

YIFAN SUN, Ph.D.¹
E-mail: m18705423520@163.com
CHAOZHONG WU, Prof.¹
E-mail: wucz@whut.edu.cn
HUI ZHANG, Prof.¹
(Corresponding author)
E-mail: zhanghuiits@whut.edu.cn
WENHUI CHU, Ph.D.^{1,2}
E-mail: chuwh@whut.edu.cn
YIYING XIAO, M.Sc.¹
E-mail: shadowxyy@outlook.com
YIJUN ZHANG, M.Sc.¹
E-mail: zyjits@whut.edu.cn

¹ Intelligent Transportation Systems Research Centre
Wuhan University of Technology
1178 Heping Dadao, Wuchang District, Wuhan, Hubei
Province, China

² School of Automobile and Traffic Engineering
Wuhan University of Science and Technology
Special 1, Huangjiahu University Town, Hongshan
District, Wuhan, Hubei Province, China

Traffic Engineering
Original Scientific Paper
Submitted: 20 July 2020
Accepted: 11 Dec. 2020
DOI: 10.7307/ptt.v33i4.3668

EFFECTS OF INDIVIDUAL DIFFERENCES ON MEASUREMENTS' DROWSINESS-DETECTION PERFORMANCE

ABSTRACT

Individual differences (IDs) may reduce the detection-accuracy of drowsiness-driving by influencing measurements' drowsiness-detection performance (MDDP). The purpose of this paper is to propose a model that can quantify the effects of IDs on MDDP and find measurements with less impact by IDs to build drowsiness-detection models. Through field experiments, drivers' naturalistic driving data and subjective-drowsiness levels were collected, and drowsiness-related measurements were calculated using the double-layer sliding time window. In the model, MDDP was represented by $|Z\text{-statistics}|$ of the Wilcoxon-test. First, the individual driver's measurements were analysed by Wilcoxon-test. Next, drivers were combined in pairs, measurements of paired-driver combinations were analysed by Wilcoxon-test, and measurement's IDs of paired-driver combinations were calculated. Finally, linear regression was used to fit the measurements' IDs and changes of MDDP that equalled the individual driver's $|Z\text{-statistics}|$ minus the paired-driver combination's $|Z\text{-statistics}|$, and the slope's absolute value ($|k|$) indicated the effects of ID on the MDDP. As a result, $|k|$ of the mean of the percentage of eyelid closure (MPECL) is the lowest (4.95), which illustrates MPECL is the least affected by IDs. The results contribute to the measurement selection of drowsiness-detection models considering IDs.

KEYWORDS

traffic safety; drowsiness-detection models; non-intrusive measurements; naturalistic driving study; individual differences.

1. INTRODUCTION

Drowsiness driving is still a widespread traffic problem and always leads to serious injuries [1]. Studies pointed out that about 15–30% of all crashes were contributed by drowsiness driving [2]. According to the government report of America, drowsiness-related crashes caused approximately 1550 deaths and 12.5 billion dollars of losses each year [3]. Unlike drunk driving, due to the lack of objective criteria of drowsiness occurrence, many drowsiness-related accidents have not been reported, and the actual harm of drowsiness driving might be more serious. Thus, studies of drowsiness driving have always been a hotspot in traffic safety research.

Anti-drowsiness driving assistance systems can warn drivers when they are drowsy, which is considered to be an effective countermeasure for drowsiness-driving [4]. Anti-drowsiness driving assistance systems rely on accurate and reliable

drowsiness-detection models that have been widely studied [5, 6]. And according to the intrusiveness of data collection, drowsiness-detection models can be divided into two types: intrusive and non-intrusive [7]. Although intrusive models using data like electroencephalograms [8], electromyography [4] have good accuracy, these are only available in laboratory conditions [7]. As opposed to that, the non-intrusive models based on driving behaviour data [5, 9] or eye movement data [10] are more practical because the data collection process causes little disturbance to drivers and fewer restrictions [11]. Thus, non-intrusive drowsiness-detection models have become the interest of traffic safety research.

However, the accuracy and reliability of non-intrusive drowsiness-detection models are significantly affected by individual differences (IDs) [6]. IDs generally refer to differences among drivers in behavioural performance, physiological characteristics, and cognitive abilities, etc., which are crucial issues in current driving behaviour research [12-14]. To date, most non-intrusive drowsiness-detection models belong to generalised models that use the mixed measurement data of all drivers to train models and utilise the same drowsiness threshold to detect the drowsiness of each driver [1]. Many scholars have studied the mechanism of IDs affecting the drowsiness-identification of generalised models. Inger [15] pointed out that the drowsiness thresholds of generalised models were the average value of all drivers rather than the individual driver's threshold, and using the average threshold made the drowsiness-detection at the individual driver level suffer from systematic error. Similarly, Philip et al. [16] reported that there were major IDs in driving performance under drowsy state, and IDs might reduce the accuracy of drowsiness-identification based on average damage of drowsiness on driving performance. Yan [13] explored IDs in drowsiness-detection and found that IDs in measurement distribution among drivers could exceed differences caused by drowsiness, which decreased the correlation between measurements and drowsiness. These studies illustrated that it is necessary to consider IDs when establishing drowsiness-detection models.

Currently, the drowsiness-detection models considering IDs include two categories. One is the driver-specific drowsiness-detection model without IDs, which is trained using the driving behaviour data of the individual driver [7]. And the other is the generalised drowsiness-detection model consider-

ing IDs, in which measurements that contain fewer IDs or were less affected by IDs are preferentially selected to build models [17]. For driver-specific models, Wang [7] used 23 non-intrusive measurements of individual drivers to establish the driver-specific drowsiness-detection model based on the multilevel logit model. The results showed that the drowsiness-detection accuracy of the personalised model considering IDs was higher. Chu et al. [18] recorded driving behaviour data through field experiments, using individual drivers' data, and driver-specific drowsiness-detection models were established based on the RBF neural network and support vector machine. The results showed that driver-specific models could improve drowsiness-detection accuracy by eliminating IDs. You [19] concentrated on IDs at driver's eye level and used eye landmarks of individual drivers to train the driver-specific drowsiness-detection model based on support vector machines. The accuracy reached 94.8%, which outperformed the models neglecting IDs. The above studies demonstrated that driver-specific drowsiness-detection models are more accurate at the individual driver level, which is attributed to using the individual driver's unique drowsiness-detection criteria. However, driver-specific drowsiness-detection models rely on the training of the individual driver's data to obtain the driver's unique drowsiness-threshold, which cannot accurately detect a new driver's drowsiness. Consequently, the practicalities and generalisations of driver-specific models are weak and drowsiness-detection efficiencies are low.

To address the shortcomings of driver-specific drowsiness-detection models, some researchers explored the method to deal with IDs in generalised drowsiness-detection models. Generalised models considering IDs made a balance between reducing the impact of IDs and improving the practicalities of drowsiness-detection. On the one hand, IDs in measurement distribution are analysed, and measurements with small IDs are chosen to build models for reducing the impact of IDs on drowsiness-detection accuracy. On the other hand, the generalised model only needs to be trained once using all drivers' data, which does not need to respectively utilise the data of each driver to train multiple driver-specific models. All drivers can use the generalised model to detect drowsiness. Therefore, the practicalities and efficiencies of drowsiness-identification models are improved.

