Empirical Evaluation of Multi-modal Mental Fatigue Assessment using Low-cost Commercial Sensors

Wang Heqiao
National Junior College
Singapore 288913

Aung Aung Phyo Wai
Institute for Infocomm Research, ASTAR
Singapore 138632

Abstract—**Drowsy driving poses a dangerous threat to road safety. This study aims to compare and evaluate the accuracy afforded by different modalities in assessing drowsiness, including: electroencephalogram (EEG), eye tracker, photoplethysmogram (PPG), and video recording. The work also investigates the impact of two parameters in the driving tasks on subjects' level of alertness, presence of road-mark and frequency of lane-departure events. Ten healthy subjects without sleep deprivation participated in individual 1.5-hour experiments. A threshold of 80th percentile of reaction time was taken as ground truth to label data as Drowsy or Non-Drowsy. Using supervised learning random forest algorithm and stratified 10-fold cross validation, the results suggest that EEG features achieved highest classification accuracy: 0.957 and 0.864 for individual and combined sessions respectively, followed by eye tracker (0.821, 0.755 respectively). The highest accuracy of all modalities fell in the section that has the longest reaction time. These experiments additionally show that an absence of road-marks does indeed increase subjects' reaction time, though they may not necessarily become drowsier. Further, low frequency of lane-departure events did make subjects drowsier as hypothesized. Since existing commercial products that claim to detect drowsiness are very expensive and target at vehicle transportation companies only, they are not available to daily private car users. As such, this type of study deserves attention so that drowsiness detection products could be made affordable and accessible to both professional drivers and daily private car users.**

Keywords-drowsy driving; multi-modal; low-cost; driving simulator; reaction time; random forest classifier

I. INTRODUCTION

Drowsy driving is a great concern to road safety. 25% of the deadly traffic accidents on the highway are caused by the temporary drowsiness of drivers, according to the German Road Safety Council [1]. Likewise, based on police reports, the US National Highway Traffic Safety Administration (NHTSA) estimated that driver drowsiness is the culprit of at least 100,000 vehicle crashes each year [1]. To truly prevent these devastating accidents, the driver's state of drowsiness should be monitored.

Current measures to monitor a driver's drowsiness can be mainly categorized into two broad types: vehicle-based and user-based, which can, in turn, be further divided into behavioural and physiological measures. This study focuses on user-based measures - both behavioural (eye tracker [2] and video recording [3]) and physiological (EEG [4] and PPG [5, 6]). Compared to vehicle-based measures that analyse the external effects of drowsiness on vehicle motion, user-based measures directly monitor the internal state and behaviour of drivers. The onset of drivers' drowsiness is a gradual, which constitutes a cumulative process that often takes sometime before it manifests into noticeable changes in vehicle motion. In this study, we hypothesise that user-based measures are able to detect this earlier, and more reliably, than vehicle-based measures. Consequently, there is a high chance that the warnings afforded by vehicle-based measures may be too late to prevent the impending accident as the tragedy may have already taken place when the vehicle begins to move abnormally.

There are already several commercial driving fatigue monitoring and management systems available that claim to be able to detect the early onset of drowsiness in real-time. The ground rule in such systems is that they must not distract the drivers from their driving task. A few examples of such systems are SmartCap, Optalert and GuardVant, which utilize EEG (SmartCap Tech) [7], eye tracking (Optalert) [8] and real-time video facial detection (GuardVant) [9] respectively to detect drowsiness. However, aside from technical difficulties such as being able to operate accurately over expected range of vehicle temperature, humidity, and vibration conditions, each system only makes use of a single modality.

This study aims to compare and evaluate the accuracy of different modalities – EEG, eye tracker (E.T.), PPG, and video recording – in assessing drowsiness. This study also investigates the impact of two parameters in the driving simulator, presence of road-mark and frequency of lane-departure events, on subjects' level of alertness, by making an inter-section comparison. It was hypothesised that the presence of white road-marks as moving and repetitive visual stimuli induces mental fatigue and thus drowsiness from drivers and that low frequency of lane-departure events would cause drivers to react longer and become drowsier.

