

# Driver State Detection Based on Cardiovascular System and Driver Reaction Information Using a Graphical Model

Thanh Tung Nguyen<sup>1\*</sup>, Hirofumi Aoki<sup>2</sup>, Anh Son Le<sup>2</sup>, Hirano Akio<sup>2</sup>, Kunimoto Aoki<sup>2</sup>, Makoto Inagami<sup>2</sup>, Tatsuya Suzuki<sup>1</sup>

<sup>1</sup>Graduate School of Engineering, Nagoya University, Nagoya, Japan

<sup>2</sup>Institute of Innovation for Future Society, Nagoya University, Nagoya, Japan

Email: \*tung.nguyenthanh.hust@gmail.com

**How to cite this paper:** Nguyen, T.T., Aoki, H., Le, A.S., Akio, H., Aoki, K., Inagami, M. and Suzuki, T. (2021) Driver State Detection Based on Cardiovascular System and Driver Reaction Information Using a Graphical Model. *Journal of Transportation Technologies*, 11, 139-156.

<https://doi.org/10.4236/jtts.2021.112009>

**Received:** January 29, 2021

**Accepted:** February 28, 2021

**Published:** March 3, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Traffic accidents are mainly caused by human error. In an aging society, the number of accidents attributed to elderly drivers is increasing. One noteworthy reason for this is operation misapplication. Studies have been conducted on the use of human-machine interfaces (HMIs) to inform the driver when he or she makes an error and encourage appropriate actions. However, the driver state during the erroneous action has not been investigated. The purpose of this study is to clarify the difference in the driver's state between normal and surprising situations in a misapplication scenario, utilizing multimodal information such as biometric information and driver operation. We found significant changes in the interaction of components between the normal and the surprised driving state. The results could provide basic knowledge for the future development of a driver assistance system and driver state estimation using data acquired from multiple sensors in the vehicle.

## Keywords

Human Engineering, Driver Behavior, Bio-Signal, Safety

## 1. Introduction

A report released by the World Health Organization [1] in 2020 showed that there are around 1.35 million fatalities every year because of road traffic accidents. The factor that contributes most to a large number of accidents is human error: speeding, driving under the influence of psychoactive substances, distracted driving, etc. The development of the technology related directly or indirectly to vehicles has helped to reduce the physical stress imposed on drivers but

has also led to more distraction during the driving task. Klauer *et al.* [2], in a report on the impact of driver inattention on near-crashes, showed that secondary tasks are the factors that contributed most to inattention-related accidents. The driver inattention defined in the report including secondary tasks, drowsiness, drive-related inattention to the roadway, and non-specific eye glance toward the roadway.

Another emerging crash factor that is gaining attention is the driver's internal state [3] [4] [5] [6]. The sources of the internal state can come from the outside environment or inside the driver's mind. Startle response or surprise state are quite common in driving: a pedestrian sudden pass by, unintended acceleration or lough horn of a large truck, etc. Startle is a reaction to sudden, intense threatening stimulus while surprise is inclined toward cognitive and emotional response. These terms are often used interchangeably. Research about the effects of them on operation performance has been done in aviation [7] [8] [9]. In some extreme cases, the surprise may impair the pilot's troubleshooting capabilities. This is also true for the car driver. Subtle surprise events appear more often when driving on urban roads. More extreme surprises such as near hit can develop fear or panic feeling. In a report, K. Lococo *et al.* [10] stated that startle or panic was common associated with pedal misapplication in North Carolina, United States. The report also mentioned that 57% of the crashes were in the parking lots or driveways. In Japan, the Institute for Traffic Accident Research and Data Analysis (ITRADA) has published reports about driving operation errors and pedal misapplication [11] [12]. As for driving operation errors, drivers aged 24 or under and 75 or over are the groups who cause the highest number of accidents [11]. Additionally, in [12] "flustered/panicking" is the most factor common for all operation errors.

In order to prevent the source of operation errors, many studies have attempted to detect inattention [13] [14] [15], cognitive distraction [16] [17] or driver emotion [4] [18] [19] [20] [21]. As for the pedal misapplication due to the surprise event, the previous studies focused on the foot behavior [22], foot placement [23] or interruption of the other task [24]. There is limited knowledge of what happens with the internal states and what should be improved when countermeasures fail to prevent the error. In this study, focusing on a specific unforeseen event, human-machine interaction, and pedal operation, we proposed a multimodal model to detect the driver's state in normal driving and surprise driving caused by an unexpected situation. By using data collected from the cardiovascular system and driving responses, the proposed model could lead to a better understanding of the interaction between those components. The results of this study can be used in the future development of assistance systems for vehicles.

## 2. Literature Review

### 2.1. Related Research

Previous studies have investigated the relationship between surprising situations

and driving performance. When investigating pedal misapplications, Schmidt *et al.* [25] mentioned that other operation errors might cause unexpected accidents, including human error (going in the wrong direction) and using the wrong gear (either reverse or drive) when starting to drive. Upon analyzing a large number of data sets, Green [26] pointed out that, in unexpected and surprising events, the human perception-brake reaction time increases significantly compared to in fully aware situations (0.7 - 0.75 seconds in fully aware situations, 1.25 seconds in common ones and 1.50 seconds in surprise events). Fitch *et al.* [27] published research about braking performance and surprise events in 2010. The results also agreed with Green's conclusion that surprised driver responses are slower than those of an aware driver, but these performances vary depending on other factors such as age, gender, vehicle, etc. B. Freund *et al.* [28] suggests that a decrease in cognitive capability may be an important contributor to the pedal error of an older driver. In the report of K. Lococo *et al.* [10], drivers between the ages of 70 and 74 are 1.8 times more likely to be involved in an accident due to pedal error than any other vehicle accident, and between the ages of 80 and 84 the number is four times more.

## 2.2. Driver State Assessment

To prevent unsafe situations, many studies have focused on detecting driving situations and the driver's state. The approaches come from many fields, including mechanical engineering, those relating to driving dynamics and driver performance; medical and bioengineering; computer science, using image recognition; etc. The methods used by the above approaches are mainly based on the assumption that the parameters inspected will show different trends or patterns in normal and abnormal driving conditions. This assumption was applied to steering behavior as an index of workload by Boer [29]. As for bio-signals, Choi *et al.* [30] used a principle dynamic model to predict the activation level of two autonomic branches (sympathetic and parasympathetic) for emotional stress. Gao *et al.* [4] used a camera to obtain facial information and input the data into supervised learning for emotion detection. Meanwhile, Solovey *et al.* [31] combined both driving performance data and bio-signals and applied different machine learning methods for real-time detection.