With regards to generalised drowsiness-detection models considering IDs, effects of IDs on the measurements' drowsiness-detection performance (MDDP) are the major factors influencing the accuracy of drowsiness-detection [13]. Thus, some scholars explored IDs in drowsiness measurements of generalised models [17, 20, 21]. The obvious IDs in the distribution of measurements, such as the standard deviation of steering wheel movement [12], percentage of eyelid closure (PERCLOS), gaze time ratio [19], steering reversal rate (SRR), and standard deviation of lane position (SDLP) [21] have been frequently mentioned in previous studies. And scholars have compared the magnitude of IDs in various measurements. Xu, et al. [17] calculated 23 non-intrusive measurements using simulated driving experimental data and calculated IDs of measurements based on the Kruskal-Wallis test. It was found that IDs of PERCLOS and measurements derived from the steering wheel angle were higher and lower, respectively. Furthermore, the researchers analysed the impact of IDs on the MDDP. Niu [20] used the F-statistics of variance analysis to represent the MDDP and studied changes of the MDDP with IDs by comparing the individual driver's F-statistics with the F-statistics of the paired driver combinations. The results indicated that IDs weakened the MDDP. Zhang [21] calculated the Pearson correlation coefficients of SDLP, SRR, and drowsiness respectively for each participant and pointed out that IDs reduced the correlation between measurements and drowsiness levels. The above results could provide significant basis for developing generalised drowsiness-detection considering IDs.

Previous studies proved that IDs of measurements existed and found that the IDs of various measurements were different. However, the quantitative analysis on the effect of IDs on MDDP has not been reported in current studies, which is crucial for the generalised drowsiness-detection models to choose measurements less affected by IDs. In previous generalised drowsiness-detection models without considering the effect of IDs on the MDDP, the drowsiness measurements whose drowsiness-detection abilities were greatly affected by IDs were chosen to train models, which caused the IDs to sharply reduce the drowsiness-detection accuracy. The prerequisite of improving the drowsiness-detection accuracies of the models is to select measurements that can distinguish awake state and fatigue state effectively, that is, the MDDP is ideal [22]. Therefore,

it is significant to analyse the effects of IDs on the MDDP when choosing measurements to build generalised drowsiness-detection models.

In the present paper, a new model was established to quantify the effect of IDs on the MDDP for finding measurements with fewer effects by IDs. The significance of the results is to provide a basis for generalised drowsiness-detection models to select measurements that are less affected by IDs. As a result, we can weaken the effects of IDs on the drowsiness-detection accuracy of generalised models and ensure the accuracy and efficiency of drowsiness-detection. In this paper, first the drivers' subjective-drowsiness levels and multi-source sensor data were collected through naturalistic driving experiments on the motorway. Then, a double-layer sliding time window was used to compute nine non-intrusive measurements. Next, the MDDP was evaluated by $|Z\text{-statistics}|$ of the Wilcoxon-test and IDs of measurements were calculated. Then, changes of the MDDP with respect to IDs were analysed by linear regression, and the model to quantitatively analyse the effect of IDs on MDDP was built. Finally, the effect of IDs on MDDP was quantified and verified. According to the results, measurements whose drowsiness-detection performance is less affected by IDs could be selected to establish generalised drowsiness-detection models, which could ensure the average drowsiness-detection accuracy and improve practicalities of generalised drowsiness-detection models.

The remainder of this paper is organised as follows. The experimental details of data collection were presented in section 2. The formulas of selected measurements, data pre-processing, and the models to quantify the effects of IDs on the MDDP were introduced in Section 3. In Section 4, the quantitative effects of IDs on the MDDP, its verification analysis, and other relative results were displayed. Finally, this is followed by the discussion and conclusions in Section 5 and 6, respectively.

2. EXPERIMENTS

2.1 Participants

Forty professional drivers (including 5 female drivers) with proficient driving skills were recruited and numbered. The ages of drivers ranged from 34 years to 57 years (mean=46.83, SD=5.62). The years of driving experience ranged from 3 years to 32 years (mean=16.53, SD=6.10). They had good

driving skills and stable driving habits. The participants were in a good psychological and physiological state, with normal work and rest, and did not take any drugs for 3 weeks before experiments. Every participant signed the informed consent and was told the possible experimental risks and details before the experiments. Each experiment was accompanied by a safety officer who was an instructor with 30 years of driving experience. After training, the safety officer was responsible for inquiring and recording the subjective-drowsiness level of participants using the Karolinska Sleepiness Scale (KSS) and taking emergency measures to ensure safety when necessary.

2.2 Apparatus

Participants drove a real vehicle that was integrated with on-board sensors including the three-way HD camera, the Mobileye, the Driver State Sensor (the sampling frequency is 60 HZ), the inertial navigation equipment, and the steering wheel angle sensor (the sampling frequency is more than 20 HZ), etc. because turning behaviour caused by corners and traffic congestions can interfere in the study of the correlation between drowsiness and steering measurements. In order to avoid the influence of road alignment and other factors on driving behaviour, the Hanshi Motorway, which has straight road alignment, better road surface, and less

traffic volume, was chosen as the experimental road. The experimental routes and sensors were shown in *Figure 1*. The high definition cameras collected the outside driving environment, the driver's facial features, and driving manipulation. The Mobileye gathered distances from the vehicle to the lane line. The driver state sensor gathered eye movement data. The inertial navigation equipment recorded velocity, acceleration, and GPS. Besides, we also used materials such as the participant's demographic information scale and the KSS, etc.

2.3 Experimental procedure

Before the experiments, the participants spent at least 10 minutes to familiarise themselves with the experimental vehicle. During the experiments, the participants departed the Fuhe Toll Station, entered the Suizhou Service Area after 2 hours, and spent one hour to have meals and rest. Then, participants drove to the Xiangyang North Toll Station and turned back to the starting point. One experiment took about 6 hours which did not include the rest time, and the total distance was about 600 kilometres. Participants were required to drive at approximately 100 km/h. To avoid interference with driving drowsiness, participants were not allowed to listen to songs and nor to communicate. Experiments were conducted during the non-traffic peak period. During the experiments, the traffic environment was

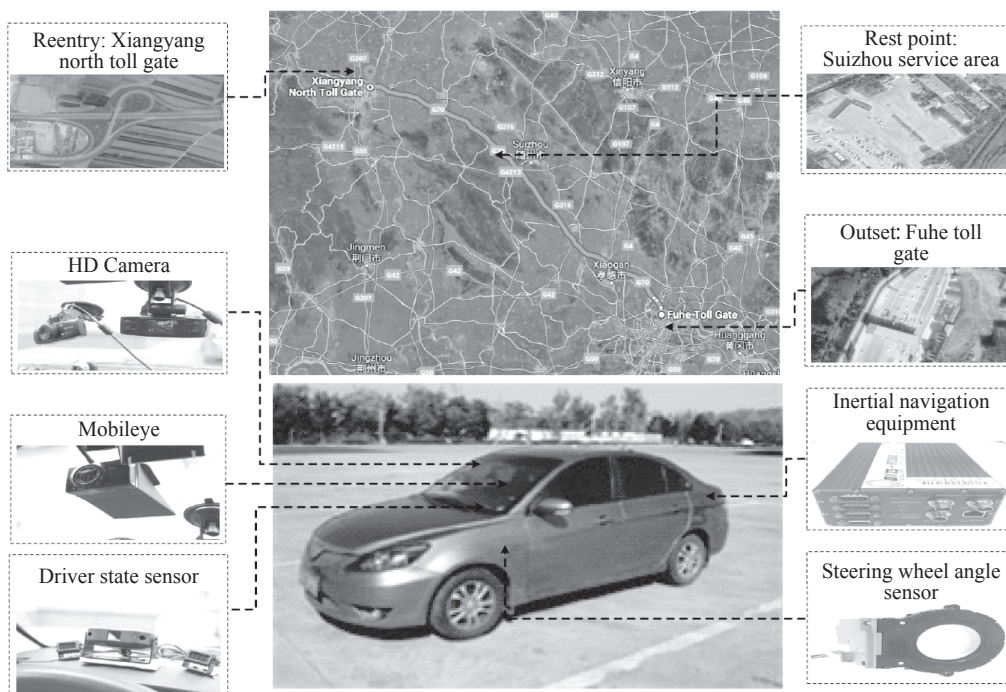


Figure 1 – Experimental routes and apparatus

relatively simple and there was little traffic congestion due to the low traffic volume. Besides, the traffic conditions outside the vehicle were recorded by the camera in real-time. Therefore, some data under abnormal traffic scenes could be deleted based on videos outside the car. The safety officer recorded the participant's self-report KSS every 5 minutes, and after the experiments, the final KSS was determined by considering the videos of the participant's face and the experts' opinions. Participants received certain compensation after finishing the whole experiment.