II. EXPERIMENT DESIGN

A. Driving Simulator

The driving simulator comprises of a fixed base, a gaming steering wheel, and a desktop screen (see Fig. 1). The simulation created a monotonous driving environment, which is more likely to cause mental fatigue [10]. There was a single car travelling at fixed speed on a three-lane highway, with no random or discrete visual stimulus that may disturb the monotonous pattern.

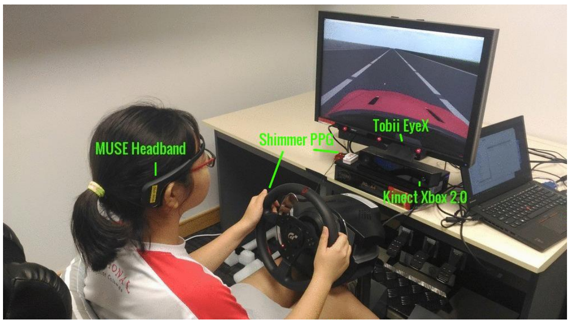

Figure 1. Experiment Setup

B. Sensors

The sensors used in the experiments included: a MUSE EEG Headband, a Tobii EyeX Eye Tracker, a Shimmer PPG, and a Kinect Xbox 2.0 (see Table 1).

TABLE I. USE OF SENSORS IN THIS STUDY

Modality	**Model (Cost)**	***Features used in this study***	***Sampling rate***
EEG	MUSE Headband (US$250)	Raw EEG singals from 4 channels: Tp9, FP1, FP2, Tp10, with 3 reference channels	256 Hz
E.T.	Tobii EyeX (US$80)	Gaze position	60 Hz, with eye to application latency 15 ms+/−5 ms
PPG	Shimmer	Calibrated PPG values	256 Hz
Video	Kinect Xbox 2.0	Real-time video recording	Around 15fps

C. Procedure

All experiments started at 530 pm at the lab at ASTAR Institute for Infocomm Research, Singapore. Subjects were asked to perform five driving tasks in a pre-determined sequence: circuit driving, followed by Section 1 to 4 (see Table II), with no intermission in between. Subjects were taught how to move from one section to another prior to the tasks, and then performed the tasks alone in the lab.

TABLE II. PROCEDURE

Task index	**Task details**				
	Task type	***Duration***	***Time interval between events***	***Presence of road-marks***	***Time recorded for below events***
Circuit	Free driving	15 min	-	Present	Start and end of the section
Section 1 (S1)	Car-deviating (Left 50% Right %)	15 min	Short, 2-8 s in random	Present	Start and end of the section; Deviation onset; Response onset (correct, wrong, no response); False turn
Section 2 (S2)		15 min		Absent	
Section 3 (S3)		15 min	Long, 10-19 s in random	Present	
Section 4 (S4)		15 min		Absent	

For the 5-min circuit driving task, subjects were able to control the car movement freely. This resembles arcade car games. This was to set contrast with the latter sections in which subjects had no control. By using the Karolinska Sleepiness Scale [11], all subjects reflected no increase in their level of sleepiness after the completion of this task. Therefore all data obtained during this time was treated as Non-drowsy data.

As indicated in Table 1, the tasks corresponding to S1 to S4 each last 15 minutes. For these tasks, the simulator automatically and randomly causes the car to deviate away from the central lane, which was considered as a lane-departure event. Subjects were instructed to keep the car in the central lane using the steering wheel whenever a lane-departure event occurred. The probability of a left and a right deviation was kept uniform (i.e. at 50% to 50%) for all four car-deviating tasks. The time interval between each adjacent event varied from task to task and from event to event. The time of a correct response onset was recorded when the subjects turned the steering wheel in the expected direction; conversely it was recorded as the time of a wrong response onset. Finally, the time of no response onset was recorded if subjects made no response within the maximum reaction time allowed of 5.9 seconds after deviation onset. In this third case, the car was automatically returned to the central lane. If the subject turned the steering wheel when the car was supposed to be travelling straight on the central lane, the car would not deviate, though a false turn would be recorded.

III. DATA PROCESSING AND ANALYSIS

A. Data Labelling

RT for each lane-departure event was calculated as the time latency between deviation onset and response onset. Number of errors and misses were also counted.

