Received 19 March 2021; revised 1 June 2021 and 25 July 2021; accepted 29 July 2021. Date of publication 3 August 2021; date of current version 12 August 2021.

Digital Object Identifier 10.1109/OJITS.2021.3102125

Toward Health–Related Accident Prevention: Symptom Detection and Intervention Based on Driver Monitoring and Verbal Interaction

HIROAKI HAYASHI [1], MITSUHIRO KAMEZAKI [2] (Member, IEEE), AND SHIGEKI SUGANO [3] (Fellow, IEEE)

[1] Department of Modern Mechanical Engineer, Graduate School of Creative Science and Engineering, Waseda University, Tokyo 162-0044, Japan

[2] Research Institute for Science and Engineering, Waseda University, Tokyo 162-0044, Japan

[3] Department of Modern Mechanical Engineering, Waseda University, Tokyo 169-8555, Japan

CORRESPONDING AUTHOR: H. HAYASHI (e-mail: h_hayashi@sugano.mech.waseda.ac.jp)

This work was supported in part by JST SCORE (STSC19015) and in part by the Research Institute for Science and Engineering (RISE), Waseda University.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Committee of the Waseda University, under Application No. 2019-276.

This article has supplementary downloadable material available at https://doi.org/10.1109/OJITS.2021.3102125, provided by the authors .

ABSTRACT Professional drivers are required to safely transport passengers and/or properties of customers to their destinations, so they must keep being mentally and physically healthy. Health problems will largely affect driving performance and sometimes cause loss of consciousness, which results in injury, death, and heavy compensation. Conventional systems can detect the loss of consciousness or urgently stop the vehicle to prevent accidents, but detection of symptoms of diseases and providing support before the driver loses consciousness is more reasonable. It is challenging to earlier detect symptoms with high confidence. Toward solving these problems, we propose a new method with a multi-sensor based driver monitoring system to detect cues of symptoms quickly and a verbal interaction system to confirm the internal state of the driver based on the monitoring results to reduce false positives. There is almost no data that records abnormal conditions while driving and tests with unhealthy participants are dangerous and ethically unacceptable, so we developed a system with pseudo-symptom data and did outlier detection only with normal driving data. From data collection experiments, we defined the confidence level derived from cue signs. The results of evaluation experiments showed that the proposed system worked well in pseudo headache and drowsiness detection scenarios. We found that signs of drowsiness varied with individual drivers, so the multi-sensor based driver monitoring system was proved to be effective. Moreover, we found that there were individual differences in how the cue signs appeared, so we can propose an online re-training method to make the system adapt to individual drivers.

INDEX TERMS Symptom detection system, health-related accident prevention, human factors, multi-sensor systems, road safety.

I. INTRODUCTION

DRIVER'S health problems can drastically decrease driving performance and increase accident risk. On Oct. 20, 2020, in the United States, a school bus driver suffered a stroke while driving and caused a crush, resulting in two children were injured. On Jan. 4, 2021, in Japan, a taxi driver suffered a stroke while driving and plunged into a pedestrian crossing, then one died and five were seriously injured. Professional drivers, who engage to drive buses, trucks, and taxis, play an important role in safely transporting passengers and properties of customers to their destinations, so health is important for the drivers to keep driving safely. The ministry of land, infrastructure, transport, and tourism (MLIT) of Japan [1] reports

The review of this article was arranged by Associate Editor Emmanouil Chaniotakis.

that hundreds of health-related accidents occur in transport industry every year in Japan and that number is increasing year by year. Obstructive Sleep Apnea Syndrome (OSAS), which is deeply influenced by lifestyle-related diseases, recently became a fatal public health problem. Previous studies have indicated that patients with OSAS are at an increased risk of vehicle accident since OSAS degrades the overall sleep quality, which causes daytime sleepiness [2]. Moreover, cerebrovascular and cardiovascular diseases occupied more than 60% of the diseases that caused the death of the drivers in the past ten years in Japan [1]. Also, in other countries, many cases of cerebrovascular and cardiovascular caused traffic accidents were reported [3]–[5].

Toward prevention of health-related accidents, companies in the transport industry and MLIT are seeking effective ways to achieve the following possible solutions [6].

1) Encourage drivers to keep a healthy condition outside working hours to reduce the risk of infection.
2) Conduct regular medical examinations to manage drivers' health and employment.
3) Check drivers' health and confirm if the driver is ready to work in the roll call before driving.
4) Handle problems when the driver suffers from some health problem while driving.

For solution (a), work environment and drivers' lifestyle greatly affects drivers' health condition, so company administrators make effort to improve the work environment such as reducing overwork and provide advice and support for the drivers' health management, including diet, sleep, and exercise. For solution (b), screening test and medical opinions can effectively reduce the risk of health-related accidents. Company administrators manage drivers' working time based on the result of the health examination. For solution (c), company administrators check drivers' health state by their voice and complexion and so on during the roll call just before starting to drive. Here, solution (d) can be an important one to deal with the onset of diseases while driving, which is basically difficult to predict and prevent. However, there are still no effective methods for the solution (d). Several systems are currently adopted to detect driver's 'loss of consciousness' while driving. DriverKarte [7] monitors driver's face and gaze direction while driving to estimate if the driver has lost consciousness and fell. In addition, recent vehicles are equipped with autonomous brake systems to avoid vehicle collisions. These supportive systems would be useful to avoid or reduce the damage of accidents, but its detection time when the driver becomes incapable of driving is 'too late' to satisfactorily prevent the accidents.

In most cases, some 'symptoms' of diseases would appear before drivers lose consciousness. Table 1 lists the symptoms, which are medically recognized, of cerebrovascular diseases, cardiovascular diseases, and OSAS. As stated above, fatal accidents are mostly caused by the cerebrovascular and cardiovascular diseases and OSAS had increased accident risks, so we focus on the above three diseases in this study. The

TABLE 1. Symptoms of diseases.

Diseases: Categories	Diseases: Types	Symptoms
Cerebrovascular disease	• Subarachnoid hemorrhage • Hemorrhaging of brain • Stroke	• Facial paralysis • Hemiplegia • Language disorder • Visual disorder • Headache
Cardiovascular disease	• Cardiac infarction • Angina	• Chest pain
Respiratory disease	• Obstructive sleep apnea syndrome (OSAS)	• Sleepiness

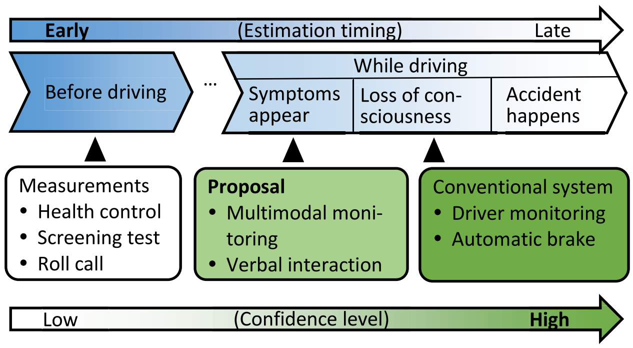


FIGURE 1. Confidence level and proposed system.

America Stroke Association (ASA) lists five main categories of stroke symptoms [8]: (i) sudden numbness or weakness of face, arm, or leg, especially on one side of the body, (ii) sudden confusion, trouble speaking or understanding speech, (iii) sudden trouble seeing in one or both eyes, (iv) sudden trouble walking, dizziness, loss of balance or coordination, (v) sudden severe headache with no known cause. In [9], key symptoms of heart attack were revealed as chest pain or discomfort. Furthermore, symptoms of OSAS include excessive daytime sleepiness, fatigue, impaired cognition and so on [10]. Basically, it is desirable that drivers recognize symptoms and understand problems in their health condition, then stop driving by themselves. However, symptoms are apt to appear suddenly and disappear immediately, so drivers do not recognize them or regard them as nothing serious. Moreover, the mission of transportation may make some professional drivers ignore the symptoms and/or endure the ache. Therefore, a symptom detection system can help drivers to recognize their health problem and provide support before the health condition gets worse.

Note that, as shown in Fig. 1, the relationship between 'estimation timing,' a timing when the system tries to detect symptoms, and 'confidence level,' accuracy of the symptoms which the system detected, is a trade-off. It is not difficult to judge if a driver has lost consciousness. We might predict the onset of disease from detected symptoms before loss of consciousness, but its accuracy is not guaranteed because such symptoms have large variations with individuals. Toward the above problems, the purpose of this study is to develop a system that detects symptoms before loss of consciousness with a high confidence level.

The proposed symptom detection system has two practical subsystems; the multi-sensor based driver monitoring system and verbal interaction system. Conventional monitoring systems focus on monitoring driver's drowsiness, cognitive burden, and situational awareness [11]–[13]. On the other hand, the proposed system focuses on evaluating driver's health state. The multi-sensor based driver monitoring system could detect symptoms quickly by monitoring the driver's face, body posture, physiological data, and driving maneuver. The verbal interaction system could reduce false positives by asking drivers whether or not they have any symptoms?