Due to equipment failure and other reasons, experimental data of 5 participants (No.10, 13, 17, 25, 32) were not collected. Finally, the investigators obtained valid original experimental data of 35 participants. The participants with invalid experimental data were excluded in the following analysis.

3. ANALYSIS METHODS

3.1 Drowsiness measurements

According to references [1, 10, 11, 21], nine non-intrusive drowsiness-driving measurements were chosen. These measurements were computed using original data from various sensors, which reflects the impact of drowsiness on the driver's physiological characteristics, driving operations, and vehicle movement. The information on non-intrusive measurements was presented in *Table 1*.

The calculation formulas of some complicated measurements were as follows.

②SRR reflects the stability of the steering wheel control. Referring to research studies on drowsiness-driving in a real vehicle environment [21], 6° was chosen as the threshold.

$$SRR = \frac{N_{thr}}{N_{all}} \quad (1)$$

where N_{thr} is the number of sampling points exceeding the threshold and N_{all} is the total number of angle sampling points in the time window.

④VSA eliminates the effect of road curvature on the steering wheel angle [23] and reflects the relationship between drowsiness and steering wheel angle.

$$VSA = \frac{MSWA}{SDSWA} \quad (2)$$

where $MSWA$ and $SDSWA$ are the mean value and standard deviation of the steering wheel angle in the time window, respectively.

PERCLOS is the standard measure for drowsiness-detection [10]. The ⑨SDPE describes the variation of PERCLOS in the time window, which can reflect the blink frequency of drivers.

$$SDPE = \sqrt{\sum_{i=1}^{n_{pe}} \left(PE_i - \frac{1}{n_{pe}} \sum_{i=1}^{n_{pe}} PE_i \right)^2 \frac{1}{n_{pe} - 1}} \quad (3)$$

where PE_i is the PERCLOS of the driver at sampling point i , and n_{pe} is the total number of PERCLOS sampling points in the time window.

3.2 Data pre-processing

The time window to calculate the measurement is considered sensitive to drowsiness-detection [21]. According to reference [24], the double-layer sliding time window was proposed to compute measurements. And due to a monotonous environment, long duration of driving, and high speed, drowsiness-driving in the continuous driving scenario on the motorway is more dangerous [7]. Therefore, we focused on the continuous driving

Table 1 – The non-intrusive drowsiness measurements

Data source	Symbol	Description	unit
Mobileye	①SDLP	Standard deviation of lane position	m
Steering wheel angle sensor	②SRR	Steering reversal rate	/
	③SDSWM	Standard deviation of steering wheel movement	°
	④VSA	Variation of steering wheel angle	/
Inertial navigation equipment	⑤SDLA	Standard deviation of longitudinal acceleration	m/s ²
	⑥MATV	Mean absolute value of the transverse velocity	m/s
	⑦SDTV	Standard deviation of transverse velocity	m/s
Driver state sensor	⑧MPECL	Mean of PERCLOS	/
	⑨SDPE	Standard deviation of PERCLOS	/

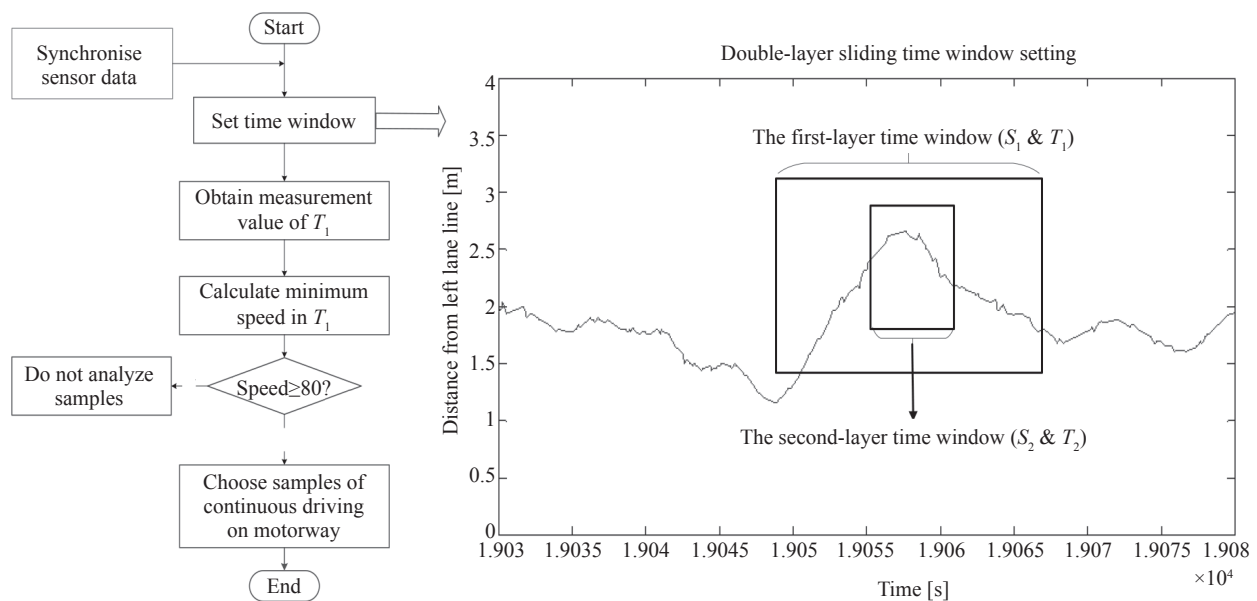


Figure 2 – The data pre-processing flow

scene on the motorway. The pre-processing flow of setting time window and selecting continuous driving data is shown in *Figure 2*.

First of all, multi-source data were synchronised according to the timestamp. Because the driver's self-report drowsiness level was recorded every 5 minutes, in order to avoid sample crossing of two KSSs, the original data were divided into samples within 5 minutes. For each data sample, if the captured videos have demonstrated that the drivers' behaviour is interrupted by any traffic incidents, this sample was excluded.

Set double-layer sliding time window. It is found that in a real vehicle environment, the drowsiness-driving state generally lasts 15–75 seconds, while the duration of typical drowsiness operation characteristics is generally 5–20 seconds [24]. In order to avoid the calculation of the average effect covering the drowsiness operating characteristics, the double-layer sliding time window was proposed to calculate measurements. Firstly, the first-layer sliding step (S_1) and time window (T_1) were set within every 5 minutes. The data in T_1 were a sample that was a fundamental unit for obtaining measurements and detecting drowsiness driving. Secondly, within each T_1 , the second-layer sliding step (S_2) and the time window (T_2) were set. The data in T_2 were used to calculate measurements, the maximum measurement for all T_2 in every T_1 was chosen as the final measurement value of the sample (T_1). For instance, when calculating

SDLP, if $T_1=60s$, $S_2=5s$, $T_2=20s$, T_1 included nine T_2 and the lane position data in each T_2 can be used to calculate the SDLP. The maximum SDLP of all nine T_2 in T_1 was chosen as the final SDLP of the sample (T_1).