Using physiological information to detect human states has been attempted for many years. Various types of experiments have been conducted to ensure the safety of drivers [13] [14]. To detect the driver's state, sensors (electrodes, light-based sensors) and cameras are widely used because they are easy to set up. Other objective methods include output performance, which will evaluate the result of the task. Subjective methods use questionnaires or rating scales in between tasks or after the experiment. In many studies, individual measures can be used alone or mixed features can be used to improve the result.

The advantage of using sensors and cameras is that they can continuously monitor the behavior of the subject, which is essential in real-time applications. These measures assume that in normal driving conditions human responses stay

in stable patterns or distributions. Changes in physiological response are often acquired from the medical or human factor fields. Changes in performance come from preset experiments or long-term research. Commonly, physiological indexes are related to human sensing resources: vision, auditory, space, and movement.

The main advantage of subjective methods is that they are simple and easy to deploy in an experiment. Some of the famous techniques that are widely used are NASA-Task Load Index (NASA-TLX) [32] and the instantaneous self-assessment workload scale (ISA) [33]. The disadvantage of these methods is that the answers heavily rely on the subject's memory. Furthermore, self-report measures cannot be obtained continuously like the above-mentioned measures. However, they are easy to complete before or after an experiment.

### 2.3. Graphical Models

As mentioned in the previous section, studies on the relationship between physiological data and driving behavior are mostly based on statistical analysis and assumptions. This study will adopt a different approach by applying a Gaussian graphical model (GGM) for the detection of drivers' states and inspecting the potential relationships of the components in the human factors.

Graphical models are popular in the fields of chemistry, genomics, neurology, psychology, and social interaction [34] [35] [36]. They provide a principled approach for dealing with uncertainty problems through the use of probability theory and an effective approach for handling complexity through the use of graph theory. A graphical model is a family of probability distribution defined in terms of a directed graph (Bayesian network) or an undirected graph (Markov network). The components of a graphical model are nodes and edges. Nodes are represented by random variables, which are the investigated features. Influences between one node and the others are represented as edges. Edges can be directed or undirected. Among graphical models, the Gaussian graphical model (GGM) is famous for its exploratory data analysis and can be used for both cross-sectional and time-series data.

Let  $X$  be a set of  $p$ -dimensional Gaussian random vectors defined as:

$$X = \{X_1, \dots, X_p\} \sim N(0, \Sigma) \quad (1)$$

where all the variables are centered and normally distributed (or have been standardized to have a mean of 0) and have the variance-covariance matrix  $\Sigma$ .

By preparing the variables satisfied (1), the inverse of  $\Sigma$ , the precision matrix  $K$ ,  $K = \Sigma^{-1}$ , is focused. The precision matrix can be standardized; the partial correlation coefficients of two variables can be encoded by:

$$\text{cor}(X_i, X_j | X_{-(i,j)}) = \frac{-K_{ij}}{\sqrt{K_{ii}} \sqrt{K_{jj}}} \quad (2)$$

where  $K_{ij}$  denotes an element of  $K$  and  $X_{-(i,j)}$  denotes a set of variables without  $i$  and  $j$  elements.

When creating a GGM as a network (a partial correlation network), each variable  $X_i$  is represented as a node and a partial correlation between two variables is represented as an edge. Typically, positive partial correlations are visualized as blue or green edges, and negative partial correlations are red edges. When there is no relation (the partial correlation is zero), no edge is drawn.

### 3. Method

#### 3.1. Participants

As mentioned in Section 1, elderly citizens (aged 75 or over) are among the group that causes the highest number of accidents due to operation errors. All 35 subjects (20 males and 15 females) who participated in the study, with ages from 65 to 85 years old (mean age = 74.3), had valid driving licenses. The subjects were informed that they must give their consent before they participated in the study. The study protocol was approved by Nagoya University's Institute of Innovation for Future Society Ethical Review Board.

The participants were asked to drive while wearing bio-sensors to collect physiological data. Due to the noise in the data (bad contact or loose electrodes) or data corruption, some of the data had to be marked as unusable. The recorder sometimes lost its time system, which led to us being unable to merge the timing between the physiological data and the driving data. Those data were marked as not merged with the driving simulation data. For those above reasons, only 8 subjects' data were used.

#### 3.2. Apparatus

Due to the high risk of accidents in surprising situations, the experiments were conducted in an advanced simulation room in the NIC Building, Institute of Innovation for Future Society, Nagoya University. The simulator was a 5-screen 4K projector with a stereoscopic-view driving simulator that incorporates numerous elements, including driving simulation, traffic simulation, and vehicle dynamics and performance, by building upon the UC-win/Road software (FORUM8 Co. Ltd.). The system was optimized to take into account human perceptions and traits by incorporating complex mathematical models, high-luminance and high-definition visual cues, realistic cockpit modules, and a highly responsive motion platform. The logging function of UC-win/Road recorded the driver's operation and the vehicle's dynamic values.

The human-machine interface (HMI) system alerted the driver about the danger distance between the vehicle and the obstacle, which here was a building. The system consisted of a screen put on the instrumental dashboard in front of the driver, a speaker put under the driver seat, and a vibration motor put under the brake pedal. Upon receiving a warning signal about the distance, the HMI system would issue one of a combination of 4 types of warning—a message on the display, a high-pitched warning sound, a human warning voice, and a vibration—named pattern 0 (P0) to pattern 6 (P6).

To provide additional information besides driving information, monitoring cameras were placed inside the driver cockpit: one for monitoring the driver and one for monitoring the driver's foot movement. The recording videos were synchronized with the UC-win/Road time system.