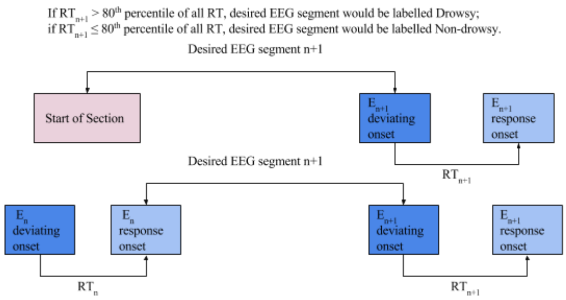

Figure 2. Illustration of data labelling process

As seen in Fig. 2, EEG segment between deviation onset of E_n and response onset of E_n was chosen where subjects were in the natural state in absence of visual stimuli. This was when subjects started to feel drowsy, which caused them to respond slower to the next event. EEG segment between deviation onset of E_n and response onset of E_n was not taken into consideration because that segment involves information processing and execution of attentional network of subjects [12], in fulfilment of the task demand. It was assumed that subjects would become slightly more alert when they visualized each lane-departure event and responded to it [13]. The same data labelling method applies to eye tracker data.

B. Feature Extraction

TABLE III. FEATURE EXTRACTION

Modality	Data	Feature extracted
EEG	Raw EEG Data	Moving average of delta (0.3-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), low beta (12-18 Hz), high beta (18-30 Hz), beta (12-30 Hz), and gamma (30-40 Hz) band energy Start and end of the section
Eye Tracker	Gaze Position	Gaze position and their changing velocity* with a new sampling rate of around 1.33 Hz
PPG	Calibrated PPG Data	Fast Fourier Transform (FFT) and AutoRegression (AR) heart rate variability** (HRV) spectrograms for each section [14]

*Velocity＝Displacement÷Time

** HRV is a measure of the beat-to-beat (R-R Intervals) changes in the heart rate. In HRV the low (LF) and high (HF) frequencies fall in the range of 0.04–0.15 Hz and 0.14–0.4 Hz, respectively.

C. Classifier Training and Validation

Data was then applied to the supervised learning random forest algorithm [15] in software Orange3 in order to train a classifier that categorised if a subject is Drowsy or Non-drowsy. Ten trees were used and subsets smaller than five were not split. Stratified 10-fold cross validation [16] was then employed to evaluate viability and classification accuracy (CA).

IV. DISCUSSION

EEG features achieve higher CA, precision and recall than E.T. features across all sections (see Table III). This is because EEG measures the internal state of subjects, while E.T. measures their external behaviour. The onset of drowsiness may be so subtle that it could not be reflected in eye movement. Furthermore, EEG has seven features from each independent channel. However, only two features were extracted from E.T. data, i.e. gaze position and its dependent velocity. This smaller number and dependency of E.T. features may translate to lower CA.

TABLE IV. CA COMPARISON BETWEEN EEG AND E.T.

Section	Feature	CA	Precision	Recall
S1	EEG	0.93	0.93	0.93
	E.T.	0.73	0.65	0.73
S2	EEG	0.96	0.96	0.96
	E.T.	0.82	0.76	0.82
S3	EEG	0.95	0.95	0.95
	E.T.	0.77	0.72	0.77
S4	EEG	0.88	0.84	0.94
	E.T.	0.69	0.67	0.69
Avg.	EEG	0.93	0.92	0.95
	E.T.	0.75	0.70	0.75
All	EEG	0.86	0.87	0.86
	E.T.	0.76	0.68	0.76

The CA of EEG features decreases when data of all four sections were combined for training (see Table III): overall accuracy is lower than any single section's accuracy. In contrast, while accuracy of E.T. features also falls, it is even higher than the average accuracy. For E.T. features, highest accuracy was achieved when both gaze position and velocity features were applied (see Table IV). In particular, gaze position feature gave a higher accuracy than velocity feature. Unlike velocity and EEG where the accuracy fell when data were combined across sections for training, the gaze position accuracy increased.

TABLE V. CA COMPARISON BETWEEN EEG AND E.T.