Chose measurement samples of consecutive driving data on the motorway. According to the external videos captured by the HD camera, the driving duration of the continuous driving scenario on motorways was extracted. Through statistical analysis of the vehicle speed of these driving durations of the continuous driving scenario on motorways, and according to the reference [21], we chose T_1 with the minimum speed above 80 km/h as the sample in consecutive driving data on the motorway.

3.3 Quantitative model building

As shown in *Figure 3*, this paper proposed a model based on the Wilcoxon-test (W-test) and simple linear regression to quantitatively analyse the effect of IDs on the drowsiness-detection performance of measurements. Nine non-intrusive drowsiness-driving measurements were input into the model one by one to calculate the effect of IDs on their drowsiness-detection performance.

In Part A, based on the result of the Wilcoxon-test for the individual driver, paired-driver combinations having various IDs between two drivers' measurements were formed. We made the Wilcoxon-test

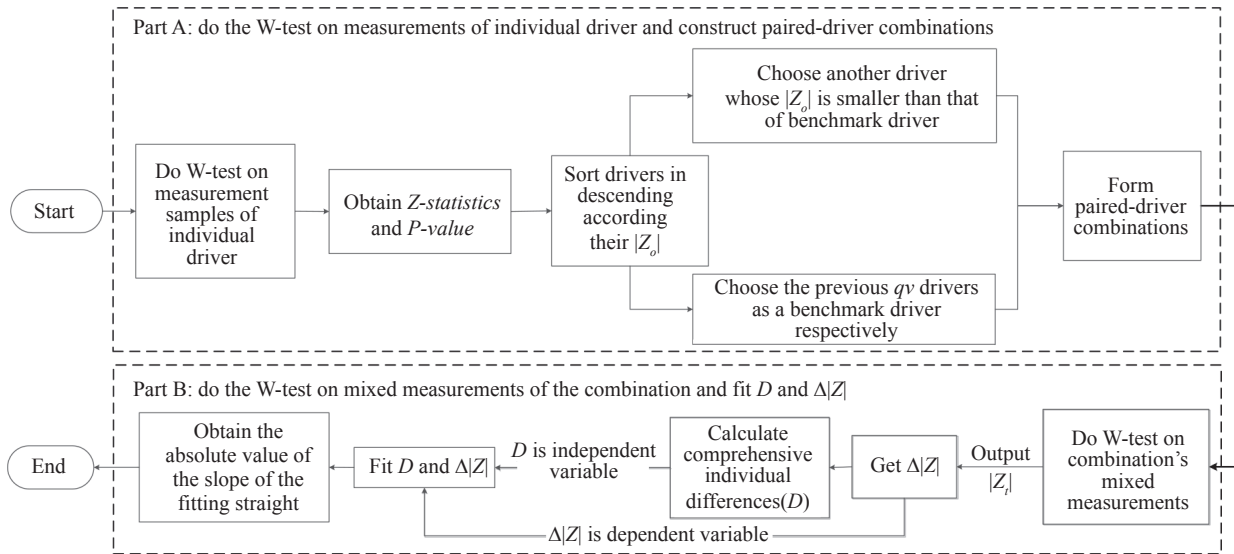


Figure 3 – The model for analysing the effects of IDs on measurements

on the measurement samples of individual drivers and obtained the P -value (P_o) and Z -statistics (Z_o) of the Wilcoxon-test. Then, drivers were arranged in descending order according to their $|Z_o|$, the previous qv drivers were respectively chosen as the benchmark driver (qv needed to be optimised, which will be introduced in *Formula 7*). Next, the benchmark driver was combined with the other driver whose $|Z_o|$ is smaller than that of the benchmark driver to form some combinations including two drivers. Finally, mixed measurement samples of the paired-driver combinations were obtained by mixing measurement samples of the two drivers.

In part B, firstly, the Wilcoxon-test was performed on the paired-driver combination's mixed measurement samples and the combination's Z -statistics (Z_l) was obtained. $\Delta|Z|$ of each paired-driver combination equalled to $|Z_{ob}|$ of the benchmark driver minus $|Z_l|$ of this paired-drivers combination, which represented the change of the MDDP. Secondly, the measurement's comprehensive IDs (D) between the two drivers in the combination were computed using *Formula 5*. Finally, simple linear regression was used to fit D (independent variable) and $\Delta|Z|$ (dependent variable), and the absolute value of the slope of the fitted line indicated the effect degree of IDs on the MDDP.

Wilcoxon-test

The Wilcoxon-test is usually used to analyse the differences between the unpaired samples from the two groups and the data need not follow a normal distribution [25]. The bigger the $|Z$ -statistics, the greater the difference between the measurements of the sober state and the drowsy state. P -value<0.05 means

that there are statistically significant differences between the measurement samples of the sober state and the drowsy state. Thus, $|Z$ -statistics was used to represent the drowsiness-detection performance of measurements.

We supposed there are n measurement samples $S_1, \dots, S_j, \dots, S_n$ in the sober driving state and there are m measurement samples $F_1, \dots, F_j, \dots, F_m$ in the drowsy driving state. The measurement samples from the sober state and the drowsy state are mixed, the amount of the mixed samples is $T=n+m$, and the mixed samples are arranged in ascending.

$$W_s = \sum_{i=1}^n R_i, \quad W_f = \sum_{j=1}^m R_j, \quad W_{FS} = W_s - \frac{m \cdot (m+1)}{2},$$

$$W_{SF} = W_f - \frac{n \cdot (n+1)}{2}, \quad U = \min\{W_{FS}, W_{SF}\}, \tag{4}$$

$$Z = \frac{U - \frac{n \cdot m}{2}}{\sqrt{\frac{n \cdot m \cdot (T+1)}{12}}} \sim N(0, 1)$$

R_i is the rank of the measurement sample S_i of the sober driving state in mixed samples, R_j is the rank of the measurement sample F_j of the drowsy driving state in mixed samples.

Calculation of IDs of measurement

We proposed the following formulas to calculate the comprehensive IDs between the drivers' measurements.

$$ICC_s = \frac{SSA_s}{SST_s}, \quad ICC_f = \frac{SSA_f}{SST_f}, \quad ICC = \frac{ICC_s + ICC_f}{2}$$

$$M_b = M_{bs} - M_{bf}, \quad M_r = M_{rs} - M_{rf}$$

$$M_{2s} = M_{bs} - M_{rs}, \quad M_{2f} = M_{bf} - M_{rf} \tag{5}$$

$$D = \lambda \cdot ICC + \eta \cdot |Z_{ob} - Z_{or}| + a \cdot |M_r - M_b| + \beta \cdot (|M_{2s}| + |M_{2f}|) / 2$$

In the paired-driver combinations, each driver's measurement is a group sample. ICC_s and ICC_f are the intra-class correlation coefficient of the sober state and the drowsy state, respectively. SSA_s and SSA_f are the between-group variances of rank in the sober state and the drowsy state, respectively. SST_s and SST_f are the total variances of rank in the sober state and the drowsy state, respectively. M_{bs} and M_{bf} are the measurement median of benchmark driver in the sober state and the drowsy state, respectively. M_{rs} and M_{rf} are the measurement median of the other driver of a paired-driver combination in the sober state and the drowsy state, respectively. $\lambda, \eta, \alpha, \beta$ are weights of the four IDs source factors ($ICC, |Z_{ob}-Z_{or}|, |M_r-M_b|, (|M_{2s}|+|M_{2f}|)/2$).