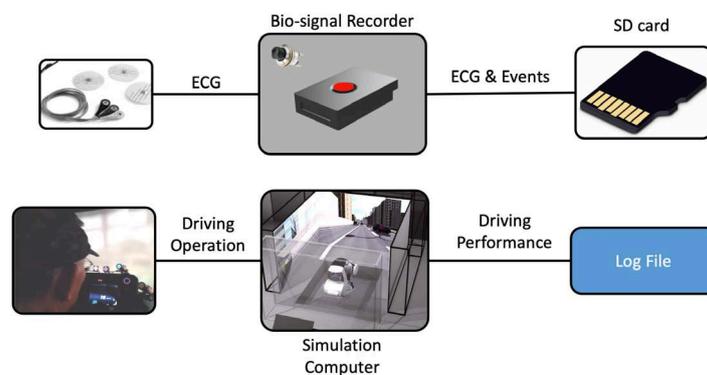
Physiological data were recorded by a portable recorder which has the same function with Livo TM4488 (Livo) (Toyota Technical Development Corporation, TTDC, Japan). Livo is a biomedical signal recording system that can record physiological activities such as electromyography (EMG) and electrocardiogram (ECG). ECG recordings used a lead II configuration at a sample rate of 1000 Hz. Depending on the subject's medical history, isopropyl alcohol or non-alcohol cleaner was used to clean the skin and standard pre-gelled disposable electrodes (Ag/AgCl paste, Vitrode Bs-150) were applied. The recorder did not have an internal real-time clock, so the time system was synchronized with the UC-win/Road time system through the wireless network. The bio-signal was recorded continuously without interruption, and the experiment events were marked by a button event operated by a monitoring operator seating behind the driver seat. The data acquisition system diagram is showed in **Figure 1**.

### 3.3. Experimental Procedure

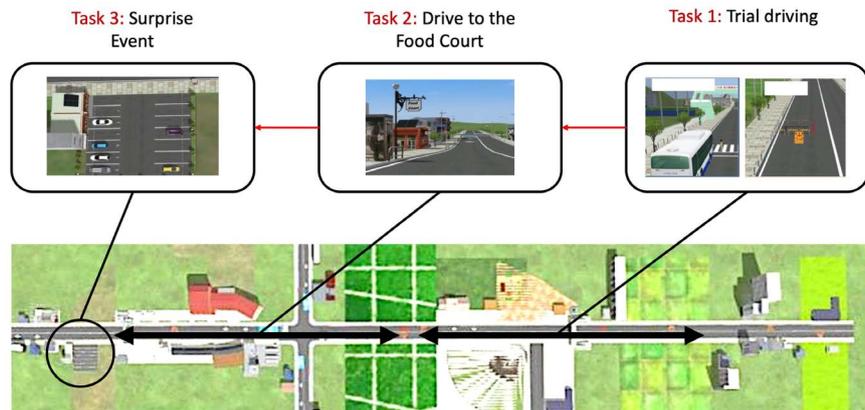
The tasks were designed to evaluate various driver response features to the surprising situation and alarm sources. The experimental detail is showed in **Figure 2**.

The first task was a trial drive which allowed the subjects to become familiar with the driving environment. In this task, the driver would drive through a straight road and pass a bus stop in the same lane, which required the driver to slow down and change lane. Right before passing the bus, there was a zebra crossing and a person tried to pass through, which required the subject to stop the vehicle. This setup helped the subjects to become used to the feeling of driving with the simulator system. Due to the nature of the instruction task and the subject's nervousness, the data variation of the first task was large and excluded in the analysis of this research.

The second task was to drive through an intersection with traffic control and then drive into a parking lot in front of a food court.



**Figure 1.** Data acquisition system.



**Figure 2.** Driving scenario and Tasks.

The third task was the main focus of this experiment, which was to create a surprising scenario. At the parking lot, the subject was asked to move out; the gear shift was intentionally reversed by the operator, causing the vehicle to move toward the food court instead of moving backward. The second task and third task were carried out continuously without a break.

The subjects were informed about the overall objective of the experiment but were not told when the surprise event would occur. The task sequence was carried out in an orderly fashion by the operator. To prevent any negative influence on the driving state, the subjects were asked to take 5 minutes resting before starting the test. The subjects were asked to drive in their normal driving style. The normal driving state was considered to be all the data collected throughout the normal driving situations, including the second task (driving along the street) and part of the third task (before the surprise event). The surprise state was considered to be the data recorded in the latter part of the third task when the subject reacted to the unexpected movement. Because the subjects expected to move backward but instead moved forward toward the store in front of where they were parked. Some drivers could realize the situation, release the gas pedal and press the brake pedal in time; others could not realize or could not react fast enough and hit the wall in front of the vehicle. This setup was considered a surprising event.

Besides the human factors and driving performance, the effectiveness of the alert source in the last task (reverse gear) was taken into consideration to improve the driver's reaction. To keep the surprise feeling intact, each subject only experimented once with one of the six alert patterns. The subjects were not informed about the alert pattern before it happened and were asked about their awareness of the alert after the test. For other details of the experimental setup, refer to [37].

### 3.4. Physiological Measures

Among the features most commonly used to explore the human state, the cardiovascular system and related features have been used for a long time. The res-

ponses of the cardiovascular system are controlled by the autonomic nervous system (ANS). The sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) both have influences on the heart rate. SNS increases the heart rate by increasing the firing rate of pacemaker cells, while the PNS decreases it through the influence of the vagal nerve and is known as rest and digest. The increase in, decrease in, and trend of heart rate are mostly affected by physical activities. Recent studies [38] [39] show that the heart rate and heart rate variability (HRV) can be used to detect and predict different human states.

The 3-lead ECG data went through a preprocessing procedure, including a noise filter and the extraction of the time elapsed between two successive R waves of the QRS complex on the electrocardiogram (RR interval). Then, the processed data were divided into windows (10 seconds, 30 seconds, 60 seconds). Since the window sizes were all under 5 minutes, they were considered to be an ultra-short analysis of HRV. According to recent research on ultra-short HRV analysis [40], the features which are used for investigation for an extremely short period of RR series are the mean RR, square root of the mean squared differences between successive RR intervals (RMSSD), low-frequency power (LF), high-frequency power (HF), and standard deviation of the Poincaré plot perpendicular to (SD1) and along (SD2). The meaningfulness of each feature depends on the nature of the statistic index they are based on. As for this result, to ensure the integrity of the study only mean RR, RMSSD, SD1, and SD2 were used as input data for the investigation.