Section	Feature	CA
S1	Gaze position	0.73
	Velocity	0.67
	Both features	0.75
S2	Gaze position	0.75
	Velocity	0.74
	Both features	0.81
S3	Gaze position	0.76
	Velocity	0.73
	Both features	0.78
All	Gaze position	0.77
	Velocity	0.73
	Both features	0.78

This finding suggests that gaze position was not affected by the parameters in the driving tasks. This may be because the road pattern on the simulator screen was not changing in different sections: always monotonous with three driving lanes on the highway. Subjects had adopted their own way that they were comfortable with to look at the screen throughout the experiment.

Additionally, for each subject, the section that obtained the highest accuracy from both modalities (in this case S2) had the longest RT, which also had the largest range and standard deviation (SD) (see Table IV). When RT spans over a longer range and varies more greatly from each other, as depicted in the box plot in Fig. 3, the labelling method can recognize drowsy data better and more accurately, which then gave the modalities higher CA.

It was also observed that for each subject, the ranking of CA of EEG and E.T. features is identical in Table III. This consistency suggests data labelling and both methods reliable.

TABLE VI. CA COMPARISON BETWEEN EEG AND E.T.

Section	RT Range	RT SD
S1	1.79	0.263
S2	2.61	0.356
S3	0.98	0.184
S4	1.50	0.258

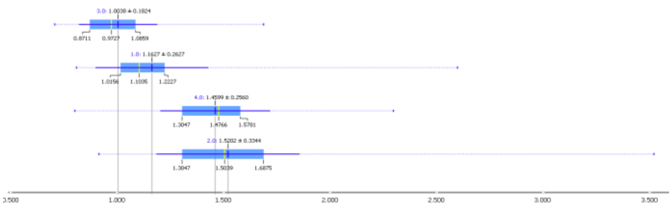

Figure 3. RT Box Plot (from top to bottom S3, S1, S4, S2 respectively)

It was observed that both FFT and AR spectrum patterns of S1 and S3 is similar (see Figure 4). Similarly, circuit driving, S2 and S4 also share similar patterns. These interesting findings may be due to the absence of road-marks. Since data in circuit driving was treated as Non-drowsy data, the similarity in pattern suggests subjects were rather alert in S2 and S4 as compared to S1 and S3. Thus the absence of road-marks actually caused subjects to become more active, instead of drowsier.

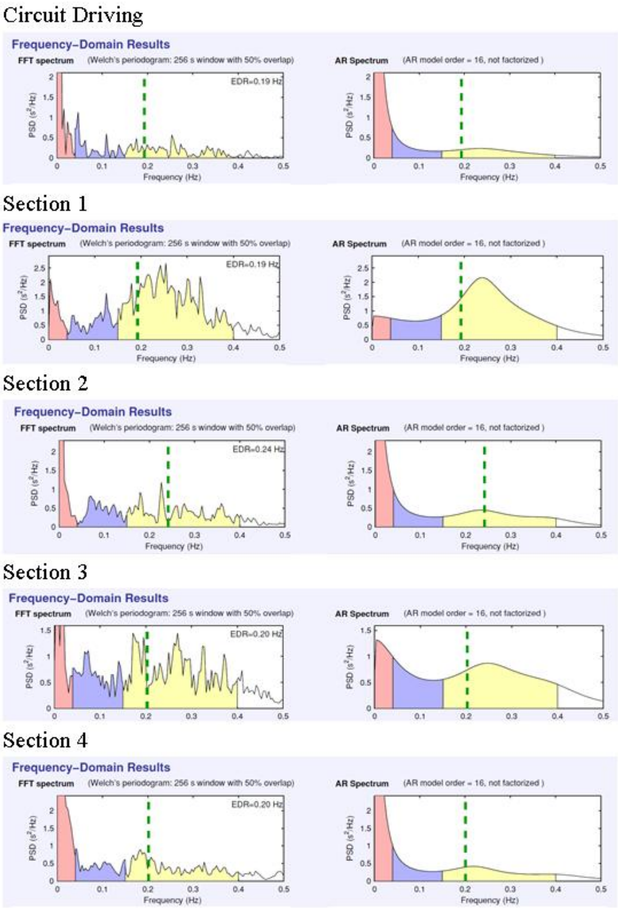

Figure 4. HRV spectrums for all sections

RT is longer when road-marks are absent; this holds true for all subjects studied. This is evident in Fig. 5, where colour for S2 and S4 is much deeper compared to that for S1 and S3 (S1 and S2 are longer because the events were more frequent).