Simple linear regression

The simple linear regression model was used to fit $\Delta|Z|$ and D , and the model was as follows:

$$\Delta|Z| = k \cdot D + b \tag{6}$$

Formula 6 reflects the relationship between the change of the drowsiness-detection performance of the measurement and IDs. $|k|$ indicates the effect degree of IDs on the drowsiness-detection performance of the measurement. The smaller $|k|$, the smaller the effect degree of IDs on the MDDP. The coefficient of determination (R^2) evaluates the fitting goodness of the linear model [26]. In this paper, R^2 is no less than 0.4, which means that the fitting function reflects the relationship between $\Delta|Z|$ and D .

In the model shown in Figure 3, the number of benchmark drivers (qv) and $\lambda, \eta, \alpha, \beta$ of Formula 5 affect the R^2 . For improving the fitting effect, the objective function was established to optimise $\lambda, \eta, \alpha, \beta$, and qv . R_i^2 was the R^2 of i -th measurement, which is determined by $\lambda, \eta, \alpha, \beta, qv$.

$$R_i^2 = g(\lambda, \eta, \alpha, \beta, qv)$$

$$h = \sum_{i=1}^9 [R_i^2 - 0.4] \cdot R_i^2 + df_i \cdot 5, \quad df_i = \begin{cases} 1 & R_i^2 \geq 0.4 \\ 0 & R_i^2 < 0.4 \end{cases} \tag{7}$$

Through iterating, we obtained the optimal parameter values that maximise h namely, $\lambda=0.6, \eta=0.1, \alpha=0.1, \beta=0.2, qv=3$.

4. RESULTS

4.1 The Wilcoxon-test for individual participants

For individual participants, according to KSS, the measurement samples were divided into two groups: the sober state and the drowsy state. According to the literature [7], measurement samples with $KSS \leq 3$ belonged to the sober driving state, and measurement samples with $KSS \geq 6$ belonged to the drowsy driving state. The Wilcoxon-test was used to analyse the differences between the measurement samples in the sober state and the drowsy state.

The design of the time window has an important effect on the drowsiness-detection performance of the measurement [20, 24]. This paper aims to guide to select measurements for the generalised drowsiness-detection model. The optimal time window should enable as many drivers as possible to use the measurement to detect drowsiness. Thus, the Wilcoxon-test was performed on each participant's measurement and optimised time windows to maximise the number of participants whose $P_o < 0.05$. We designed the range of time windows [20, 24] and the best time windows for each measurement were acquired by iterating. The optimal time window setting and Wilcoxon-test outcomes were shown in Table 2.

Table 2 – Wilcoxon-test results of drowsiness measurements

Measurements	S_1 [s]	T_1 [s]	S_2 [s]	T_2 [s]	The number of participants whose $P_o < 0.05$	$ Z_{omax} $
①SDLP	5	40	3	12	26	13.62
②SRR	8	60	4	14	28	7.64
③SDSWM	5	38	3	8	29	14.20
④VSA	5	35	3	12	29	10.97
⑤SDLA	5	38	3	8	29	19.00
⑥MATV	5	35	3	10	31	18.16
⑦SDTV	5	44	3	12	29	13.75
⑧MPECL	5	44	3	10	34	29.54
⑨SDPE	5	38	3	12	32	12.27

Notes: $|Z_{omax}|$ was the maximum of $|Z_o|$ for all participants.

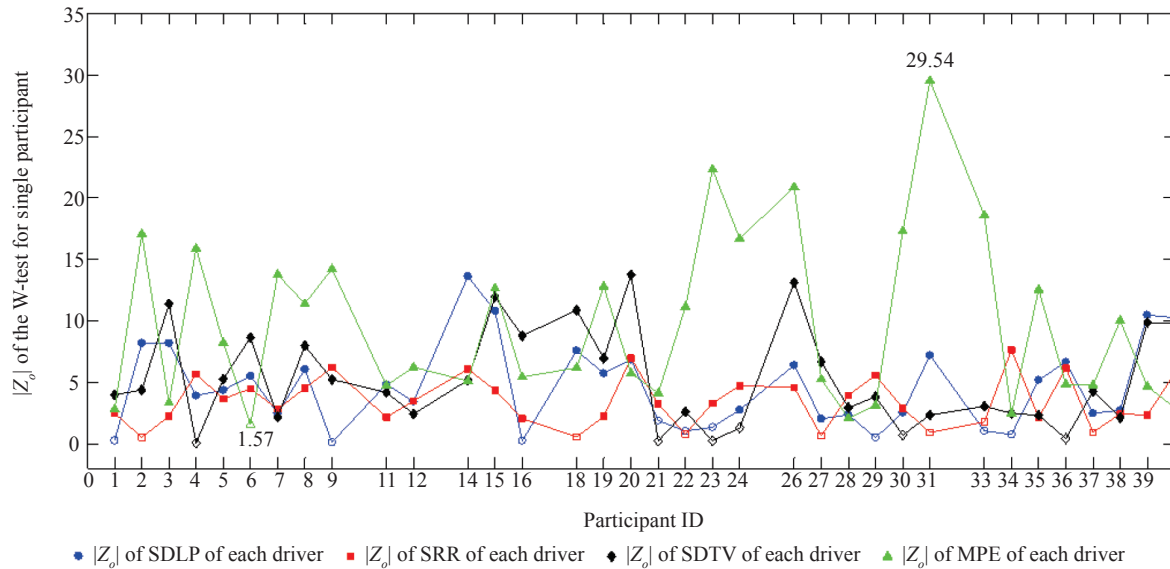


Figure 4 – The $|Z_o|$ of Wilcoxon-test for individual participants

In Figure 4, we took some measurements as examples and listed $|Z_o|$ of all participants. The hollow points are $|Z_o|$ of the participants whose $P_o \geq 0.05$. There were obvious differences among the participants' $|Z_o|$. Taking MPECL as an example, the $|Z_o|$ of the participant No. 31 was the highest with a value of 29.54, which indicated that No. 31 was the participant who used MPECL to detect drowsiness most effectively. Whereas $|Z_o|$ of the participant No. 6 was the lowest minimum with a value of only 1.57. It was found that for different participants, the performances of the same measurement to detect drowsiness-driving were significantly different. It could be speculated that if the MPECL samples of participant No. 31 and No. 6 were mixed, compared with participant No. 31, the performance of MPECL to detect the drowsiness driving might be reduced for the mixed measurement data of No. 31 and No. 6. Besides, $|Z_o|$ of No. 36, No. 37, and No. 39 was similar, which indicated that the drowsiness-detection performance of the MPECL of these participants was similar.

4.2 IDs in measurement distribution

To analyse the IDs in the distribution of the measurements between participants, we took SDSWM as an example and drew the box plot of SDSWM in the sober state and the drowsy state. In Figure 5, we only chose these participants whose $P_o < 0.05$. For most participants, the median of SDSWM in the drowsy state was bigger than that in the sober state and the change in steering wheel movement

was more drastic in the drowsy state. It was indicated that drowsiness could decrease the stability and safety of steering operations for most participants, which was consistent with previous studies [5, 6].