### 3.5. Driving Response Measures

Researchers have used various driving performance measures and criteria for evaluating drivers' performance and state. Engström *et al.* [41] have released a comprehensive report on driving performance assessment including the use of mean speed, lateral position variation, time headway, brake reaction time, steering wheel reversal rate, etc.

In the scope of this study, the driving scenario mostly focused on the longitudinal dynamics of the vehicle, the driver's reaction to the driving scenes, and the safety evaluation of the outcome. For this reason, only the reaction time during the transition between the acceleration pedal and the brake pedal was considered.

Three types of alerts were inspected in this study: no alert situation (no alert source, called pattern 0, P0), alert with display and voice (called pattern 3, P3), and alert with display and integrated alarm (called pattern 6, P6). To ensure the surprise condition of the experiment, each subject only received one type of alert one time. In total, two subjects had P0, three subjects had P3, and three subjects had P6. The experiment period for the driving state and surprise state for each subject was about 2 to 3 minutes for each state. Since we did not know which type is more effective for the driver's state, this feature was encoded as a one-hot feature. The alert type was considered as normal input. During normal driving, the alert input was encoded as no alert. During the surprise events, the alert in-

puts were encoded respectively to their alert patterns.

As mentioned above, the reaction time (RT) was a feature that was investigated in this study. Due to the discrete nature of the pedal applied, the reaction time could only be calculated in a specific period. In the driving task, some drivers used the brake pedal more often, and others just released the gas pedal to slow down and only fully stop in case of stopping. Thus, for a short time (10 and 30-second window) some of the extracted data had no reaction operation. Consequently, the representative RT feature was extracted and used only in a 60-second window analysis.

Distinguishing between two driving states is one of the main concerns of this study; for normal driving without any negative influence and the surprise state when driving, details are described in the previous section. The labeling of the drivers' state was confirmed by carrying out a questionnaire after the experiment. The subjects were instructed to provide accurate answers about whether they were aware of the alert pattern and whether they were surprised by the situation. Besides the questionnaire, the drivers' reactions were also double-checked by the video recorded inside the driving cabin.

### 3.6. Detection Method Based on a Graphical Model

In this study, GGM and a Logistic Regression Classifier were combined in sequence to form a detection model. Before adding them to the model, all inputs (features) had to be standardized. In the learning phase, the driving data were used to estimate Driving GGM, and the surprise data were used to estimate Surprise GGM. The labeled inputs were grouped, and we used QuickGraphicLasso from “skggm: Gaussian graphical models using the scikit-learn API” [42] to estimate the respective GGM models.

The original aim of the model was to combine HRV indexes, reaction time, and alert patterns to classify the driver's state and use the structures acquired from GGM to explain the relationship between these inputs with structural changes in the models. For this reason, only 60-second windows with the reaction time data were used. The model inputs include meanRR, RMSSD, SD1, SD2, alert type, and reaction time. As the number of input data was small, the whole data set was used for the learning phase.

The trial data  $Z = \{Z_1, \dots, Z_p\}$ —the sampled data contained  $p$  features  $\times$   $m$ , where  $m$  is the length of time—is assumed to belong to one of the GGM models discussed above. The score function is the log-likelihood value of the covariance of the Gaussian trial data with the covariance of the estimate graph model. In the proposed model, assuming that the graph models are correctly estimated, the log-likelihood value is used to represent how “likely” it is that the test data belong to one of the graph models. The training trial data are extracted by a segmentation data set, as shown in **Figure 3**. Since the data set was small, synthetic test trial data were created to test the performance of the detection method. The test trial data were randomly selected from the data set by the group labels “Drive data” and “Surprise data”. The test trial data were then labeled according

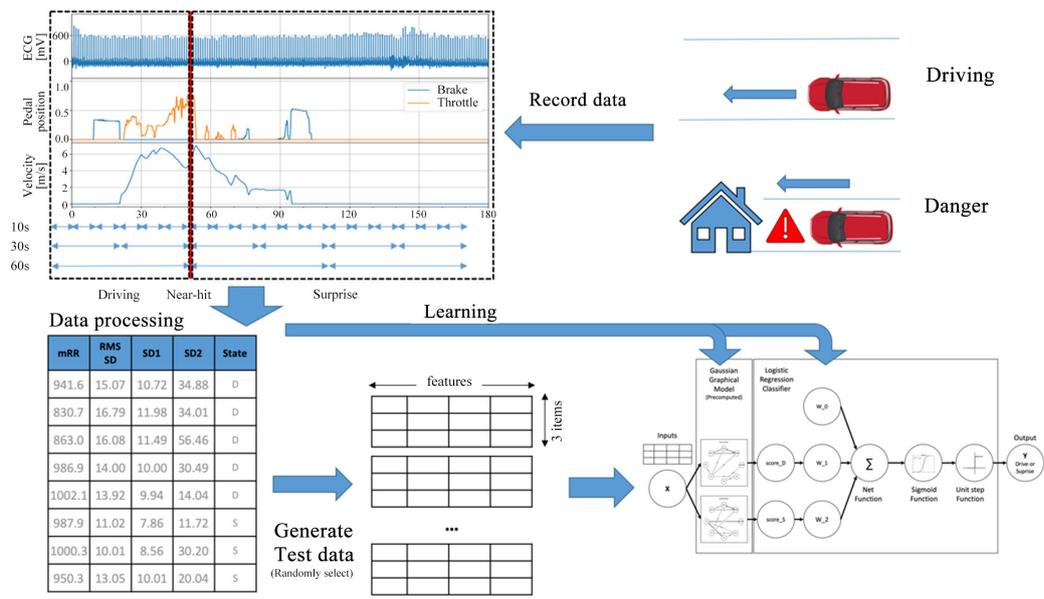
to their group.

Later, the score values obtained by the precomputed GGMs were used as the input of the Logistic Regression Classifier. The Logistic Regression Classifier (weights of the classifier,  $w_i$  with  $i \in (0, 2)$ ) was trained by the train trial data. The detailed structure of the graphical detection method is shown in **Figure 4**.