The original contribution of this paper is as follows. We present a multi-sensor based monitoring system to quickly detect symptoms and a verbal interaction system to reduce false positives. We develop such a system using only pseudo-symptom data and normal driving data due to the lack of abnormal data. We also define cue signs by referring to symptoms of cerebrovascular disease, cardiovascular disease, and SAS, which are the major diseases causing health-related accidents. We then define the confidence level by using detected cue signs to determine how to interact with the driver. We finally confirm from the evaluation experiments that the monitoring and verbal interaction system worked effectively. Extending [14], we improve the accuracy of the posture and eyes-state classification systems by selecting optimal pre-trained models and increasing the amount of data. Moreover, we analyze the effectiveness of the multi-sensor system over a single-sensor system and evaluate drivers' mental demand to discuss if the verbal interaction system interferes with driving tasks or not. From the results, we found that different drivers showed different cue signs of drowsiness, so the multi-sensor based driver monitoring system is effective, and that the verbal interaction system could increase drivers mental demand when the driver is at an abnormal state, but the increased mental demand would offer a positive effect such as making the drivers to be aware of their abnormality. Finally, we provide insight of online re-training method to further improve the confidence level in symptom detection.

The rest of this paper organized as follows; Section II provides the related works. The proposed system is described in Section III. Sections IV and V show the experiments and results. Section VI provides further analysis of proposed system. Finally, conclusions are provided in Section VII.

II. RELATED WORKS

In this section, we analyze conventional studies on health-related vehicle accident prevention systems.

A. MONITORING

Generally, a screening test is useful to detect potential health disorders or diseases for peoples who do not have any symptoms yet, including several detailed and invasive examinations such as a blood test and X-ray examination, which are difficult to be installed inside vehicles. Recent development of in-cabin sensor technologies has enabled to measure the driving environment and driver behavior on a large scale. For example, '100-Car Naturalistic Driving Study' uses cameras and many vehicle state and kinematic sensors to collect large-scale and naturalistic driving data [15]. Previous studies have revealed that a system with multiple sensors has several advantages including improving signal to noise ratio, reducing ambiguity and uncertainty, and increasing confidence and robustness against interference [16]. In [17], a multi-sensor system consisting of camera, depth sensor, and radar was developed for driver's hand-gesture recognition. In [18], multiple on-board sensors, including ocular sensor, performance sensor, and clock, were developed for driver drowsiness detection. As for health monitoring, multiple sensors were used together for measuring physiological information [16], [19]. The physiological data is useful for monitoring driver's health condition. However, the symptoms of diseases do not always appear in physiological data. For example, symptoms of stroke often appear even in a normal blood pressure states [20]. On the other hand, most of the symptoms listed in Table 1 are accompanied by pain, paralysis, and others, which will cause physical movement of body. For example, drivers may massage their head and lean their body forward when they have headache. Thus, body movement could be an important information for the detection. Moreover, bad health conditions would affect driving performance [21] and cause depression of driver's physical and cognitive functions and unstable maneuvers or none of maneuvers. Thus, driving maneuvers also could be important for detecting symptoms. To reduce the omission of detection, a multi-sensor based monitoring system, which could monitor not only physiological information but also body movement and driving maneuvers, is required.

B. INTERACTION

The advantage of the monitoring system is to observe the driver non-invasively, so there is no burden on the driver. Moreover, it can monitor continuously for a long time, so it can quickly detect symptoms at any time. On the other hand, a bi-direction interaction will help to increase the confidence level. There are two kinds of interaction, which are physical and verbal interaction. One famous example of physical interaction is FAST, i.e., face drooping, arm weakness, speech, and time to call 9-1-1 [22]. If a driver could not perform the FAST successfully, it is high probability that he/she has stroke. However, it is dangerous to perform non-driving related tasks while driving. As stated in the previous section, driving performance is affected by health condition, so we here assume the driving maneuvers as physical interaction. As for verbal interaction, medical doctors often use interview as the primary way of obtaining comprehensive information about a patient [23]. In addition, a communication robot estimates user's health condition from body temperature, blood pressure, weight, and interview responses, to support the users [24]. As for driving situation, as the solution (b) stated

TABLE 2. Symptoms and cue signs.

Symptoms	Cue signs			
	Posture and facial	Physiological		Maneuver
		Heart rate	Blood pressure	
Hemiplegia	One-hand hang	Tachycardia	High	Unstable maneuver, Non-maneuver
Headache	Massage head	Tachycardia	High	
Facial paralysis	Face distorts	Tachycardia	High	
Language disorder	No response, Fault response	Tachycardia	High	
Visual disorder	Abnormal gaze behavior	Tachycardia	High	
Chest pain	Hold chest	Tachycardia	High	
Sleepiness	Eyes closed	Bradycardia	Low	

in Section I, before driving, managers can measure driver's physiological information and ask some questions about the driver's health state based on the physiological information. Moreover, as a kind of hand-free interface, verbal-based human-machine interfaces are introduced to modern intelligent vehicles. In [25], voice alert systems were designed as the most effective form of alert in conveying information to the driver while driving. In [26], a navigation dialogue system was proposed so that drivers could tell the system the destination while driving. In addition to the monitoring system, we propose a verbal interaction system to improve the system confidence by reducing false positives of the monitoring system.

III. SYMPTOM DETECTION SYSTEM

A. REQUIRED WORKS

As stated in the previous section, when a driver feels a headache, which is a symptom of stroke, the driver may massage his/her head. Then the headache may fasten heartbeat (tachycardia) and raise blood pressure, and the driver could become unable to drive the vehicle as usual. These phenomena can be cue signs of headache (symptom). Table 2 lists these and other considerable cue signs. For example, a driver may hold his/her chest when he/she has heart pain, or the driver hangs one hand when he/she has hemiplegia. Sleepiness state will make a driver close his/her eyes for a long time, increase the blink period, and slow heart rate (bradycardia) [27], so 'eyes closed' and 'bradycardia' can be cue signs of sleepiness. These cue signs could be an indicator to estimate symptoms, so a monitoring system is required to detect these cue signs. Moreover, a verbal interaction that comprehends drivers' state could improve the system confidence by reducing the false positives of the monitoring system. Here, the timing and context of the interaction is important. If the system just repeats the verbal interaction at a certain time interval, without regard for the driver's state, drivers will find it annoying and ignore it. Thus, the verbal interaction system should adequately interact with drivers based on the results of monitoring. From the above analyses, as Fig. 2 shows, the proposed symptom detection system is required to have the following two functions.

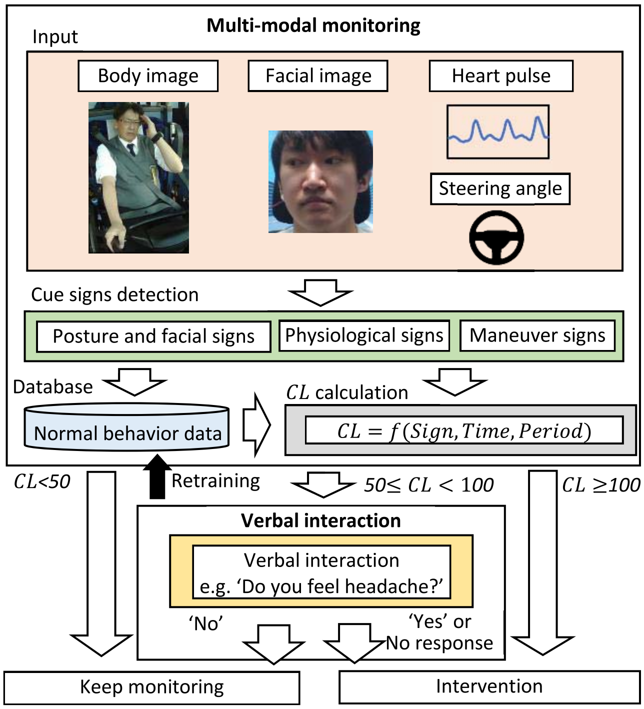


FIGURE 2. Symptom detection system.

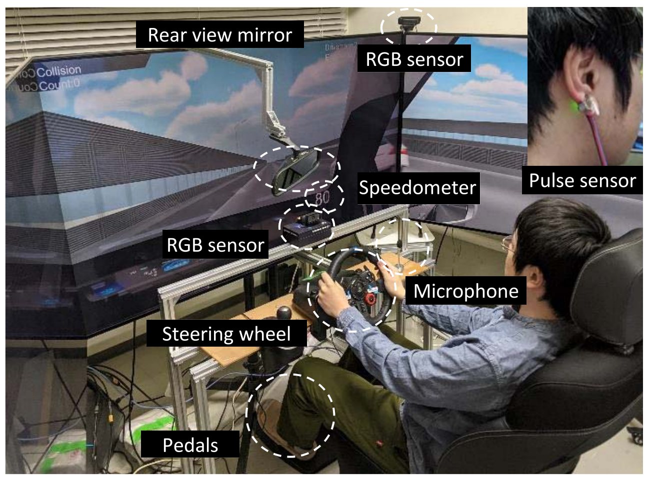


FIGURE 3. Driving simulator with monitoring and verbal interaction system.

