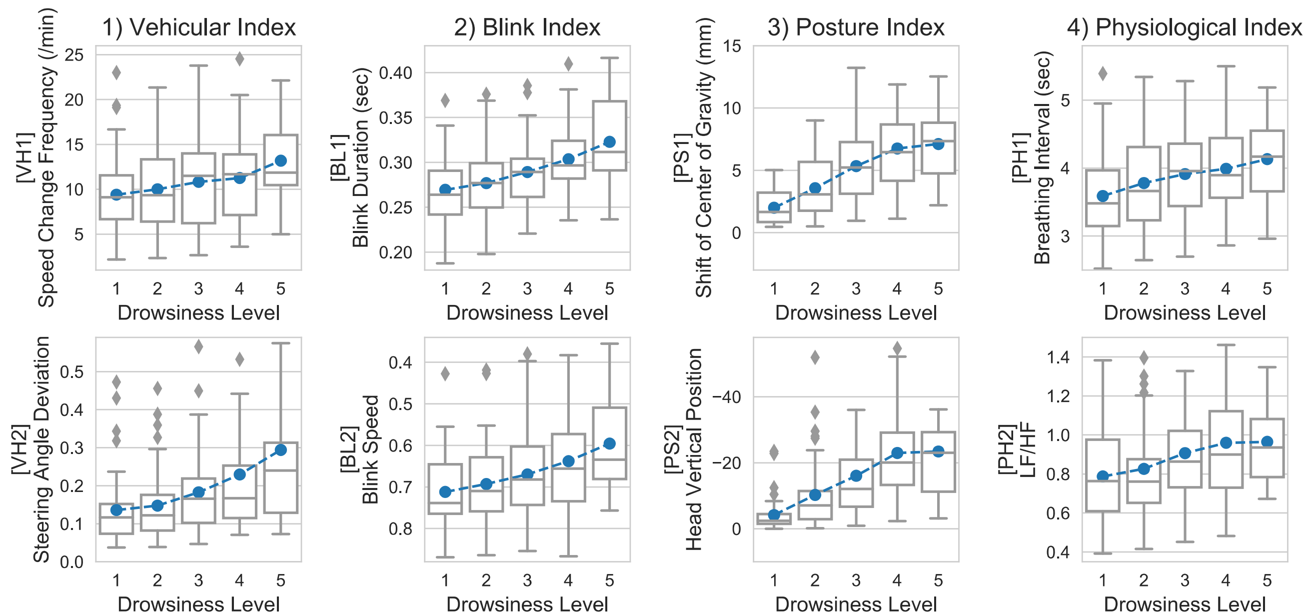
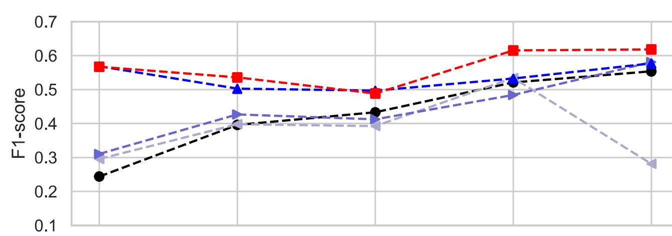


Fig. 5. Plots showing typical index values of categories by drowsiness level. Blue dashed lines and dots show the averaged values by drowsiness level. Diamonds in boxplots show outliers, and boxes show 25-75% value ranges. (VH: Vehicular, BL: Blink, PS: Posture, PH: Physiological.)