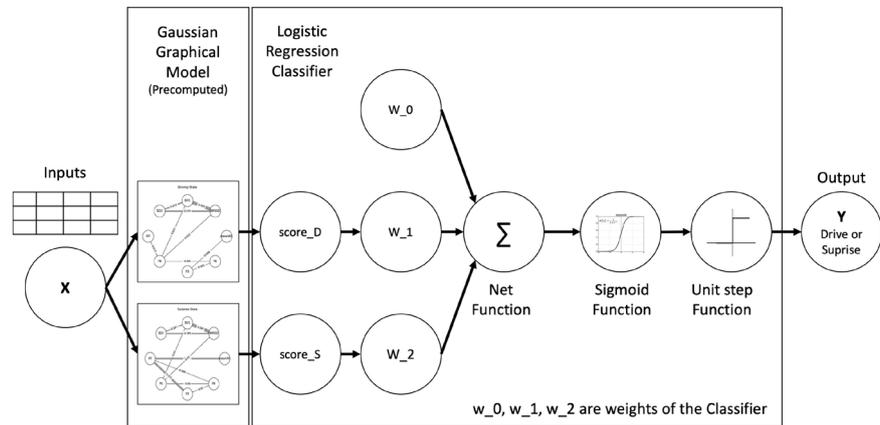
## 4. Results

### 4.1. Canonical Machine Learning Methods

As mentioned in Section 3.3, the data, including the simulation data and RR interval, were segmented into time windows of 10 seconds, 30 seconds, and 60 seconds to create three data sets. Details of the data collection and process are shown in **Figure 3**. The data set extracted with the 60-second time window was excluded in this section and only used with the proposed model due to there being such a small sample of data. The collected data set was then divided into a training set and a test set to evaluate the performance of the conventional machine learning models. Since the amount of the data was relatively small, the ratio 80:20 was used. Cross-validation was also used in the training session to prevent overfitting. Three canonical classification models were trailed in this study due to their popularity in small data sets: support vector machine (SVM), random forest (RF), and multilayer perceptron (MLP). The preprocessing, learning, prediction, and cross-validation were carried out using the APIs of scikit-learn [43] in Python. Due to the limitation of the collected data, the number of surprise state data was significantly lower than the number of drive data. The imbalance of the data will have affected the performance of the model. For that reason, the synthetic minority over-sampling technique (SMOTE) [44] was used to balance the number of two classes.



**Figure 3.** Data collection, training, and evaluating the proposed method.



**Figure 4.** Schematic of the proposed detection method.

Overall, the accuracy of RF was the best among the canonical methods (0.98 - 0.99 in the training set, 0.71 in the test set). The accuracy of SVM and MLP was barely better than that of a random guess (0.64 - 0.65 in the training set, 0.48 - 0.50 in the test set). Other indexes reflecting the effectiveness of the models are shown in [Table 1](#).

## 4.2. Graphical Model Method

The performance of the proposed graphical model was shown to be favorable in the case of the 60-second data set with RT information. The performance in the case of the 60-second data set without RT was still better than that of SVM and MLP. Furthermore, the two GGMs provide additional information about how driving performance interacts with the physiological indexes and alert type in each state.

## 5. Discussion

The basic statistic of the cardiovascular shows in [Table 2](#). It can be seen that the mean of heart rate in the surprise was lower than the mean of heart rate in normal driving. This result seems somewhat unexpected as typically one would assume that the heart rate should increase to response to a surprise event. It is true that most of the case the cardiac activities increase right after the stressful event then gradually decrease back to normal. Hence the mean heart rate of the surprise data might be lower than the mean heart rate of the normal driving. A previous study [45] also mentioned that cardiac acceleration or deceleration was due to individual differences. In another study [46] about the psychological responses to facial expression in which the image appears as a kind of sudden event, the heart rate decrease and startle reflexes increase indicated that the rising of attention response and preparation for fight-or-flight response of the SNS system.

Among the inspected machine learning methods, RF shows the highest accuracy in both the training set and test set. The gap between the training and test accuracy might be the result of overfitting. The cross-validation and averaging of

the model could reduce this phenomenon. The high precision shows that the RF model used in this study is good at detecting the surprise state. However, the low recall reflects that the RF model is bad at detecting the driving state and tends to mistake the driving state for the surprise state. This might be the effect of the synthetic data creation by SMOTE to balance the data. Although not reported in this study, we have tested those machine learning methods without SMOTE, the accuracies were better but the performance indexes (precision, recall, and F1-score) were worst. The limited sample size has significant effects on the performance of the machine learning methods, especially with MLP which we found out during the training model that it did not converge. It can be seen that the performance indexes of SVM and MLP decrease correlated with the decrease of sample size. Surprise in driving involves not only the cognitive function but also the sensorimotor function (bodily movement, steering, and leg movement). Using only physiological measures will not be as effective in detecting driver states under varying external conditions as the addition of other information on driver behavior and the driving context. Similar results have been found by Solovey *et al.* [31] and McDonald *et al.* [47], and the performance of machine learning methods is lower when using only physiological features as inputs.

**Table 1.** Performance of support vector machine (SVM), random forest (RF), multilayer perceptron (MLP), and the proposed model.

Model	Window size (secs)	Number of Samples	Accuracy on train set	Accuracy on test set	Precision	Recall	F1-score
SVM	10	278	0.65	0.48	0.58	0.54	0.56
	30	95	0.65	0.50	0.50	0.43	0.46
RF	10	278	0.99	0.71	0.85	0.63	0.72
	30	95	0.98	0.71	0.88	0.50	0.64
MLP	10	278	0.64	0.48	0.58	0.52	0.55
	30	95	0.64	0.50	0.50	0.36	0.42
GGM-L	10	278	0.73	0.57	0.77	0.21	0.32
	30	95	0.67	0.59	1.00	0.18	0.31
	60 <sup>1</sup>	48	0.80	0.66	0.77	0.47	0.58
	60 <sup>2</sup>	48	0.94	0.94	0.94	0.97	0.96

<sup>1</sup>Without reaction time information; <sup>2</sup>with reaction time information.