- Multi-sensor based driver monitoring system. It monitors not only the driver's physiological data but also posture, facial data, and driving maneuver data.
- Verbal interaction system. To obtain internal state of drivers with higher confidence level, it communicates by speech with the drivers based on the result of monitoring.

B. OVERVIEW OF METHODOLOGY

In this paper, we develop and evaluate a fundamental framework with multi-sensor based driver monitoring and verbal interaction systems, but the framework does not cover all

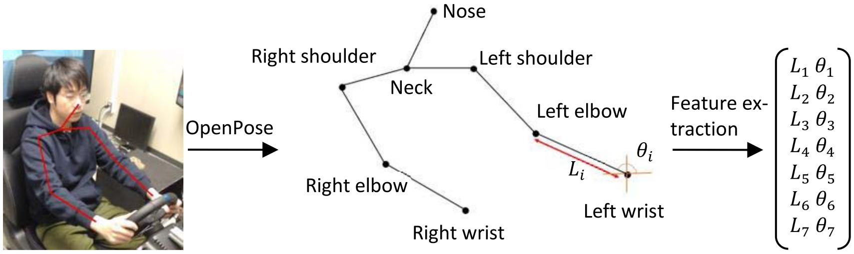


FIGURE 4. Feature values extracted from driver's posture.

considerable symptoms such as blood pressure and unstable maneuver due to a very preliminary study on symptom detection. The multi-sensor based monitoring system captures drivers' body and face, and measures the heart pulse and steering angle as system inputs. If cue signs are detected, the system will calculate the confidence level, and verbal interaction will be triggered based on it.

C. DATA SAMPLE

Anomaly detection systems generally need an enormous quantity of data in both normal and abnormal behaviors to define a boundary to classify them. However, collecting abnormal data in this study is very challenging. There is almost no data that records health-related accidents. Not all vehicles are equipped with devices to monitor the driver's condition, so only little health-related accidents are recorded. In recent years, personal data is strictly protected, so it is difficult to access the recorded data of health-related accidents. Thus, we propose to create pseudo-symptom database for developing and evaluating a cue sign detection system and conducted outlier detection only with normal data. To make the pseudo-symptom database, in exps-1 and 3 (explained later), we asked healthy persons to feign symptoms (e.g., headache). In connection with this, we used a driving simulator to observe driver's behavior and tested the effectiveness of developed systems. Toward future experiments with professional drivers in a real environment, in this study, we performed fundamental experiments with general drivers in a simulated environment. The simulator was developed using Unity (details are described in [28]). The simulator included four screens to display the virtual driving environment, a rearview mirror, a steering wheel, pedals, a real driver seat, two red-green-blue (RGB) cameras for driver monitoring, and a microphone for driver's speech input, as shown in Fig. 3. One RGB camera for capturing body posture was attached on the upper right of the middle screen and the other RGB sensor for capturing face was attached in front of the driver seat.

D. CUE SIGN DETECTION

We focused on seven cue signs in this study: (a) massage head, (b) hold chest, (c) one-hand hang, (d) eyes closed, (e) tachycardia, (f) bradycardia, and (g) non maneuver.

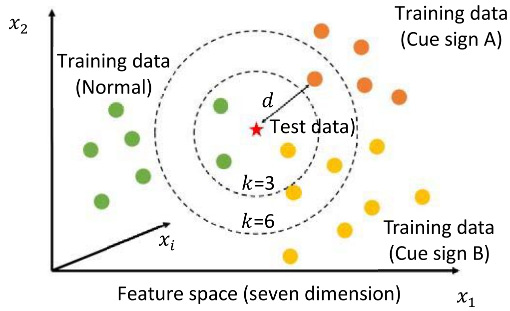


FIGURE 5. Algorithm of k-NN, the test data will be classified to Normal when $k = 3$, and Cue sign B when $k = 6$.

1) Body posture: We developed a posture classification system to estimate driver's upper body posture. A RGB sensor was attached on the upper right of the middle screen, as shown in Fig. 3, to capture the driver's upper body. We adopted OpenPose [29], which is a software library for real time detecting the key-points of the human body, hand, face, and foot from a single image. There are several pre-trained models in OpenPose. Compared with our previous study, we have changed the 'mobilenet_thin model' to the 'cmu model' which has a better benchmark in accuracy [30]. The system detects seven skeletons (eight key-points) of driver's upper body and classifies body postures according to closest training examples in the feature space by using k-nearest neighbor (k-NN), which is an easy-to-implement supervised machine learning method. The system calculates the product of the length and angle of each skeleton as feature points, as shown in Fig. 4. The system measures the distance between the test data and each training data based on Euclidean distance d, which was given by

$$d = \sqrt{\sum_{i=1}^{7} \left(L_{t,i}\theta_{t,i} - L_{s,i}\theta_{s,i}\right)^2}, \tag{1}$$

where $L_{t,i}$ and $\theta_{t,i}$ are the i th joint's length and angle of test data and $L_{s,i}$ and $\theta_{s,i}$ are the ones of training data, as shown in Fig. 5. There are four classes of posture, including normal, massage head, hold chest, and one-hand hang. We conducted an experiment to collect training data. We asked participants to provide the four kinds of posture by imagining when they have the symptoms. The details of data collection and system evaluation are described in Sections IV and V.

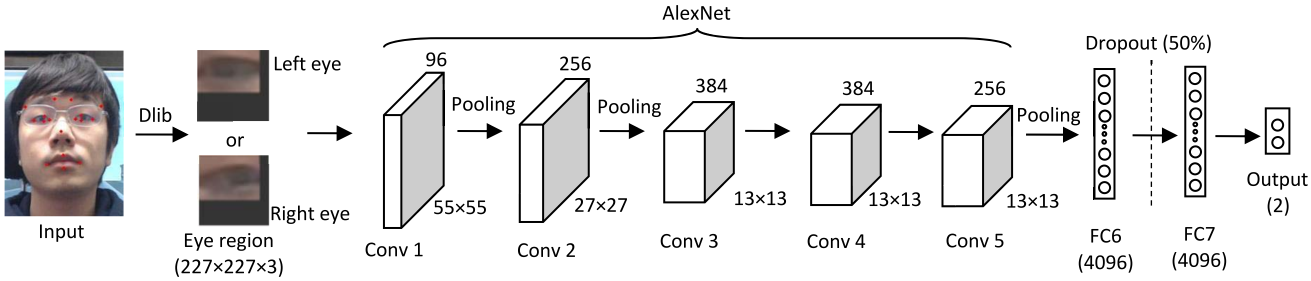


FIGURE 6. Architecture of eyes state classification system.

2) Eyes state: We developed an eyes-state classification system to estimate if driver's eyes are open or closed. The system structure is shown in Fig. 6. A RGB sensor was attached in the front of the driver seat, as shown in Fig. 3, to capture the driver's face. We adopted a convolutional neural network (CNN) to estimate each eye state. A transfer learning method was introduced because it could achieve high accuracy with a bit of learning. The eyes-state classification task is simple (only two classes) but required real-time processing, so a pre-trained network with fast processing speed is required. Thus, we chose AlexNet [31] which has less layers so that it could run fast. Furthermore, the input is drivers' eyes region which is extracted by Dlib library [32] and reshaped with pixel size of 227×227. The last three fully connected layer's unit was set as 4096, 4096, and 2, respectively. The two classes of output are open and closed. Here, a dropout layer was implemented in the fully connected layer to prevent the system from overfitting. We performed experiments to obtain training data in the same process as developing the posture classification system. The details of data collection and system evaluation are described in Sections IV and V.

3) Heart rate: The system measures driver's heart pulse by using an ear-mounted pulse sensor, as shown in the right side of Fig. 3. Heart rate variability are often quantified by using R–R interval, which is the time elapsed between two successive heartbeats on the electrocardiogram. The heart rate HR (beats/min) can be derived by dividing 60 s into R–R interval. According to [33], HR always varies ±10% from the mean value even in a resting state. Thus, the system classifies the HR into three classes (normal, tachycardia, and bradycardia) with a margin, and which is given as

$$\begin{cases} \text{Normal}, & \text{if } HR_{mean} \times 0.9 \le HR(t) \le HR_{mean} \times 1.1 \\ \text{Tachycardia}, & \text{if } HR(t) > HR_{mean} \times 1.1 \\ \text{Bradycardia}, & \text{if } HR(t) < HR_{mean} \times 0.9 \end{cases} \quad (2)$$

where HR_{mean} means the mean heart rate when a driver is relaxing and $HR(t)$ means the driver's heart rate of at time t.

4) Vehicle maneuver: The system detects if drivers maneuver the vehicle or not. 'Non-maneuver' can be easily detected from the angle of steering wheel. If there is no input of steering, the system identifies the state as non-maneuver, and otherwise, the state as maneuver.