TABLE III
STATISTICAL SIGNIFICANCE OF THE TYPICAL INDICES

Compared levels		VH1	VH2	BL1	BL2	PS1	PS2	PH1	PH2
Inter low levels	Levels 1–2	n.s.	n.s.	*	n.s.	*	*	n.s.	n.s.
	Levels 1–3	n.s.	*	*	*	*	*	*	*
	Levels 2–3	n.s.	*	*	*	*	*	n.s.	*
Inter low – high levels	Levels 1–4	*	*	*	*	*	*	*	*
	Levels 1–5	*	*	*	*	*	*	*	*
	Levels 2–4	*	*	*	*	*	*	*	*
	Levels 2–5	*	*	*	*	*	*	n.s.	*
Inter high levels	Levels 3–4	*	*	*	*	*	*	*	n.s.
	Levels 3–5	*	*	*	*	*	*	n.s.	n.s.
	Levels 4–5	*	*	*	*	n.s.	n.s.	n.s.	n.s.

*:p<0.05, n.s.: not significant (Steel-Dwass test). Abbreviations correspond to Fig. 5.



Combined Index	Level 1	Level 2	Level 3	Level 4	Level 5
Blink	0.244	0.396	0.433	0.521	0.554
Blink + Physiological	0.296	0.398	0.392	0.534	0.281
Blink + Vehicular	0.310	0.427	0.413	0.484	0.581
Blink + Posture	**0.569**	0.503	**0.497**	0.532	0.577
Blink + Posture + Physiological + Vehicular	0.567	**0.536**	0.489	**0.615**	**0.618**

Fig. 6. F1 score by drowsiness level (when other modalities were combined with the blink index, the posture index was particularly effective at improving the score in drowsiness levels 3 and below). The best performance for each level is highlighted in bold.

improved the F1 score over the range of drowsiness levels 1 through 3. When all the indices were combined, the overall F1 score and RMSE improved slightly, as shown in the bottom line of Table IV.

V. Discussion

In this study, posture measures were used in addition to conventionally used blink measures to detect not only deep drowsiness but also slight drowsiness before degradation of driving performance. By combining these indices, the detection accuracy improved in the lower range of drowsiness (levels 1 through 3), indicating that the proposed method of combining the blink and posture indices is effective to cover all drowsiness levels.

As the position of the center of gravity showed significant differences in the range of slight drowsiness, i.e., drowsiness levels 1 through 3, this information improved the estimation accuracy of drowsiness levels 3 and below. One of the

TABLE IV
RMSE AND F1 SCORE OF ALL MODALITY COMBINATIONS

Used Index Category				F1 score	RMSE
Previous method [11]				-	0.91
BL				0.429	0.775
	PS			0.236	0.749
		PH		0.121	1.018
			VH	0.227	0.932
BL	**PS**			**0.536**	**0.620 ***
BL		PH		0.380	0.771
BL			VH	0.443	0.744
	PS	PH		0.252	0.784
	PS		VH	0.343	0.740
		PH	VH	0.213	0.926
BL	**PS**	**PH**		**0.553**	**0.610 ***
BL	**PS**		**VH**	**0.544**	**0.599 ***
BL		PH	VH	0.445	0.740
	PS	PH	VH	0.272	0.780
BL	**PS**	**PH**	**VH**	**0.565**	**0.601 ***

BL: Blink, PS: Posture, PH: Physiological, VH: Vehicular.
The F1 score is the average F1 scores of the drowsiness levels.
The RMSE is the average RMSE of subjects.
The conditions with significantly smaller RMSE than Blink only (Dunnett test, $p<0.05$) are highlighted in bold and with a *.

possible reasons why posture information showed different characteristics from blink information is the difference in the aspect of restraint. There is no restrictive equipment on the drivers' face, but their body position is softly restricted by the position of equipment such as the steering wheel, pedals, and seat. Because the most comfortable seated position is largely determined by the seat shape, the seating position is almost the same between individuals and is therefore independent of individual differences. In particular, during long-term driving which is most likely to cause drowsiness, drivers rarely sit on unstable edges, thereby limiting the range of variation in the position of the center of gravity. These restrictions stabilize the measure, making the position of the center of gravity a useful index. However, as there is concern that the change of posture would be induced by relaxing as well as drowsiness, the cause of posture changes requires further investigation. Nevertheless, because the drivers practiced before the main trial and could be considered adequately relaxed, the posture change during driving in our study was likely the effect of drowsiness.