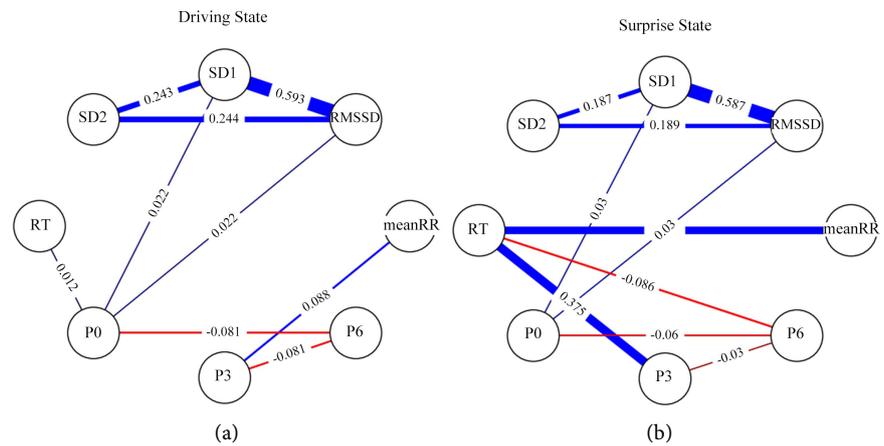
**Table 2.** Mean and standard deviation (SD) of the physiological variables and reaction time in each state.

Variable	Normal driving state		Surprise state		Unit
	Mean	SD	Mean	SD	
RR interval	911.3	64.5	934.7	55.2	ms
Heart Rate	66.3	4.5	64.6	3.8	bpm
Reaction Time	1792.4	286.3	463.8	129	ms

The result of the proposed model can be explained by the evaluation method: the score value of the graphical model is the log-likelihood of the given test data for the parameter values of the estimated model. This means that it calculates the probability of a set of data to decide whether it fits with the model or not. Generally, if this value is high the test data has a high probability of belonging to that graphical model. The technique of creating test data is similar to the bootstrap technique, which means that the data which were used in the learning phase will be used in the evaluation phase but in random order. Despite the fact of small sample data, our proposed model yield the best performance in detecting driver state compared to the other machine learning methods. In reality, the number of usable data points for human behavior in driving is limited. The performance of the model with a small data set is an advantage. Overcome limitation of sample data will be considered in our future work.

One of the most important points when approaching graphical models is inference. The two graphical models (GGM for driving state and GGM for surprise state) included in the learning phase have similar features. Because the conditions for each model are controlled and different, we acquired two models with different structures. The changes in structure will help us to gain a better understanding of the interaction between the models' features. **Figure 5** shows the results of the two models estimated from the data. Clear changes in the structure of those two models can be seen. In the case of the drive state, reaction time has almost no relationship with the other features.

The partial correlation with P0 (no alert) is small, which is understandable because there is no alert source to affect other human factors. The change in the model structure from the driving state to the surprise state is noticeable. While the partial correlation between the pair "RT - mean RR" and "RT - P3" is positive, the partial correlation of the pair "RT - P6" is negative. There is little change in the partial correlation between the nodes SD1, SD2, and RMSSD and the other nodes. These results can be interpreted as meaning that the reaction time and the inner system have less influence in normal driving, though in a surprising situation they affect each other closely. Furthermore, the three features SD1, SD2, and RMSSD have no interaction with the other features, and they can be omitted in similar studies in the future. The interesting point here is that the alert type also affects the reaction time but in an opposite way. P3 and P6 use a similar alert source—auditory (voice in P3 and alarm in P6) and visual (display in both patterns)—but P3 tends to have a positive influence on RT while P6 tends to have a negative effect. This can be interpreted as the alarm (in P6) reducing the reaction time, which means that it increases the driving performance. Meanwhile, the voice (in P3) tends to prolong the reaction time. It can be inferred that the human sensing system and brain respond in different ways to the alert sources: when a human hears a sound, if it is a human voice the perception process needs time to interpret the information and make a decision. Consequently, the human/dialogue sound appears to be effective in the case of giving instructions, such as in take-overs [48]. On the other hand, the response



**Figure 5.** Graphical structures for (a) driving state and (b) surprise state.

to the alarm takes less time; the human brain can automatically acknowledge the serious problem and take action in a shorter period [49] [50]. Although not reported in this study, the questionnaire result also showed that the alarm sound was more effective than the other warning sources. As for the display, the size is important due to the visual problems of old people.

There are still some limitations in this study. The first one is the number of samples. The actual number of subjects involved was higher, but only eight ECG data were used for this study. The second limitation is the assumption in the GGM model—the multivariate normality of the data set—which means that the features need to carry out standard scaling.

## 6. Conclusion

This study investigated different classification methods for a specific case: normal driving and surprise driving. The accuracy and effectiveness of three common classification methods have been investigated, along with the effect of different window sizes on the outcome of the prediction. The graphical-based detection model shows a good potential in both the prediction performance and exploratory attribute. Although there are some limitations regarding the amount of data gathered, the results showed a conclusion that was consistent with those of previous studies. The favorable outcome shows that this approach can be applied not only to this problem but also to other human behavior studies that inspect the internal interaction of system components. In the future, our focus will be on extending the number of samples collected and integrating other features to gain a better understanding of the interactions not only in surprising situations but also in other critical driving scenarios.

## Acknowledgements

The authors would like to thank Mr. Sho Tsujita, Ms. Ami Azuma, Mr. Tomoki Ikoma (DENSO CORPORATION, Japan) for their contribution in the experiment, the technical support division for preparing the simulation room and the