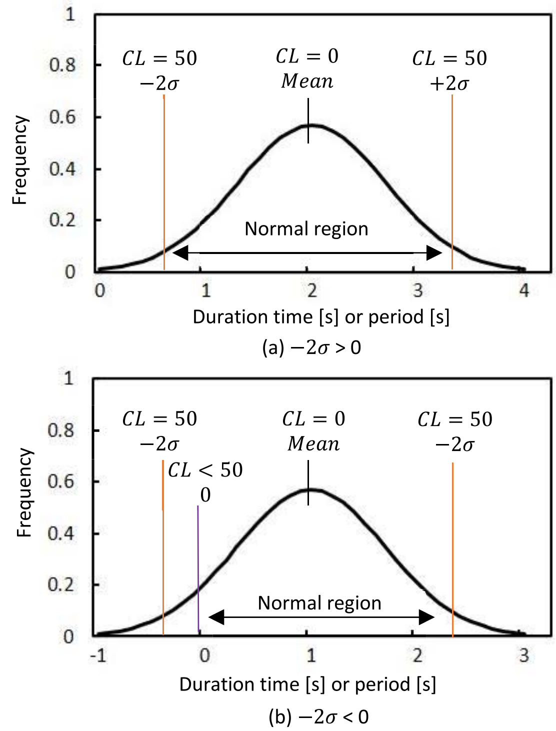


FIGURE 7. Gaussian distribution and confidence level (*CL*).

E. CONFIDENCE LEVEL (CL) CALCULATION

Drivers sometimes express cue signs about body posture even when they are in a normal state. For example, they rub their head when their head is itchy. Distinguishing itchy head from headache is a challenging task. Headache, hemiplegia, and other symptoms often continue for tens of minutes, so we adopted the duration time for distinguishing them. We thus proposed a confidence level (CL) that is calculated from cue sign data including the type, duration time, and period.

In general, there are three types of anomaly detection methods: 'supervised anomaly detection,' 'semi-supervised anomaly detection,' and 'unsupervised anomaly detection.' The 'supervised one' and 'semi-supervised one' require a labeled data set which includes both normal and abnormal data. On the other hand, many studies [34], [35] rely on the 'unsupervised one' because it works with unlabeled data set. When most of the instances in the data set are normal, it can

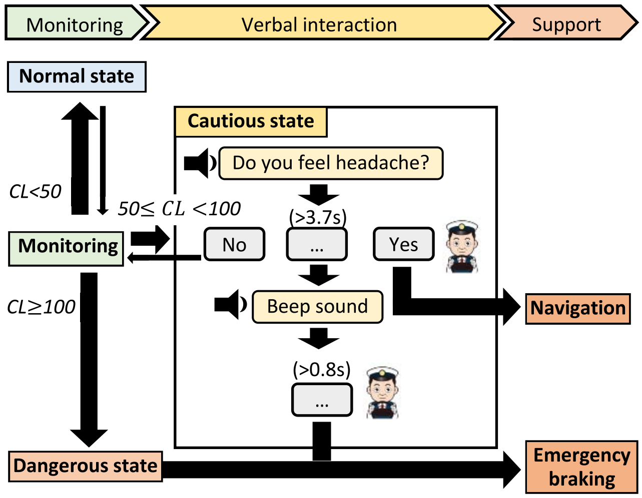


FIGURE 8. Flowchart of monitoring and verbal interaction systems.

detect anomalies by just checking how much a data deviates from the normal region. As stated in Section III, due to the difficulty of collecting symptom data, we chose the 'unsupervised one.' There are several types in unsupervised anomaly detection: (a) one for data following a normal distribution, e.g., Hotelling's T2 statistic [36] and Mahalanobis Taguchi system, and (b) the other for data not following a normal distribution, e.g., k-means clustering, local outlier factor, and isolation forests. The (b) needs to manually define a threshold, so we chose the (a) to achieve a more objective system. The Hotelling's T2 statistic is used for univariate data and the Mahalanobis-Taguchi system is for multivariate data. To perform verbal interaction, the system needs to understand which features are abnormal, so a univariate analysis is required to explore each feature in the data set, separately. From the above, we designed the CL calculation formulas based on the Hotelling's T2 statistic.

We first conducted experiments to investigate how the cue signs appear while normal driving. We then made a Gaussian distribution of the duration time and period of each detected cue sign, on the basis of the probability density function f.

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3}$$

$$\mu = \left(match \sum x\right)/n, \quad \sigma = \sqrt{\sum (x-\mu)^2/(n-1)}, \tag{4}$$

where $\sum x$ means the sum of duration time T or period P, n means the number of detection of each cue sign, and μ and σ means the mean and standard deviation of T or P, respectively. Examples of distribution is shown in Fig. 7. A mean of the $\pm 2\sigma$ region (95.45%) was set as a normal region. CL is set to 0 when a value is just the same as the mean and is set to 50 when a value is just $+2\sigma$ or -2σ, as shown in Fig. 7 (a). When -2σ is smaller than zero, the normal region is set to from 0 to 2σ, as shown in Fig. 7 (b). As state above, headache, hemiplegia, and other symptoms often continue for tens of minutes, so we assume that the cue signs will appear at a longer time and higher frequency than those when the driver is in a normal state. Thus, when T comes to be longer or P comes to be shorter than the normal range, CL will increase more than 50 and the system judges that the driver has a symptom.

By using T, P, and the mean duration time T_{mean} and period P_{mean} at the distribution, CL from duration time CL_t and period CL_p are given by

$$\begin{cases} CL_{t,i} = k_{t,i} \times (T_i - T_{mean,i}), \text{ if } T_i > T_{mean,i} \\ CL_{p,i} = k_{p,i} \times (P_{mean,i} - P_i), \text{ if } P_i < P_{mean,i} \end{cases}, \tag{5}$$

where k_t and k_p are the coefficient and i is the type of cue sign. Each k can be calculated by substituting $CL = 50$ and $T{=}T_{2\sigma}$ or $P{=}P_{2\sigma}$ in (5), and is given by

$$\begin{cases} k_{i,t} = 50/(T_{2\sigma,i} - T_{mean,i}) \\ k_{i,p} = 50/(P_{mean,i} - P_{-2\sigma,i}) \end{cases}, \tag{6}$$

where $T_{2\sigma}$ and $P_{-2\sigma}$ mean the duration time and period at the $+2\sigma$ and -2σ point. Finally, the system should intervene if one of the CL s exceeds 50, so the system derives the maximum value as the final confidence level CL as

$$CL = \max\left(CL_{i,t} \cup CL_{i,p}\right). \tag{7}$$

Note that the CL calculation method for 'eyes closed' differs from the above method. Compared with our previous study, we reconsidered the special characteristics of 'eyes closed' and redesigned its CL calculation method. As defined above, if the period of a cue sign is shorter than the normal region, it will be estimated as a symptom. But in the case of 'eyes closed,' humans sometime blink twice rapidly at less than 1 second, even when they are in a normal state. On the other hand, humans will repeat blink rapidly more than three times in 1–2 seconds, called blink burst, when they are drowsy [37]. In this study, a rapid twice blink will be estimated as normal, but a repeated rapid blink, i.e., more than three time, will be estimated as a symptom, e.g., sleepiness.

F. VERBAL INTERACTION SYSTEM

When the system detects a fallen state or long-time eyes closed state, it can definitely recognize that the situation is dangerous, so the system must immediately make CL higher and intervene vehicle control, e.g., performing emergency braking, to prevent accident. However, when CL is a medium level (lower than dangerous but higher than normal), the system should ask the driver to respond his/her conditions by performing verbal interaction, to judge if CL should be higher or lower. The system's speech was prepared in advance as WAV files. The speech recognition part was developed by using Google Speech Recognition API [38], Fig. 8 shows the flowchart of monitoring and verbal interaction with three states according to CL score.

1) Normal state: When CL is low ($CL < 50$), the system defines the state as normal and continues monitoring the driver.

2) Cautious state: When CL is reasonably high ($50 \leq CL < 100$), the system defines the state as cautious and performs verbal interaction. For example, when the system detected 'massage head,' it would ask the driver 'Are you OK? Do you feel headache?' If the driver responds 'No,' the system finishes verbal interaction and continues monitoring. If the driver responds 'Yes,' the system induces the driver to stop driving while navigating the vehicle the nearest parkable area. If there was no response from the driver within a certain time, the system beeps for confirmation. We set the time as 3.7 s by taking into consideration that the driver might not have heard the system's voice. If there was even no response within a different certain time (0.8 s), the system judges that the driver is in an extremely dangerous state and performs emergency braking. We chose 3.7 and 0.8 s by reference to necessary time for responding to information and warning reported by MLIT [39].

3) Dangerous state: When CL is sufficiently high ($CL \geq 100$), the system defines the state as dangerous, which means that it can definitively recognize the driver as in imminent danger. Therefore, the system immediately performs emergency braking to avoid vehicle collisions.

IV. EXPERIMENT

We describe about three experiments including the scenarios of experiments, data acquisition, participants, and experimental procedures. Compared with our previous study, we have increased the number of participants and evaluated the system by detecting pseudo-symptoms.