Conversely, the vehicular index, which showed similar characteristics to the blink index and may have little additional information, improved the estimation accuracy only marginally. Moreover, the physiological index, which did not show significance in higher drowsiness levels, reduced the accuracy in drowsiness level 5. This reduction may have been caused by overfitting to training dataset subjects because the physiological index has a large variation between subjects.

Combining all indices improved the accuracy over the blink and posture indices model, especially in drowsiness levels 4 and 5, which might be a contribution of the vehicular index. However, the overall improvement was small and the worst F1 score by drowsiness levels (0.489) was not improved compared with the blink and posture combination (0.497). Therefore, for practical use, in which fewer sensors and non-intrusive measures are desirable to reduce cost, the blink and posture indices model is the most feasible and effective way to detect all levels of drowsiness.

VI. CONCLUSION

In this paper, we proposed a drowsiness detection model designed to cover all drowsiness levels, from slight to severe. The posture information was particularly useful in conjunction with blink information because the posture index showed higher sensitivity to weak drowsiness than conventional information and was able to compensate for the shortcomings of the blink information. Since blink and posture information can be obtained even while not driving, this knowledge has the potential to contribute to drowsiness detection for occupants during automated driving in addition to manual driving.

Future studies will focus on the development of arousing and arousal-maintenance technologies after drowsiness detection. Of note, people in a state of slight drowsiness are likely to be aroused by a relatively weak stimulus. Our achievement of drowsiness detection over a wide range of levels, including slight drowsiness, will enable the provision of interfaces that allow the selection of appropriate stimuli optimized to match the driver's degree of drowsiness and driving conditions.

REFERENCES

[1] National Highway Traffic Safety Administration. (Oct. 2017). *Traffic Safety Facts Crash Stats: Drowsy Driving 2015*. [Online]. Available: http://www.nhtsa.gov/risky-driving/drowsy-driving

[2] European New Car Assessment Program. (Sep. 2017). *Euro NCAP 2025 Roadmap*. [Online]. Available: https://cdn.euroncap.com/media/30700/euroncap-roadmap-2025-v4.pdf

[3] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 6937–16953, Dec. 2012.

[4] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[5] C. Bila, F. Sivrikaya, M. A. Khan, and S. Albayrak, "Vehicles of the future: A survey of research on safety issues," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1046–1065, May 2017.

[6] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jul. 1990.

[7] W. W. Wierwille and L. A. Ellsworth, "Evaluation of driver drowsiness by trained raters," *Accident Anal. Prevention*, vol. 26, no. 5, pp. 571–581, 1994.

[8] H. Kitajima, N. Numata, K. Yamamoto, and Y. Goi, "Prediction of automobile driver sleepiness: 1st Report, rating of sleepiness based on facial expression and examination of effective predictor indexes of sleepiness," *Trans. Jpn. Soc. Mech. Eng. C*, vol. 63, no. 613, pp. 93–100, Sep. 1997.

[9] P. Philip *et al.*, "Fatigue, sleep restriction and driving performance," *Accident Anal. Prevention*, vol. 37, pp. 473–478, May 2005.

[10] K. Kaida *et al.*, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, Jul. 2006.

[11] S. Hachisuka, T. Kimura, K. Ishida, H. Nakatani, and N. Ozaki, "Drowsiness detection using facial expression features," SAE Tech. Paper 2010-01-0466, Apr. 2010.

[12] Y. Saito, M. Itoh, and T. Inagaki, "Driver assistance system with a dual control scheme: Effectiveness of identifying driver drowsiness and preventing lane departure accidents," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 660–671, Oct. 2016.