assistant team for taking care of the elder subjects during the experiment. This research was supported by JST-OPERA Program Grant Number JPMJOP1612, Japan.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] World Health Organization (2020) Road Traffic Injuries. World Health Organization, Geneva.
- [2] Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D. and Ramsey, D.J. (2006) The Impact of Driver Inattention on Near-Crash/Crask Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. DOT HS 810594, National Highway Traffic Safety Administration, Washington DC.
- [3] Chan, M. and Singhal, A. (2013) The Emotional Side of Cognitive Distraction: Implications for Road Safety. *Accident Analysis and Prevention*, **50**, 147-154. <https://doi.org/10.1016/j.aap.2012.04.004>
- [4] Gao, H., Yuce, A. and Thiran, J.-P. (2014) Detecting Emotional Stress from Facial Expressions for Driving Safety. 2014 *IEEE International Conference on Image Processing*, Paris, 27-30 October 2014, 5961-5965. <https://doi.org/10.1109/ICIP.2014.7026203>
- [5] Chan, M. and Singhal, A. (2015) Emotion Matters: Implications for Distracted Driving. *Safety Science*, **72**, 302-309. <https://doi.org/10.1016/j.ssci.2014.10.002>
- [6] Scott-Parker, B. (2017) Emotions, Behaviour, and the Adolescent Driver: A Literature Review. *Transportation Research Part F: Traffic Psychology and Behaviour*, **50**, 1-37. <https://doi.org/10.1016/j.trf.2017.06.019>
- [7] Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F. and Yeh, M. (2014) Startle and Surprise on the Flight Deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **58**, 1047-1051. <https://doi.org/10.1177%2F1541931214581219>
- [8] Landman, A., Groen, E. L., van Paassen, M. M. (René), Bronkhorst, A.W. and Mulder, M. (2017) Dealing with Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Human Factors*, **59**, 1161-1172. <https://doi.org/10.1177%2F0018720817723428>
- [9] de Boer, R.J. and Hurts, K. (2017) Automation Surprise: Results of a Field Survey of Dutch Pilots. *Aviation Psychology and Applied Human Factors*, **7**, 28-41. <https://doi.org/10.1027/2192-0923/a000113>
- [10] Lococo, K.H., Staplin, L., Martell, C.A. and Sifrit, K.J. (2012) Pedal Application Errors. United States. Department of Transportation. DOT HS 811597, National Highway Traffic Safety Administration, Washington DC.
- [11] Masahide, H. (2014) Preventing Driving Operation Errors. No. 107, Institute for Traffic Accident Research and Data Analysis, Tokyo.
- [12] Akihiro, H. (2018) Accidents Due to the Misapplication of Accelerator and Brake Pedals. No. 124, Institute for Traffic Accident Research and Data Analysis, Tokyo.
- [13] D’Orazio, T., Leo, M., Guaragnella, C. and Distanto, A. (2007) A Visual Approach for Driver Inattention Detection. *Pattern Recognition*, **40**, 2341-2355.

- <https://doi.org/10.1016/j.patcog.2007.01.018>
- [14] Fletcher, L. and Zelinsky, A. (2009) Driver Inattention Detection Based on Eye Gaze—Road Event Correlation. *The International Journal of Robotics Research*, **28**, 774-801. <https://doi.org/10.1177%2F0278364908099459>
- [15] Bergasa, L.M., Almeria, D., Almazan, J., Yebes, J.J. and Arroyo, R. (2014) DriveSafe: An App for Alerting Inattentive Drivers and Scoring Driving Behaviors. 2014 *IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, 8-11 June 2014, 240-245. <https://doi.org/10.1109/IVS.2014.6856461>
- [16] Liao, Y., Li, S.E., Wang, W., Wang, Y., Li, G. and Cheng, B. (2016) Detection of Driver Cognitive Distraction: A Comparison Study of Stop-Controlled Intersection and Speed-Limited Highway. *IEEE Transactions on Intelligent Transportation Systems*, **17**, 1628-1637. <https://doi.org/10.1109/TITS.2015.2506602>
- [17] Gershon, P., Sita, K.R., Zhu, C., Ehsani, J.P., Klauer, S.G., Dingus, T.A. and Simons-Morton, B.G. (2019) Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers. *American Journal of Preventive Medicine*, **56**, 494-500. <https://doi.org/10.1016/j.amepre.2018.11.024>
- [18] Lee, K.W., Yoon, H.S., Song, J.M. and Park, K.R. (2018) Convolutional Neural Network-Based Classification of Driver's Emotion during Aggressive and Smooth Driving Using Multi-Modal Camera Sensors. *Sensors*, **18**, 957. <https://doi.org/10.3390/s18040957>
- [19] Grimm, M., Kroschel, K., Harris, H., Nass, C., Schuller, B., Rigoll, G. and Moosmayr, T. (2007) On the Necessity and Feasibility of Detecting a Driver's Emotional State While Driving. International Conference on Affective Computing and Intelligent Interaction, Lisbon, 12-14 September 2007, 126-138. [https://doi.org/10.1007/978-3-540-74889-2\\_12](https://doi.org/10.1007/978-3-540-74889-2_12)
- [20] Rebolledo-Mendez, G., Reyes, A., Paszkowicz, S., Land, J., Limited, R., Reyes, A., Domingo, M.C. and Skrypchuk, L. (2014) Developing a Body Sensor Network to Detect Emotions During Driving. *IEEE Transactions on Intelligent Transportation Systems*, **15**, 1850-1854. <https://doi.org/10.1109/TITS.2014.2335151>
- [21] Verma, B. and Choudhary, A. (2018) A Framework for Driver Emotion Recognition Using Deep Learning and Grassmann Manifolds. 2018 21st International Conference on Intelligent Transportation Systems, Maui, 4-7 November 2018, 1421-1426. <https://doi.org/10.1109/ITSC.2018.8569461>
- [22] Wu, Y., Boyle, L.N. and McGehee, D. (2018) Evaluating Variability in Foot to Pedal Movements Using Functional Principal Components Analysis. *Accident Analysis and Prevention*, **118**, 146-153. <https://doi.org/10.1016/j.aap.2018.02.011>
- [23] Wu, Y., Boyle, L.N., McGehee, D., Roe, C.A., Ebe, K. and Foley, J. (2017) Foot Placement during Error and Pedal Applications in Naturalistic Driving. *Accident Analysis and Prevention*, **99**, 102-109. <https://doi.org/10.1016/j.aap.2016.10.019>
- [24] Hasegawa, K., Kimura, M. and Takeda, Y. (2020) Pedal Misapplication: Interruption Effects and Age-Related Differences. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082093612. <https://doi.org/10.1177%2F0018720820936122>
- [25] Schmidt, R.A., Young, D.E., Ayres, T.J. and Wong, J.R. (1997) Pedal Misapplications: Their Frequency and Variety Revealed through Police Accident Reports. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **41**, 1023-1027. <https://doi.org/10.1177%2F107118139704100266>
- [26] Green, M. (2000) "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. *Transportation Human Factors*, **2**, 195-216.