A. SCENARIOS

1) Experiment 1: The aim was to obtain training data for developing body posture and eyes-state classification systems. Participants were asked to drive in a straight highway and perform three kinds of body posture including massage head, hold chest, and one-hand hang once with each hand. While the experiment, their eyes movement were also recorded.

2) Experiment 2: The aim was to observe how cue signs appear while normal driving and calculate the distribution of the duration time and period of each kind of cue sign. Participants were asked to drive in a straight highway while obeying traffic rules. Driving time was one hour. To obtain data only during normal driving, we manually removed data that participants looks sleepy or had dangerous maneuvers. The system monitors participants' body movement, eyes movement, heart pulse, and steering angle to detect cue signs.

3) Experiment 3: The aim was to test the proposed symptoms detection system. Again, it is very challenging to evaluate the system which targets unhealthy people. For a drowsiness detection system, it is possible to ask participants to drive with the driving simulator for a long time in very morning to make the participants sleepy [40]. For an alcohol detection system, it is possible to ask participants to drink alcohol and then drive on an empty road [41]. However, it is ethically unacceptable to make participants' health abnormal like headache and hemiplegia since these symptoms are life-threatening or will leave after-effects with a high probability. In this study, instead of evaluation of real symptoms detection, we only tested whether the system works as designed by detecting pseudo-symptoms. Here, we targeted the sudden headache, which is a symptom of stroke. To evaluate the system, we only asked participants to feign a sudden headache but did not tell them how to feign it. Participants were asked to drive for more than five minutes at first, and then feign a headache at any time.

B. EXPERIMENTAL SETUP AND CONDITIONS

1) Data acquirement: Driver's face and body were captured in Full HD (1080p), and all the data were sampled at 15 Hz.

2) Participants and procedure: The same twenty participants (7 females and 13 males, age 21–30 yrs.) participated the exps-1 and 2. Other five participants (5 males, age 21–30 yrs.) participated the exp-3. Since it is hard to experiment with a large group of people, we controlled the population by taking the sample from the age group of 21–30. All the experiments conducted in this study have been approved by the Ethical Review Committee of the Waseda University. Before the experiments, all participants have been informed of the experimental purpose and procedures, then signed a written consent form. At the first of each experiment, participants were well trained how to control the vehicle in the driving simulator. After that, they were asked to drive for five minutes, to measure each participants' standard mean heart rate for normal driving. For the experiment which will be described in Section VI, participants were asked to answer a questionnaire after the first five minutes of driving and each time after verbal interaction. The questionnaire asked how much they felt mental demand for the tasks on a 0–20 scale. After experiments, participants were provided with monetary compensation for their contribution.

V. RESULTS AND ANALYSIS

In this section, we describe and analyze the results of experiments, and discuss the implications in detail.

A. RESULT OF EXPERIMENT 1

We extracted images from the recorded videos at a regular frame interval, and manually labeled the images. Totally, 1600 images of body posture (20 participants$\times$4 classes$\times$20 images) and 2800 images of eyes (20 participants$\times$2 classes$\times$70 images) were obtained. 70% of the collected images were used for training and the other 30% for test. In other words, the system was tested with unknown data to evaluate its ability to generalize. By trial and error approach, we finally set k to 30 for the k-NN and the training runs for the CNN was set to

FIGURE 9. Output of posture classification system. The percentage shows the number of the closet training data out of k (k=30).

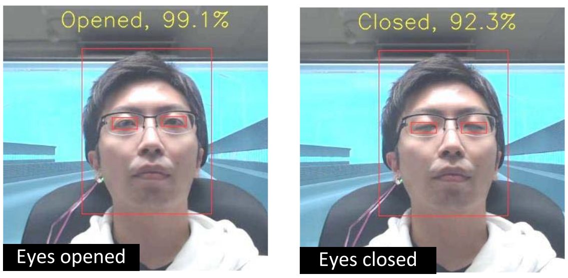
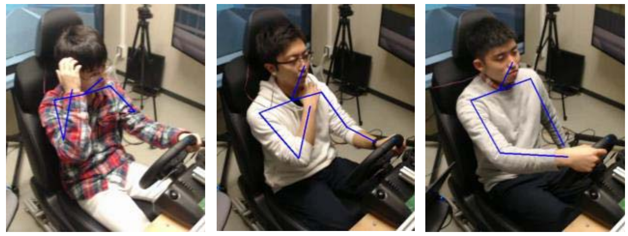


FIGURE 10. Output of eyes state classification system. The percentage shows probability of estimation.

500 epochs using back-propagation. Tables 3 and 4 list the confusion matrix which summarizes the results of each classification system. The accuracy of posture classification system resulted in 91–100%, and eyes state classification system resulted in 99–100%. Moreover, Figs. 9 and 10 show examples of detection result. We found from the results that the proposed system could adequately detect cue signs. In general, the detection accuracy is expected to be close to 100%, so in future, we will increase training data and optimize the model for improving the accuracy.

B. RESULT OF EXPERIMENT 2

We could find from the results that all the cue signs appeared while normal driving. Fig. 11 shows examples that cue signs detected in normal driving. A reason the system detected 'massage head,' 'hold chest,' and 'one-hand hang' was that participants felt itchy of their head, felt itchy of their neck, and felt tired with their hand, respectively. Here, the neck and chest are close together, so massaging neck was detected as 'hold chest.' The system detected that some participants had 'tachycardia' when they swerved out of their lane by mistaking maneuvers. Then, they felt rushed and tried to return to the lane. Moreover, the system detected that some participants had 'bradycardia' because they became calmer so that the heart rate became lower as the experiment progressed. Table 5 lists the mean and deviation of each cue sign. By following the method described in Section III-C, we calculated 2σ region and coefficients k of each cue sign, as listed in Table 6. Then, CL can be calculated by the coefficient k. As stated above, if the duration time or period of detected cue signs are not in the normal region, CL will be higher than 50 and the system will start verbal interaction.



(a) Massage head (b) Hold chest (c) One-hand hang

FIGURE 11. Cue signs detected in normal driving.

TABLE 3. Results of evaluation of posture classification system.

	Normal	Massage head	Massage chest	Hand hang
Normal	100%	0%	0%	0%
Massage head	0%	92%	0%	8%
Massage chest	4%	0%	96%	0%
Hand hang	1%	8%	0%	91%

TABLE 4. Results of evaluation of eyes-state classification system.

	Open	Close
Open	99%	1%
Close	0%	100%

TABLE 5. Distribution of detected cue signs.

Cue signs	Duration time		Period	
	Mean (in sec)	Deviation (in sec)	Mean (in sec)	Deviation (in sec)
Massage head	2.0	1.2	173.7	233.2
Hold chest	2.4	1.1	349.7	340.6
One-hand hang	3.1	1.9	497.3	556.5
Eyes closed	0.3	0.2	3.5	7.2
Tachycardia	2.3	3.7	28.3	80.4
Bradycardia	2.0	2.8	6.5	18.3
Non-maneuver	3.5	7.7	0.8	0.7

TABLE 6. Coefficients.

Cue signs	Duration time		Period	
	2σ region (in sec)	k_t	2σ region (in sec)	k_p
Massage head	0.8–3.3	40.6	0–405.6	0.2
Hold chest	1.3–3.4	47.7	11.0–688.5	0.1
One-hand hang	1.2–5.0	26.6	0–1050.7	0.1
Eyes closed	0.1–0.5	272.5	0–10.7	7.0
Tachycardia	0–5.9	13.8	0–108.3	0.6
Bradycardia	0–4.8	17.8	0–24.8	2.7
Non-maneuver	0–11.2	6.6	0.1–1.5	71.2

C. RESULT OF EXPERIMENT 3

We detected driver's pseudo headache state to test if the proposed system could work as designed. As a result, all the five participants feigned a headache by massaging their head, and all these five pseudo headaches were detected. The system also detected cue signs when the participants did not feign a pseudo headache due to itchy head, but the