Comparing the data distribution of SDSWM among participants, it was found that there were visible IDs of SDSWM in both the sober state and the drowsy state. In the sober state, the median value of the participant No. 12 was the highest with a value of 0.55, and that of the participant No. 29 was the lowest with a value of 0.27. Furthermore, for participants No. 7, No. 8, No. 9, No. 12, No. 30, and No. 31, the median of SDSWM in the drowsy state was lower than that in the sober state. This was different from the general law of the relationship between SDSWM and drowsiness. However, the mechanism of this phenomenon was not clear, it might be related to the driver's personality, operating habits, alertness, and other attributes. This phenomenon reflected that there were also obvious differences in the trend of changes in measurements from sober driving to drowsy driving between participants.

4.3 Effects of IDs on the measurements' drowsiness-detection performance

For each measurement, according to the model in Figure 3, $\Delta|Z|$ and D for each paired-participant combination were calculated. Then the linear function was used to fit D (independent variable) and $\Delta|Z|$ (dependent variable) of all paired-participant combinations. We chose SDLP, SDSWM as

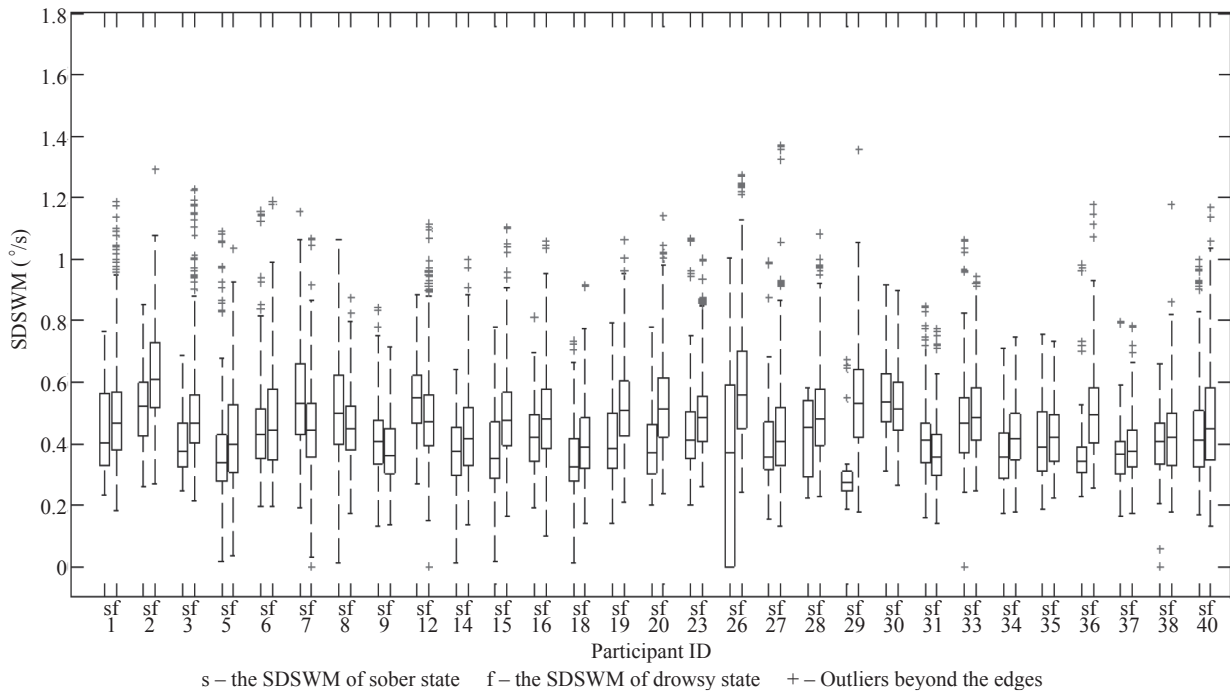


Figure 5 – The boxplots of SDSWM for individual participants

examples to show the fitted results. In Figure 6, each point represents one paired-participant combination consisting of two participants, and the scattered points are approximately distributed on both sides of a straight line. Therefore, the linear function was used to fit D and $\Delta|Z|$.

In Figure 6, $\Delta|Z| > 0$ means that compared to the MDDP for the benchmark participant, the MDDP for the paired-participant combinations decreased. The $\Delta|Z|$ and D were positively correlated. And as D increased, $\Delta|Z|$ increased to greater than zero. The result illustrated that IDs weaken the MDDP, and the greater the IDs, the greater the weakening degree in the MDDP. $|k|$ was different, which illustrated that the effect degrees of IDs on the drowsiness-detection

performance of each measurement were different. Besides, it was found that $\Delta|Z| < 0$ when the IDs were low, which meant that the MDDP for paired-participant combinations increased compared to that for a benchmark participant.

The linear fitting results of $\Delta|Z|$ and D for all measurements were summarised in Table 3. The measurements were arranged in descending order of $|k|$.

In Table 3, $|k|$ of the MATV was the highest with a value of 6.98, which indicated that its drowsiness-detection performance decreased the most rapidly due to the influence of IDs, whereas $|k|$ of MPECL was the lowest and its value was 4.95, indicating that its drowsiness-detection performance was the least influenced by IDs. Furthermore, it was

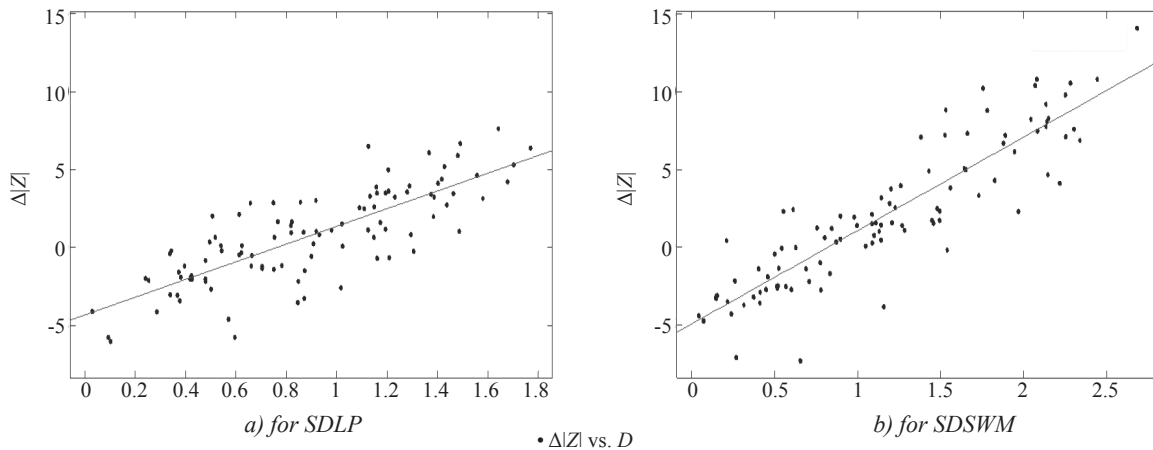


Figure 6 – The fitting lines of D and " $\Delta|Z|$ "

Table 3 – Linear regression results about D and

Measurement	Linear function	$\Delta Z =k \cdot D+b$		
		k (95% confidence interval)	b (95% confidence interval)	R^2
⑥MATV		6.98 (5.90, 8.06)	-6.70 (-8.04, -5.37)	0.59
⑤SDLA		6.93 (5.36, 8.51)	-5.70(-8.54, -2.85)	0.44
②SRR		6.30 (4.77, 7.82)	-3.57 (-4.77, -2.36)	0.45
③SDSWM		6.00 (5.42, 6.58)	-4.93 (-5.71, -4.14)	0.82
⑦SDTV		5.98 (5.31, 6.65)	-4.85 (-5.59, -4.11)	0.76
④VSA		5.88 (4.72, 7.05)	-3.06 (-4.26, -1.86)	0.51
①SDLP		5.66 (4.79, 6.53)	-4.29 (-5.15, -3.43)	0.63
⑨SDPE		5.19 (4.01, 6.38)	-2.86 (-4.56, -1.16)	0.47
⑧MPECL		4.95 (3.89, 6.01)	-5.13 (-8.17, -2.10)	0.47

found that compared to other categories of measurements, the effects of IDs on the drowsiness-detection performances of measurements derived from eye movement were lower.