- [https://doi.org/10.1207/STHF0203\\_1](https://doi.org/10.1207/STHF0203_1)
- [27] Fitch, G.M., Blanco, M., Morgan, J.F. and Wharton, A.E. (2012) Driver Braking Performance to Surprise and Expected Events. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **54**, 2075-2080. <https://doi.org/10.1177%2F154193121005402412>
- [28] Freund, B., Colgrove, L.A., Petrakos, D. and McLeod, R. (2008) In My Car the Brake Is on the Right: Pedal Errors among Older Drivers. *Accident Analysis & Prevention*, **40**, 403-409. <https://doi.org/10.1016/j.aap.2007.07.012>
- [29] Boer, E.R. (2001) Behavioral Entropy as a Measure of Driving Performance. *Proceedings of the 1st International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Colorado, 14-17 August 2001, 226-229. <https://doi.org/10.17077/drivingassessment.1043>
- [30] Choi, J. and Gutierrez-Osuna, R. (2009) Using Heart Rate Monitors to Detect Mental Stress. *2009 6th International Workshop on Wearable and Implantable Body Sensor Networks*, Berkeley, 3-5 June 2009, 219-223. <https://doi.org/10.1109/BSN.2009.13>
- [31] Solovey, E.T., Zec, M., Garcia Perez, E.A., Reimer, B. and Mehler, B. (2014) Classifying Driver Workload Using Physiological and Driving Performance Data. *Proceedings of the 32nd annual ACM conference on Human Factors in computing systems*, Toronto, April 2014, 4057-4066. <https://doi.org/10.1145/2556288.2557068>
- [32] Hart, S.G. (2006) NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society*, **50**, 904-908. <https://doi.org/10.1177%2F154193120605000909>
- [33] Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005) The Integrated Workload Scale (IWS): A New Self-Report Tool to Assess Railway Signaller Workload. *Applied Ergonomics*, **36**, 681-693. <https://doi.org/10.1016/j.apergo.2005.05.004>
- [34] Jordan, M.I. (2004) Graphical Models. *Statistical Science*, **19**, 140-155. <https://doi.org/10.1214/088342304000000026>
- [35] Dahlhaus, R. and Eichler, M. (2003) Causality and Graphical Models in Time Series Analysis. In: Green, P., Hjort, N. and Richardson, S., Eds., *Highly Structured Stochastic Systems*, Oxford University Press, Oxford, 115-137.
- [36] Epskamp, S., Waldorp, L.J., Möttus, R. and Borsboom, D. (2018) The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, **53**, 453-480. <https://doi.org/10.1080/00273171.2018.1454823>
- [37] Tsujita, S., Azuma, A., Ikoma, T., Aoki, K., Hirano, A., Son, L.A., Nguyen, T.T., Shimazaki, K. and Aoki, H. (2019) Effects of the HMI on Elderly Drivers' Behavior under a Simulated Unexpected Situation. *2019 JSAE Annual Congress*, Yokohama, 22-24 May 2019.
- [38] Vicente, J., Laguna, P., Bartra, A. and Bailón, R. (2016) Drowsiness Detection Using Heart Rate Variability. *Medical & Biological Engineering & Computing*, **54**, 927-937. <https://doi.org/10.1007/s11517-015-1448-7>
- [39] Castaldo, R., Montesinos, L., Melillo, P., James, C. and Pecchia, L. (2019) Ultra-Short Term HRV Features as Surrogates of Short Term HRV: A Case Study on Mental Stress Detection in Real Life. *BMC Medical Informatics and Decision Making*, **19**, Article No. 12. <https://doi.org/10.1186/s12911-019-0742-y>
- [40] Pecchia, L., Castaldo, R., Montesinos, L. and Melillo, P. (2018) Are Ultra-Short Heart Rate Variability Features Good Surrogates of Short-Term Ones? State-of-the-Art Review and Recommendations. *Healthcare Technology Letters*, **5**, 94-100.

- <https://doi.org/10.1049/htl.2017.0090>
- [41] Engström, J., Markkula, G., Keinath, A., Horst, D., Juch, S., Mattes, S., Foehl, U., Östlund, J., Peters, B. and Thorslund, B. (2005) Driving Performance Assessment—Methods and Metrics.
- [42] Laska, J. and Narayan, M. (2017) Skggm 0.2.7: A Scikit-Learn Compatible Package for Gaussian and Related Graphical Models. <https://doi.org/10.5281/zenodo.830033>
- [43] Pedregosa, F., Weiss, R., Brucher, M., Varoquaux, G., Gramfort, A., Michel, V., *et al.* (2011) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.
- [44] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2011) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. <https://doi.org/10.1613/jair.953>
- [45] Nomikos, M.S., Opton, E. and Averill, J.R. (1968) Surprise versus Suspense in the Production of Stress Reaction. *Journal of Personality and Social Psychology*, **8**, 204-208. <https://doi.apa.org/doi/10.1037/h0025274>
- [46] Gantiva, C., Araujo, A., Calderón, L., Gómez, P. and Reina, F. (2019) Psychophysiological Responses to Facial Expression of Surprise, Sadness, and Disgust. *Australian Journal of Psychology*, **71**, 100-107. <https://doi.org/10.1111/ajpy.12222>
- [47] McDonald, A.D., Ferris, T.K. and Wiener, T.A. (2020) Classification of Driver Distraction: A Comprehensive Analysis of Feature Generation, Machine Learning, and Input Measures. *Human Factors*, **62**, 1019-1035. <https://doi.org/10.1177%2F0018720819856454>
- [48] Forster, Y., Naujoks, F., Neukum, A. and Huestegge, L. (2017) Driver Compliance to Take-over Requests with Different Auditory Outputs in Conditional Automation. *Accident Analysis and Prevention*, **109**, 18-28. <https://doi.org/10.1016/j.aap.2017.09.019>
- [49] Kozak, K., Pohl, J., Birk, W., Greenberg, J., Artz, B., Blommer, M., Cathey, L. and Curry, R. (2006) Evaluation of Lane Departure Warnings for Drowsy Drivers. In *Proceedings of the Human Factors and Ergonomics Society*, **20**, 2400-2404. <https://doi.org/10.1177%2F154193120605002211>
- [50] Campbell, J.L., Richard, C.M., Brown, J.L. and McCallum, M. (2007) Crash Warning System Interfaces: Human Factors Insight and Lessons Learned. DOT HS-810, National Highway Traffic Safety Administration, Washington DC.