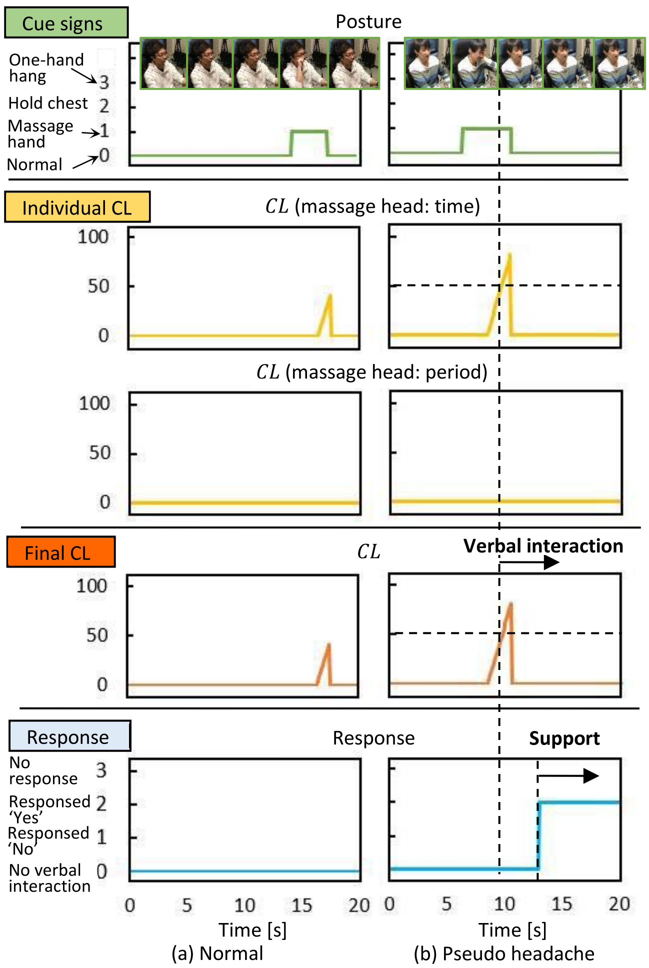
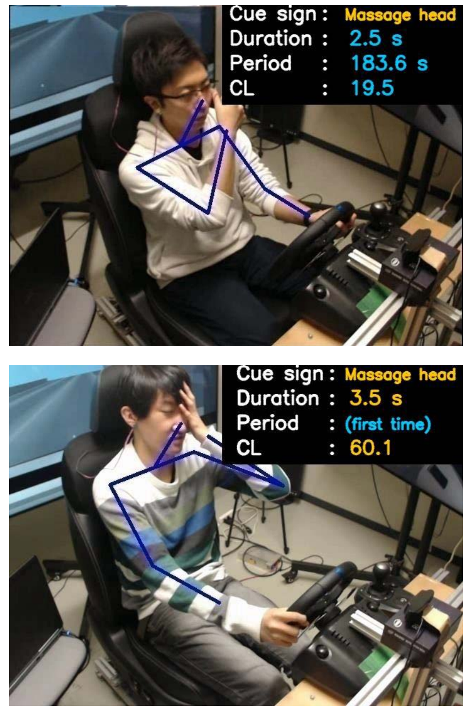


FIGURE 12. Examples of system behavior on pseudo headache detection.



(a) Normal

(b) Pseudo headache

FIGURE 13. Example scenes of 'massage head' detection.

system could judge that they were not headache as expected. Fig. 12 shows examples of how the system behaves on pseudo headache detection, including cue signs detection, *CL* calculation, and verbal interaction. The system monitored drivers' posture for cue sign detection, then individually calculated *CL* based on the detection results of each cue sign and derived the final *CL* by taking the maximum of each *CL*. Finally, the system started verbal interaction based on the final *CL*. Fig. 13 shows the detected scenes which related to Fig. 12. As for Figs. 12 (a) and 13 (a), 'massage head' was detected once but the duration time (2.5 s) was within the range of 2σ region (0.8–3.3s), so *CL* remained below 50, and system regarded it as normal behavior (e.g., 'the drive felt head itchy'). As for Figs. 12 (a) and 13 (b), 'massage head' was detected and the duration time (3.5 s) exceeded the range of 2σ region, so *CL* became higher than 50, and the system estimated that the driver felt 'headache.' Then the system started verbal interaction. In this example, the driver responded 'Yes,' meaning that the driver has headache, so the system started navigating the vehicle to the nearest parkable area. From the above results, we found that the system could work as expected to detect driver's pseudo headache states.

VI. DISCUSSION

In this section, we conducted a drowsiness detection experiment to deeply analyze the proposed system.

A. DROWSINESS DETECTION

In the Experiment 3, we tested pseudo headache state. However, the system only monitored the body posture of the driver. To deeply understand and evaluate the effectiveness of the multi-sensor based monitoring system and verbal interaction system, we conducted a drowsiness detection experiment. We asked participants to drive for one hour in a straight highway, and they were expected to fall fatigue with a monotonous drive. The room temperature was set to 25°C. The system monitored their eyes, heart rate, and maneuver then calculated *CL* based on the distribution listed in Tables 5 and 6. Once *CL* was higher than 50, the system performs verbal interaction and asks participants 'Do you feel sleepy?' Like the flowchart in Fig. 8, if they answered 'No,' we continue the experiment. If they answered 'Yes,' the system induces them to stop driving or autonomously stops the vehicle. Five participants (4 males, 1 female, age 21–26 yrs.) participated in this experiment. The procedures described at Section IV-B2) were all adopted to this experiment. At the end of experiment, participants were asked to respond questionnaire of 'mental demand,' which was scaled to 0–20 by reference to NASA task load index [34].

As a result, the system detected drowsy more than two times for some participants, and the total detection number

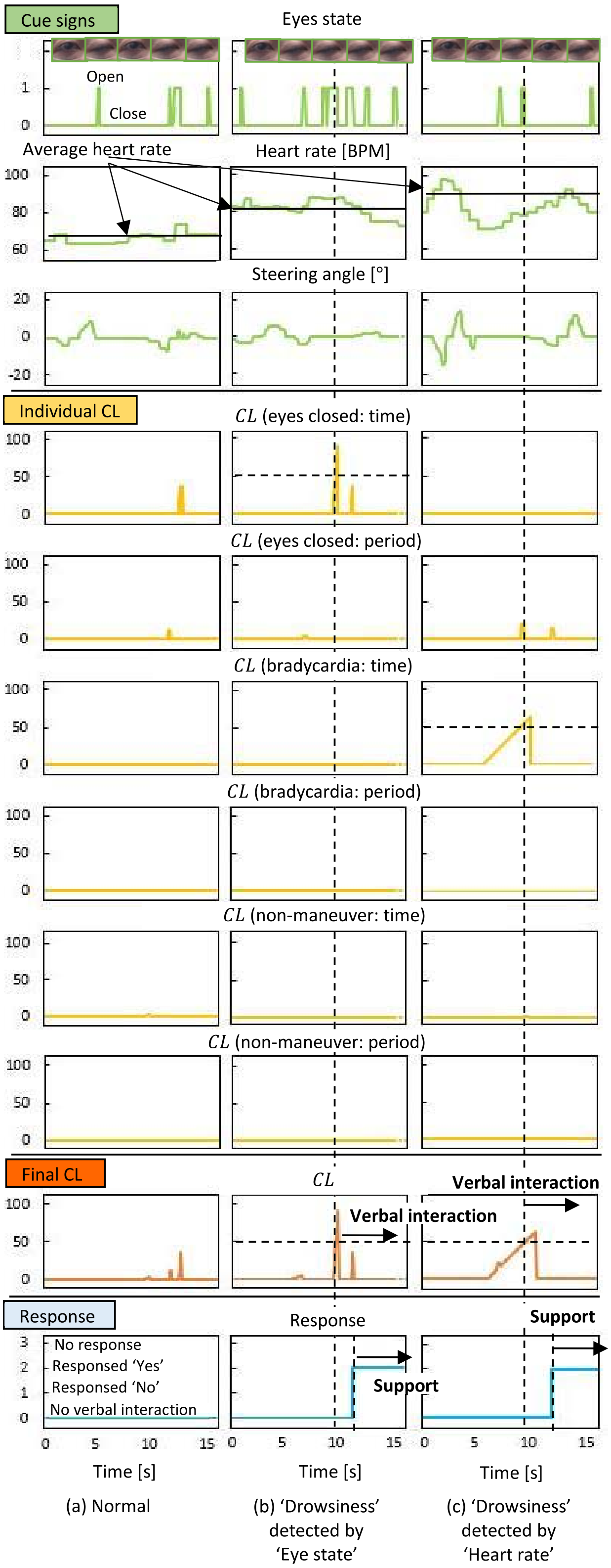


FIGURE 14. Examples of system behavior on drowsiness detection.