[13] W. W. Wierwille, L. A. Ellsworth, S. S. Wreggit, R. J. Fairbanks, and C. L. Kirn, "Research on vehicle-based driver status/performance monitoring: Development, validation, and refinement of algorithms for detection of driver drowsiness," U.S. Dept. Transp., Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Tech. Rep. DOT HS 808, vol. 247, Dec. 1994. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/2578

[14] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: A simulator study," *Accident Anal. Prevention*, vol. 35, no. 3, pp. 381–391, May 2003.

[15] P. M. Forsman, B. J. Vila, R. A. Short, C. G. Mott, and H. P. A. Van Dongen, "Efficient driver drowsiness detection at moderate levels of drowsiness," *Accident Anal. Prevention*, vol. 50, pp. 341–350, Jan. 2013.

[16] D. F. Dinges and R. Grace, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," Federal Highway Admin., Office Motor Carriers, Washington, DC, USA, Tech. Rep. FHWA-MCRT-98-006, 1998. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/113

[17] P. P. Caffer, U. Erdmann, and P. Ullsperger, "Experimental evaluation of eye-blink parameters as a drowsiness measure," *Eur. J. Appl. Physiol.*, vol. 89, nos. 3–4, pp. 319–325, May 2003.

[18] R. O. Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013.

[19] S. Elsenbrush, M. J. Harnish, and W. G. Orr, "Heart rate variability during waking and sleep in healthy males and females," *Sleep*, vol. 22, no. 8, pp. 1067–1071, 1999.

[20] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 58–75, Jul. 2012.

[21] P. K. Stein and Y. Pu, "Heart rate variability, sleep and sleep disorders," *Sleep Med. Rev.*, vol. 16, no. 1, pp. 47–66, 2012.

[22] Z. Li, S. E. Li, R. Li, B. Cheng, and J. Shi, "Driver fatigue detection using approximate entropic of steering wheel angle from real driving data," *Int. J. Robot. Autom.*, vol. 32, no. 3, pp. 291–298, 2017.

[23] Z. Li, S. E. Li, R. Li, B. Cheng, and J. Shi, "Online detection of driver fatigue using steering wheel angles for real driving conditions," *Sensors*, vol. 17, no. 3, p. 495, Mar. 2017.

[24] Z. Li, L. Chen, J. Peng, and Y. Wu, "Automatic detection of driver fatigue using driving operation information for transportation safety," *Sensors*, vol. 17, no. 6, p. 1212, May 2017.

[25] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3408–3420, Dec. 2017.

[26] C. J. De Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J. L. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accident Anal. Prevention*, vol. 126, pp. 95–104, May 2019.

[27] A. G. Correa, L. Orosco, and E. Laciar, "Automatic detection of drowsiness in EEG records based on multimodal analysis," *Med. Eng. Phys.*, vol. 36, no. 2, pp. 244–249, Feb. 2014.

[28] K. Q. Shen, X. P. Li, C. J. Ong, S. Y. Shao, and E. P. V. Wilder-Smith, "EEG-based mental fatigue measurement using multiclass support vector machines with confidence estimate," *Clin. Neurophysiol.*, vol. 119, no. 7, pp. 1524–1533, Jul. 2008.

[29] C. J. De Naurois, C. Bourdin, C. Bougard, and J. L. Vercher, "Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness," *Accident Anal. Prevention*, vol. 121, pp. 118–128, Dec. 2018.

[30] S. Barua, M. U. Ahmed, C. Ahlström, and S. Begum, "Automatic driver sleepiness detection using EEG, EOG and contextual information," *Accident Anal. Prevention*, vol. 115, pp. 121–135, Jan. 2019.

[31] D. Hallvig, A. Anund, C. Fors, G. Kecklund, and T. Akerstedt, "Real driving at night-predicting lane departures from physiological and subjective sleepiness," *Biol. Psychol.*, vol. 101, pp. 18–23, Sep. 2014.

[32] Y. Ivanenko and V. S. Gurfinkel, "Human postural control," *Front Neurosci.*, vol. 12, p. 171, Mar. 2018.