4.4 The verification of results

The measurements whose drowsiness-detection performance was affected less by IDs were chosen to build models, the decrease in drowsiness-detection accuracy was small when the IDs increased. For verifying the outcomes in Table 3, the single measurement was used to build drowsiness-detection models based on the Fisher discriminant function. We took MPECL and MATV as examples, whose drowsiness-detection performance was the lowest and the most affected by IDs, respectively. Measurement data with different IDs were used to train models. By comparing the change in the drowsiness-detection accuracy with the increase of IDs, effects of IDs on the drowsiness-detection performance of measurements were verified. The drowsiness-detection accuracy of models based on MPECL and MATV is shown in Figure 7.

In Figure 7, when IDs were 0, the data of the driver with maximum $|Z_o|$ were used to train driver-specific drowsiness-detection models. For both MPECL and MATV, the accuracies of driver-specific models using the data of the driver with maximum $|Z_o|$ were the highest (MPECL=86.47%, MATV=79.81%). When adding data of other drivers, the IDs increased from 0 to 2.5, the models became generalised drowsiness-detection models, and the accuracies of models based on MPECL and MATV both decreased. But the decrease in drowsiness-detection accuracy of MPECL (2.93%) was less than

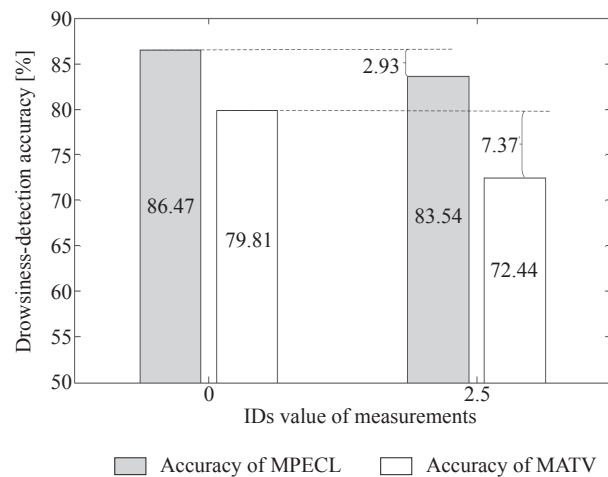


Figure 7 – The accuracy of drowsiness-detection models

that of MATV (7.37%), which verified that the effect degree of IDs on the drowsiness-detection performance of MPECL was smaller than that of MATV. When the IDs were 2.5, the drowsiness-detection accuracy of MPECL was by 11.1% higher than that of MATV. This illustrated that using measurements that were less affected by IDs could improve drowsiness-detection accuracies and adaptabilities of the generalised drowsiness-detection models.

5. DISCUSSIONS

According to the results and analyses above, some meaningful insights and contributions to the development of drowsiness-detection methods considering IDs were obtained. In sections 4.1 and 4.2, there are obvious IDs in distributions of measurements (see Figure 5), which could largely explain why the drowsiness-detection performance of the same measurement is different for different drivers

(see *Figure 4*). For different participants, the measurements with the strongest drowsiness-detection performance are different, thus, for the specific participant, it is more reasonable to choose measurements with the best drowsiness-detection performance of the participant to establish a drowsiness-detection model. Besides, it can be inferred that when the measurement data of two drivers with obvious individual differences in the measurement distribution are mixed, the differences of measurements between the sober state and drowsiness state may become less distinct, which weakens the drowsiness-detection performance of the measurements.

In section 4.3, the fitting lines in *Figure 6* illustrate that IDs have obvious influences on the drowsiness-detection performance of the measurements. IDs in the mixed data of two drivers can weaken the drowsiness-detection performance of measurements when IDs are large enough, which is consistent with the previous study [20]. However, it is found that the drowsiness-detection performance of measurements increases rather than decreases when IDs are small, which is not mentioned in previous literature [20]. The reason may be that when the IDs are small, the effects of IDs on the drowsiness-detection performance of the measurement are limited. Meanwhile, the increasing sample size can provide more information and improve the MDDP. Therefore, the drivers with small IDs in the drowsiness-driving measurements should be grouped and the mixed data of the group are used to train a generalised drowsiness-detection model, which can relatively reduce the effect of IDs and improve the accuracy and utilisation rate of the model.

The most significant achievement of this paper is to propose a quantitative analysis model for the effect of IDs on the MDDP. Using the model, the effects of IDs on the drowsiness-detection performance of non-intrusive measurements were quantified (see *Table 3*). Compared to other categories of measurements, the drowsiness-detection performance of measurements derived from eye movement data is less affected by IDs. Previous studies have only pointed out that IDs of the PERCLOS during drowsiness are low [17], but they have not quantified the effect of IDs on the drowsiness-detection performance of the PERCLOS. The IDs of the measurements derived from eye movement data are low, and the correlations between PERCLOS and drowsiness

are stronger, which may lead to the drowsiness-detection performances of MPECL and SDPE being relatively less affected by IDs.

In section 4.4, the accuracies of driver-specific models trained by individual driver's data are higher than that of generalised models trained by two drivers' data (see *Figure 7*), which is mentioned in previous studies [6]. However, the driver-specific drowsiness-identification model has weak generalisation. Thus, for improving the utilisation rate of models, it is still necessary to study a generalised drowsiness-detection model applicable to some drivers with low IDs. In order to decrease the negative influences of IDs on generalised models, this paper builds a model to find measurements less affected by IDs to train generalised models. In *Figure 7*, when the IDs are the same, the accuracies of generalised drowsiness-detection models based on measurements whose drowsiness-detection performance is less affected by the IDs are higher. Therefore, for effectively weakening the negative effects of IDs and improving accuracies and practicalities of drowsiness-detection, it is recommended to preferentially choose measurements in *Table 3* that are less affected by IDs, such as MPECL and SDPE, to train generalised drowsiness-detection models.

These results can be applied to the development of commercial anti-drowsiness systems considering IDs. For instance, in freight corporations, firstly, managers calculate IDs in the measurement between the drivers to group the drivers with smaller IDs. And then they can use the measurements that are less affected by the IDs (see *Table 3*) to establish generalised drowsiness-detection models suitable for all drivers in the group. Furthermore, the generalised drowsiness-detection models considering IDs can be embedded in anti-drowsiness systems. In this way, on the premise of reducing the effect of IDs and ensuring certain drowsiness-detection accuracy, the utilisation rate of a model can be improved as much as possible to reduce model training costs. The accuracy of the generalised drowsiness-detection model using MPECL is 83.54% (see *Figure 7*), which is higher than that of generalised drowsiness-detection models (78.01%) without considering the effects of IDs on the MDDP in the previous research [5]. Moreover, the generalised drowsiness-detection models in this paper only use individual measurements. Consequently, integrating multiple measurements that are

less affected by IDs to establish drowsiness-detection models may further improve the drowsiness-detection accuracy.

6. CONCLUSIONS

In the present work, through the field experiments on motorways, various naturalistic driving data were non-intrusively acquired. The double-layer sliding time windows were used to compute non-intrusive measurements of drowsiness driving. By iterating, the time windows for each measurement were optimised and nine non-intrusive measurements were computed. Then we proposed a method for calculating the comprehensive IDs in measurements and built a model to quantify the effect of IDs on MDDP.