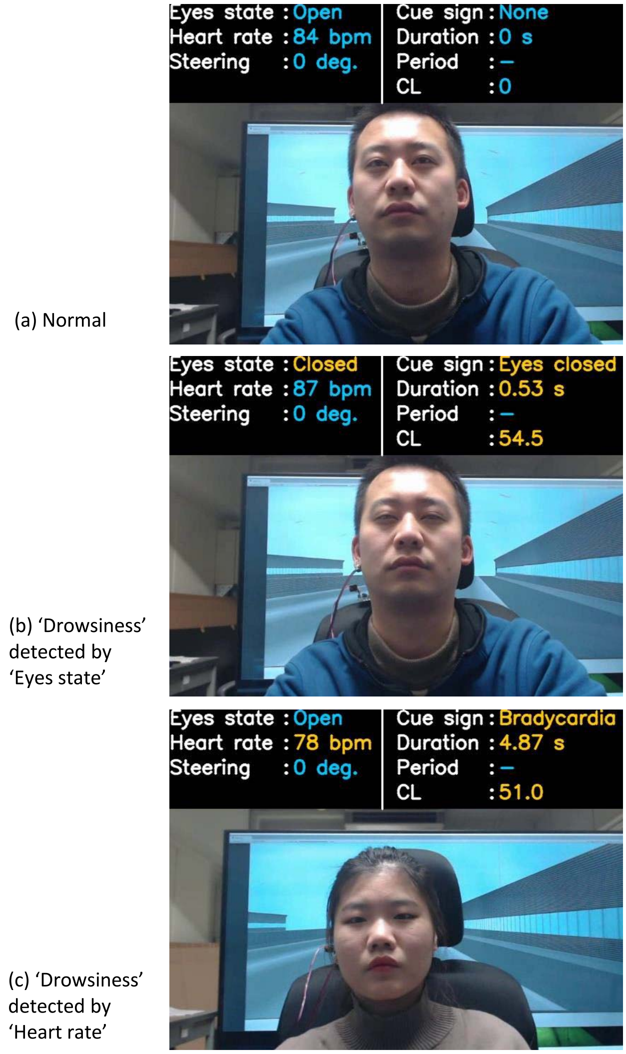


FIGURE 15. Example scenes of 'drowsiness' detection.

of times was nine. Among them, five cases of detection were based on long duration of the 'eyes closed' state, three cases were based on the 'bradycardia,' and remaining one case was based on the 'non-maneuver.' Responses to verbal interaction in each detection were listed in Table 7. From the table, we found that all the defined cue signs were important since different drivers showed different cue signs of drowsiness. Different persons and different diseases would cause different symptoms, so the proposed multi-sensor based monitoring system that monitors drivers from different aspects would be more effective than a single-sensor based monitoring system.

Fig. 14 shows examples of how the system behaves on drowsiness detection, including cue sign detection, *CL* calculation, and verbal interaction. The system monitored drivers' eyes state, heart rate, and steering angle for cue signs detection then individually calculated *CL* based on the detection results of each cue sign and derived the final *CL* by taking the maximum of each *CL*. Finally, the system started verbal interaction based on the final *CL*. Fig. 15 shows the detected scenes which related to Fig. 14. As for Figs. 14 (a) and 15 (a), all the eyes state, heart rate, and maneuver varied within the normal range, so *CL* stayed below 50. As

TABLE 7. Results of drowsiness detection.

Subject No.	Detected times	Cue sign	Response
1	1st time	Eyes closed	'Yes'
2	1st time	Eyes closed	'No'
	2nd time	Eyes closed	'Yes'
3	1st time	Eyes closed	'No'
	2nd time	Eyes closed	'Yes'
4	1st time	Bradycardia	'No'
	2nd time	Non-maneuver	'Yes'
5	1st time	Bradycardia	'No'
	2nd time	Bradycardia	'Yes'

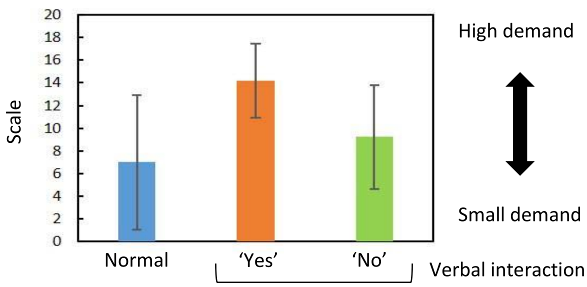


FIGURE 16. Result of mental demand questionnaire.

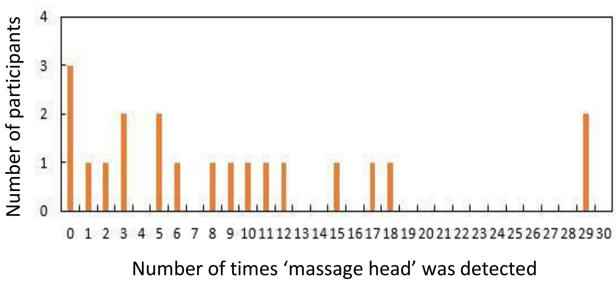


FIGURE 17. Individual differences in number of times when 'massage head' is detected.

for Figs. 14 (b) and 15 (b), the duration time of 'eyes closed' (0.53 s) exceeded the range of 2σ region (0.1–0.5 s), so *CL* became higher than 50, and the system estimated that the driver felt sleepy, and then started verbal interaction. As for Figs. 14 (c) and 15 (c), the duration time of 'bradycardia' (4.87 s) exceeded the range of 2σ region (0–4.8 s), so *CL* became higher than 50, and the system estimated that the driver feels sleepy then started verbal interaction. At both situations of Figs. 14 (b) and (c), the drivers responded 'Yes,' meaning that the drivers felt sleepy, so the system started navigating the vehicle to the nearest parkable area.

Compared to the monitoring system, the verbal interaction system could confirm driver's internal state, but it might increase driver's mental demand. To clarify it, we evaluated the mental demand by using a questionnaire. The result is summarized at Fig. 16. There are two results of verbal interaction, which are the driver response 'Yes' or 'No.' We compared these mental demand scores by using the analysis of variance (ANOVA) test and we found that there was no significant difference among each mental demand score ($p > 0.05$). The result indicates that the system did not increase driver's mental demand. However, the mental demand score of 'Yes' was apparently higher than the other two. The proposed system may increase driver's mental demand if the driver was at an abnormal state, i.e., drowsiness, but the increased mental demand is not always bad for driving. This is because it can give a sense of tension to driver and make drivers aware that they are at an abnormal state.

B. INDIVIDUAL DIFFERENCE IN CUE SIGNS

In this study, we made the distribution of each cue sign using all the 20 participants' data and we developed a fundamental symptom detection system. However, different participants may have different habits, so this leads to different distribution of each cue sign. Fig. 17 shows the number of 'massage head' detected in one-hour driving. We found that the number significantly differs among individuals. Someone performed 'massage head' more than once every three minutes, while someone did not perform it. Because of such individual variability, individual-based models, which can deal with individual driver's behavior, would be more powerful for symptom detection. However, it is difficult to collect all drivers' data, so we can propose an online re-training method. The online re-training method is to initialize the system with existing data and automatically update the system depending on the result of verbal interactions. For example, if the system detected that a driver has a symptom of disease, but the driver responded 'No' in verbal interaction, the system could modify the internal definition of the symptom by adjusting the normal region.

VII. CONCLUSION AND FUTURE WORKS

Toward prevention of health-related accidents, we proposed a new symptom detection system that could detect symptoms earlier with a higher confidence level by multi-sensor based monitoring and verbal interaction. We analyzed the symptoms of cerebrovascular disease, cardiovascular disease, and SAS, and defined cue signs accordingly. The monitoring system could detect cue signs by monitoring driver's face, body, physiological information, and maneuver, then calculate the confidence level depending on the type, duration time, and period of cue signs. On the other hand, the verbal interaction system could directly communicate with the driver based on the result of monitoring to clarify the driver's health state in detail. From data collection experiments, we could define the confidence level by using detected cue signs to determine how to interact with the driver. Using the defined confidence level, the system could effectively detect pseudo headache and drowsiness. From the results of evaluation experiments, we found that multi-sensor system had advantages over single-sensor system in dealing with individual variability and that the verbal interaction system could help drivers to be aware of their abnormal state. Furthermore, we proposed an online re-training method to improve the confidence level.

In future works, we will improve the multi-sensor system by introducing different types of sensors (e.g., a blood-pressure sensor and gaze tracking system) and analyzing which sensors are effective for symptom detection, and we will introduce a voice diagnosis system that exams driver's health by analyzing tongue and rhythm of the driver's voice. As for the eyes-state classification system, we will introduce faster computation machines to evaluate pre-trained models that have higher accuracy. Moreover, we will introduce the proposed system to a real vehicle system, to evaluate the system with professional drivers in a real environment and realize the online re-training system.

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism. (2021). *Annual Car Accident Statistics for the Car Transportation Business*. [Online]. Available: https://www.mlit.go.jp/jidosha/anzen/subcontents/statistics.html

[2] Y. Komada, Y. Nishida, K. Namba, T. Abe, S. Tsuiki, and Y. Inoue, "Elevated risk of motor vehicle accident for male drivers with obstructive sleep apnea syndrome in the Tokyo metropolitan area," *Tohoku J. Exp. Med.*, vol. 219, no. 1, pp. 11–16, 2009.

[3] P. D. Wang and R. S. Lin, "Coronary heart disease risk factors in urban bus driver," *Public Health*, vol. 115, no. 4, pp. 261–264, 2001.

[4] A. Y. X. Yu, M. K. Kapral, J. Fang, and D. A. Redelmeier, "Increased risk of traffic injury after a cerebrovascular event," *Stroke*, vol. 49, no. 12, pp. 3006–3011, 2018.

[5] P. Marx *et al.*, "Position paper: Ability to drive in cerebrovascular diseases," *Neurol. Res. Pract.*, vol. 1, p. 37, Oct. 2019.