[33] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[34] M. Ingre, T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences," *J. Sleep Res.*, vol. 15, no. 1, pp. 47–53, Mar. 2006.

[35] X. Wang and C. Xu, "Driver drowsiness detection based on non-intrusive metrics considering individual specifics," *Accident Anal. Prevention*, vol. 95, pp. 350–357, Oct. 2015.

[36] M. Thody, V. H. Gregg, and R. J. Edwards, "Reclined sitting postures: The effect on human performance of a vigilance task," in *Proc. Ergonom. Soc. Annu. Conf.*, Edinburgh, Scotland, in (Contemporary Ergonomics), E. J. Lovesey, Ed. London, U.K.: Taylor & Francis, pp. 33–39, Apr. 1993.

[37] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017.

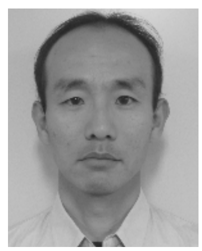
[38] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7235–7242, Jun. 2011.

[39] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[40] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. NIPS*, 1996.

Mika Sunagawa received the B.S. degree in engineering from Nagoya University, Nagoya, Japan, in 2011, and the M.S. degree in information science from Osaka University, Osaka, Japan, in 2013.

She joined Panasonic Corporation, Osaka, in 2013. Since 2016, she has been engaged in the research and development of driver monitoring systems. Her current research interests include human–machine interface using biosignals.



Shin-ichi Shikii received the M.S. degree in applied physics from Osaka University, Osaka, Japan, in 1998.

He joined Matsushita Electric Industrial Company, Ltd. (currently Panasonic Corporation), Osaka, in 2006, where he was engaged in the research and development of optical devices for novel displays. Since 2012, he has been engaged in the research and development of algorithms for human sensing based on physiological theory, especially for thermal sensation prediction for air conditioner. His current research interest includes sensing and actuating human drowsiness states.

Wataru Nakai received the B.S. degree in information engineering and the M.S. degree in information science from Osaka University, Osaka, Japan, in 2005 and 2007, respectively.

He joined Matsushita Electric Industrial Company, Ltd. (currently Panasonic Corporation), Osaka, in 2007, where he has been engaged in research and development involving sensing and actuating human states. He is currently engaged in the development of a system for sensing a driver's drowsiness states in a vehicle.

Makoto Mochizuki received the M.S. degree in precision mechanics from Tokyo University, Japan, in 2002, and the Ph.D. degree in human engineering from Kagawa University, Japan, in 2015.

He joined Matsushita Communication Industrial Company, Ltd. (currently Panasonic Corporation), Yokohama, Japan, in 2002, where he was engaged in research and development regarding driver's status estimation by using facial image and physiological information. Since 2015, he has been engaged in the development of an evaluation methodology for car human–machine interface, such as head-up-display and touch-panel screen in terms of human engineering. He is currently engaged in the development of sensing algorithms for advanced driving support systems.

Koichi Kusukame received the B.S. degree in physics and the M.S. degree in energy science from Kyoto University, Kyoto, Japan, in 2004 and 2006, respectively.

He joined Matsushita Electric Industrial Company, Ltd. (currently Panasonic Corporation), Osaka, Japan, in 2006, where he was engaged in the research and development of second harmonic generation laser for novel displays. Since 2010, he has been engaged in the research and development of algorithms for human sensing using laser, infrared radiation, and ultrasound, such as thermal sensation prediction for air conditioner and photoacoustic spectroscopy. His current research interest includes sensing and actuating human drowsiness states.

Hiroki Kitajima received the M.A. degree in experimental psychology from Keio University, Tokyo, Japan, in 1989, and the Ph.D. degree in engineering from the Chiba Institute of Technology in 2019.

He became a Researcher at The Institute for Science of Labour (currently The Ohara Memorial Institute for Science of Labour) in 1991. In this institute, he continues his research on human–machine interaction from the viewpoint of science of labor, which includes evaluation of human sleepiness.