The effects of IDs on the drowsiness-detection performance of measurements were quantified. The results illustrate that large IDs can weaken the drowsiness-detection performance of non-intrusive measurements. When the IDs become bigger, the drowsiness-detection performance of measurement decreases sharply. However, when the IDs are very small, the drowsiness-detection performance of measurement for mixed data may increase due to the increase of measurement sample size. Besides, compared with other types of non-intrusive measurements, the drowsiness-detection performance of measurements derived from eye movement is less affected by IDs.

The methods and conclusions can provide guidelines for the calculation of drowsiness measurements using naturalistic driving data, quantitative research on IDs, and the measurement selection of drowsiness-detection models considering IDs. There are also some limitations of this paper. For instance, the optimisation method and range of the IDs calculation weight coefficient are simple, which makes the linear fitting goodness of some measurements not high, the combination of participants is relatively simple, and the drowsiness assessment is relatively subjective, etc. In the future, the calculation coefficients of IDs values will be further optimised using advanced optimal algorithms. Besides, a more objective ground truth of drowsiness such as electroencephalograms will be adopted to improve the rationality of the research.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China (52072289, 51775396); the National Key Research Program of

China (2019YFB1600800); National Natural Science Foundation of China – Collaborative Foundation (U1764262); Hubei Innovation Group Project (2017CFA008).

孙一帆 博士¹
电子邮箱: m18705423520@163.com
吴超仲 教授¹
电子邮箱: wucz@whut.edu.cn
张晖 副教授¹
电子邮箱: zhanghuiits@whut.edu.cn
楚文慧 博士^{1,2}
电子邮箱: chuwh@whut.edu.cn
肖逸影 硕士¹
电子邮箱: shadowxy@outlook.com
张奕俊 硕士¹
电子邮箱: zyjits@whut.edu.cn

¹ 武汉理工大学 智能交通系统研究中心, 武汉市, 湖北省, 中国

² 武汉理工大学, 武汉市, 湖北省, 中国

个体差异对指标疲劳检测性能的影响

摘要:

个体差异可能通过影响指标的疲劳检测性能而降低疲劳驾驶的检测精度。本文的目的是提出一个可以量化个体差异对指标疲劳检测性能影响的模型, 并且找到受个体差异影响较小的指标来建立疲劳检测模型。通过实车实验, 我们采集了驾驶人的自然驾驶数据和主观疲劳水平, 使用双层滑动时间窗来计算与疲劳相关的指标。在模型中, 使用威尔科克森检验的Z统计量的绝对值 ($|Z\text{-statistics}|$) 来表示指标的疲劳检测性能。首先, 使用威尔科克森检验对个体驾驶人的指标进行分析。然后, 将驾驶人两两组合, 对两驾驶人组合的指标进行威尔科克森检验, 并且计算两驾驶人组合的指标的个体差异。最后, 使用线性回归拟合指标个体差异和由个体驾驶人的 $|Z\text{-statistics}|$ 减去两驾驶人组合的 $|Z\text{-statistics}|$ 所表示的驾驶人疲劳检测性能变化量, 使用拟合直线斜率绝对值 $|k|$ 表示指标疲劳检测性能受个体差异的影响。结果表明, 眼睛闭合时间比例均值的 $|k|$ 最小是4.95, 表明该指标受个体差异影响最小。研究结果有助于考虑个体差异的疲劳检测模型的指标选择。

关键词:

交通安全; 疲劳检测模型; 非侵入性指标; 自然驾驶实验; 个体差异

REFERENCES

- [1] Doudou M, Bouabdallah A, Berge-Cherfaoui V. Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *International Journal of Intelligent Transportation Systems Research*. 2019: 1-23.
- [2] Anund A, Fors C, Ihlström J, Kecklund G. An on-road study of sleepiness in split shifts among city bus drivers.

- Accident Analysis & Prevention*. 2018;114: 71-6.
- [3] Otmani S, Pebayle T, Roge J, Muzet A. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology Behavior*. 2005;84(5): 715-24.
- [4] Fu RR, Wang ST, Wang SW. Real-time Alarm Monitoring System for Detecting Driver Fatigue in Wireless Areas. *Promet – Traffic&Transportation*. 2017;(2): 165-74.
- [5] Li ZJ, et al. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. *Sensors*. 2017;17(3).
- [6] Liu CC, Hosking SG, Lenné MG. Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research*. 2009;40(4): 239-45.
- [7] Wang XS, Chuan X. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention*. 2016;95: 350-7.
- [8] He QC, Li W, Fan XM, Fei ZM. Driver fatigue evaluation model with integration of multi-indicators based on dynamic Bayesian network *IET Intelligent Transport Systems*. 2015;(5): 547-54.
- [9] Al-Libawy H, Al-Ataby A, Al-Nuaimy W, Al-Tae MA. Modular design of fatigue detection in naturalistic driving environments. *Accident Analysis & Prevention*. 2018;120: 188-94.
- [10] Cheng Q, et al. Assessment of Driver Mental Fatigue Using Facial Landmarks. *IEEE Access*. 2019;7: 150423-34.
- [11] Wakita T, et al. Driver identification using driving behavior signals. *IEICE Transactions on Information*. 2006;89(3): 1188-94.
- [12] Thiffault P, Bergeron J. Fatigue and individual differences in monotonous simulated driving. *Personality Individual Differences*. 2003;34(1): 159-76.
- [13] Yan RH, Wu C, Wang YM. Exploration and evaluation of individual difference to driving fatigue for high-speed railway: A parametric SVM model based on multidimensional visual cue. *IET Intelligent Transport Systems*. 2018;12(6): 504-12.
- [14] Zhao XH, Zhang XJ, Rong J. Study of the Effects of Alcohol on Drivers and Driving Performance on Straight Road. *Mathematical Problems in Engineering*. 2014.
- [15] Ingre M, et al. Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of Sleep Research*. 2006;15(1): 47-53.
- [16] Philip P, et al. Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*. 2005;(3): 473-8.
- [17] Xu C, Wang XS, Chen XH. Evaluating Performance of Non-intrusive Indicators on Drowsy Driving Detection. *Journal of Southwest Jiaotong University*. 2014;49(4): 720-6.
- [18] Chu W, et al. Driver behavior model and its application in driver fatigue identification. *China Safety Science Journal*. 2018;28(06): 43-8.
- [19] You F, et al. A Real-time Driving Drowsiness Detection Algorithm With Individual Differences Consideration. *IEEE Access*. 2019;17: 179396-408.
- [20] Niu QN. *Research on Driving Fatigue Detection Based on Hybrid Measures*. PhD thesis. Jilin, Changchun: Jilin University; 2014.
- [21] Zhang H, et al. Qiu TZ. Sensitivity of lane position and steering angle measurements to driver fatigue. *Transportation Research Record*. 2016;2585(1): 67-76.
- [22] Sikander G, Anwar S. Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*. 2018;20(6): 2339-52.
- [23] Jin LS, Li KY, Niu QN, Gao LL. A New Method for Detecting Driver Fatigue Using Steering Performance. *Journal of Transport Information and Safety*. 2014;32(5): 103-7.
- [24] Zhang XB, Cheng B, Feng RJ. Real-time detection of driver drowsiness based on steering performance. *J Tsinghua Univ*. 2010;7: 1072-6.
- [25] Kruskal WH. Historical notes on the Wilcoxon unpaired two-sample test. *Journal of the American Statistical Association*. 1957;52(279): 356-60.
- [26] Zhang D. A coefficient of determination for generalized linear models. *The American Statistician*. 2017;71(4): 310-6.