[6] Ministry of Land, Infrastructure, Transport and Tourism. (2014). *Driver's Health Management Manual for Commercial Vehicles*. [Online]. Available: https://wwwtb.mlit.go.jp/tohoku/jg/manual_kenkoukannri.pdf

[7] Omron. (2018). *DriveKarte*. [Online]. Available: https://www.oss.omron.co.jp/drivekarte/service_guide/index.html

[8] American Stroke Association. *Stroke Symptoms*. Accessed: Feb. 22, 2021. [Online]. Available: https://www.stroke.org/en/about-stroke/stroke-symptoms

[9] D. C. Goff *et al.*, "Knowledge of heart attack symptoms in a population survey in the United States: The REACT trial. Rapid early action for coronary treatment," *Arch. Int. Med.*, vol. 158, no. 21, pp. 2329–2338, 1998.

[10] M. R. Mannarino, F. D. Filippo, and M. Pirro, "Obstructive sleep apnea syndrome," *Eur. J. Internal Med.*, vol. 23, no. 7, pp. 586–593, 2012.

[11] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[12] U. E. Manawadu, T. Kawano, S. Murata, K. Kamezaki, J. Muramatsu, and S. Sugano, "Multiclass classification of driver perceived workload using long short-term memory based recurrent neural network," in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 2009–2014.

[13] H. Hayashi *et al.*, "A driver situational awareness estimation system based on standard glance model for unscheduled takeover situations," in *Proc. IEEE Intell. Veh. Symp.*, 2019, pp. 718–723.

[14] H. Hayashi, M. Kamezaki, N. Oka, and S. Sugano, "Development of an abnormal sign detection system based on driver monitoring and voice interaction for preventing medical-condition-caused car accidents," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 2020, pp. 1–8.

[15] T. A. Dingus *et al.*, "The 100-car naturalistic driving study—Phase II—Results of the 100-car field experiment," U.S. Dept. Transp., Nat. Highway Traffic Safety Admin., Washington, DC, USA, Rep. DOT HS 810 593, 2006.

[16] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks; State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

[17] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2015, pp. 1–8.

[18] R. Feng, G. Zhang, and B. Cheng, "An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster–Shafer theory," in *Proc. IEEE Int. Conf. Netw. Sens. Control*, 2009, pp. 897–902.

[19] S. Kaneko. (2019). *Improve of Traffic and Transportation Safety and Reliability-Research and Development Related to Preventive Measures Towards Secure Transportation System*. [Online]. Available: https://www.mlit.go.jp/common/001313790.pdf

[20] T. Morishita *et al.*, "Epidemiological studies on arteriosclerosis and hypertension I—The relationship between the data and cerebro-cardiovascular attack in mass screening test," *J. Japan Geriatr. Soc.*, vol. 18, no. 5, p. 324, 1981.

[21] M. M. Amick, J. Grace, and B. R. Ott, "Visual and cognitive predictors of driving safety in Parkinson's disease patient," *Arch. Clin. Neuropsychol.*, vol. 22, no. 8, pp. 957–967, 2007.

[22] J. E. Bray, B. O. Connell, A. Gilligan, P. M. Livingston, and C. Bladin, "Is FAST stroke smart? Do the content and language used in awareness campaigns describe the experience of stroke symptoms?" *Int. J. Stroke*, vol. 5, no. 6, pp. 440–446, 2010.

[23] S. B. Srivastava, "The patient interview," in *Fundamental Skills for Patient Care in Pharmacy Practice*, C. D. Lauster and B. S. Srivastava, Ed. Burlington, MA, USA: Jones Barltett Learn., 2014.

[24] K. Toda *et al.*, "A home healthcare system with communication robot technologies—Development of experimental systems and in-home verification experiments by older persons," *J. Rob. Mechatronics*, vol. 23, no. 6, pp. 951–968, 2011.

[25] S. M. Sarala, D. H. S. Tadav, and A. Ansari, "Emotionally adaptive driver voice alert system for Advanced Driver Assistance System (ADAS) applications," in *Proc. IEEE Int. Conf. Smart Syst. Inventive Technol.*, 2018, pp. 509–512.

[26] Y. Zheng, Y. Liu, and J. H. L. Hansen, "Intent detection and semantic parsing for navigation dialogue language processing," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 2017, pp. 1–6.

[27] J. M. Cori, C. Anderson, S. S. Soleimanloo, M. L. Jackson, and M. E. Howard, "Narrative review: Do spontaneous eye blink parameters provide a useful assessment of state drowsiness," *Sleep Med. Rev.*, vol. 45, pp. 95–104, Jun. 2019.

[28] U. Manawadu, M. Ishikawa, M. Kamezaki, and S. Sugano, "Analysis of individual driving experience in autonomous and human-driven vehicles using a driving simulator," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatornics*, 2015, pp. 299–304.

[29] G. Hidalgo, Z. Cao, T. Simon, S. Wei, H. Joo, and Y. Sheikh. *OpenPose Library*. Accessed: Feb. 1, 2020. [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose

[30] Z. Cao, T. Simon, S. Wei, and Y. Sheikg, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6647–6655.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[32] *davisking/dlib*. Accessed: Feb. 1, 2020. [Online]. Available: https://github.com/davisking/dlib

[33] D. J. Ewing, D. Q. Borsey, F. Bellavere, and B. F. Clarke, "Cardiac autonomic neuropathy in diabetes: Comparison of measures of R-R interval variation," *Diabetologia*, vol. 21, no. 1, pp. 18–24, 1981.

[34] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9592–9600.

[35] S. Khan, C. F. Liew, T. Yairi, and R. McWilliam, "Unsupervised anomaly detection in unmanned aerial vehicles," *Appl. Soft Comput.*, vol. 83, Oct. 2019, Art. no. 105650.

[36] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE Trans.*, vol. 27, no. 6, pp. 800–810, 1995.

[37] R. Horiuchi, T. Ogasawara, and N. Miki, "Fatigue assessment by blink detected with attachable optical sensors of dye-sensitized photovoltaic cells," *Micromachines*, vol. 9, no. 6, pp. 1–11, 2018.

[38] Google. *Cloud Speech-to-Text*. Accessed: Feb. 17, 2020. [Online]. Available: https://cloud.google.com/speech-to-text

[39] Ministry of Land, Infrastructure, Transport and Tourism. *Advanced Safety Vehicle Promotion Plan Report*. Accessed: Feb. 22, 2021. [Online]. Available: https://www.mlit.go.jp/jidosha/anzen/01asv/resourse/data/asv4pamphlet_seika.pdf

[40] X. Wang and C. Xu, "Driver drowsiness detection based on non-intrusive metrics considering individual specifics," *Accident Anal. Prevent.*, vol. 95, pp. 350–357, Oct. 2016.

[41] R. K. Jakkar, R. Pahuja, R. K. Saini, B. Sahu, and Natwar, "Drunk-driver detection and alert system (DDDAS) for smart vehicles," *Amer. J. Traffic Transp. Eng.*, vol. 2, no. 4, pp. 45–58, 2017.

HIROAKI HAYASHI received the B.S. and M.S. degrees in mechanical engineering from Waseda University in 2017, and 2019, respectively. He is currently pursuing the Graduate degree with the Department of Modern Mechanical Engineering, Waseda University. His research interests include human-machine interface systems, driving simulation, and driving automation systems. He received the Young Fellow Award from the Japan Society of Mechanical Engineers in 2020. He is a member of the JSME.

MITSUHIRO KAMEZAKI (Member, IEEE) received the B.S., M.S., and Doctor of Engineering degrees in mechanical engineering from Waseda University, Japan, in 2005, 2007, and 2010, respectively. From April 2010 to March 2013, he was a Research Associate with the Department of Mordern Mechanical Engineering, Waseda University, where he was an Assistant Professor with the Research Institute for Science and Engineering from April 2013 to March 2018. Since April 2018, he has been an Associate Professor with the Research Institute for Science and Engineering, Waseda University. His interests include intelligent machine system, man-machine interface, and operator support system. He received the Young Investigation Excellence Award from RSJ in 2016, the Best Paper Award from the IEEE/ASME AIM 2016, and the Best Paper Awards from RSJ and SICE. He is a member of the JSAE, SICE, RSJ, JSME, and IEEE RAS.

SHIGEKI SUGANO (Fellow, IEEE) received the B.S., M.S., and Doctor of Engineering degrees in mechanical engineering from Waseda University, Japan, in 1981, 1983, and 1989, respectively. From 1986 to 1990, he was a Research Associate in Waseda University. Since 1990, he has been a Faculty Member of the Department of Mechanical Engineering, Waseda University, where he is currently a Professor. From 1993 to 1994, he was a Visiting Scholar with Mechanical Engineering Department, Stanford University. He is a member of the Humanoid Robotics Institute, Waseda University. Since 2014, he has been serving the Dean of the School of Creative Science and Engineering, Waseda University. His research interests include human-symbiotic anthropomorphic robot design, dexterous and safety manipulator, and human–robot communication. He served as the General Chair of IROS2013 in Tokyo and SICE Annual Conference 2011. He is a Fellow Member of JSME, SICE, RSJ, and IEEE RAS.