Figure 5. RT Heat Map Across Sections

This is because absence of road-marks makes lane departure events more difficult to recognize. Contrary to the hypothesis that moving road-marks as repetitive visual stimuli would exacerbate mental fatigue of subjects, subjects reflected that the presence of road-marks helped recognize each lane-departure event easier. They were thus able to make faster reaction for each deviation. This study thus proves the importance and necessity of road-marks in maintaining road safety on highway. Future work can focus on the most desirable combination of length, width, colour and shape of road-marks in keeping drivers alert, so as to be implemented in reality.

It was hypothesized that low frequency of lane-departure events induces mental fatigue more greatly than high frequency. Among ten subjects, nine of them had longer RT in S3 and S4; one of them had lower RT instead. The results are thus able to prove the hypothesis. This difference in RT and SD was magnified by the absence of road-mark. The fact that there was a set of anomalous data also suggests that frequency indeed influenced RT. It may be concluded that frequency of lane-departure events had different effects on different subjects, causing them to respond either slower or faster. This phenomenon could only be explained by individual variances. The anomalous subject actually revealed that she was singing in the last two sections in order to maintain alert. This reflects one limitation of the experiment: as much as the process was kept natural and comfortable for subjects, they might choose to take effort to maintain alert instead of yielding to drowsiness. In contrast, one subject made no effort to stay alert and he actually slept for half of the last section.

V. Conclusion

This study established two comparisons: inter-modality and inter-section. EEG features obtained higher accuracy than E.T. features across all sections. Further, for each subject, the ranking of EEG and E.T. accuracy of different sections was identical. Both methods yielded highest accuracy in the section that had longest RT, which in turn had greatest range and SD. Gaze position feature afforded a higher accuracy than velocity feature; highest accuracy was obtained when both features were used. Unlike EEG and velocity features, gaze position feature actually yielded a higher accuracy when data from all four sections was used to train the classifier. Therefore, it suggests that E.T. features perform more consistently in changing road conditions, which is an advantage over EEG features despite its lower accuracy. HRV spectrums suggest that absence of road-mark made subjects more active.

Absence of road-marks increased RT of all subjects, which was not due to increased drowsiness but the increased difficulty of the tasks. This study proved the importance and necessity of road-marks in protecting road safety. Low frequency of lane-departure events induces drowsiness from drivers. The limitation of this study was mainly due to the young age of subjects, who are not real drivers. However, since younger people have a shorter reaction time than older drivers [17], it was hoped that a higher benchmark would be set in assessing driver's drowsiness. Further, as this experiment was designed by the student researcher, there could have been imperfect aspects in terms of experimental control; for example, some subjects tried their best to stay alert till the end of the experiment, while others simply fell asleep. This discrepancy caused the inter-subject variances to become very huge, which in turn diminished some significance trend that inter-subject data might unveil. Last but not least, the risk that drowsy driving carries makes this type of research dangerous to be carried out in the real-world context, where the safety of the subjects is not guaranteed. However, there are a few sensor-based studies that are tested in real-environment [18, 19]. Studies have pointed out that reaction time will be more affected in a simulated environment [18], which is likely to cause higher subjective and physiological drowsiness [19]. Hence, it is only possible for this study to be carried out in the simulated setting at this preliminary stage.

Nevertheless, this study is the first in the field that compares the accuracy of different modalities and meanwhile investigates the impacts of road-marks and frequency of lane-departure events, on driver's drowsiness, with the employment of visualisations. Future work could continue in this direction and find out the most desirable combination of length, width, shape as well as colour of road-mark and the most desirable frequency of events to optimise driver's alertness. In future work, more expressive and representative features may be extracted from data to aim for even higher accuracy. Greater number of subjects in different age brackets would be included to enhance the inclusiveness of the finding.

In conclusion, low-cost commercial sensors show high potential in assessing drivers' drowsiness. Existing products that claim to detect drowsiness target mainly truck drivers and the companies behind them. As such, they are almost inaccessible to private car owners. Although the relatively low complexity of the features is the trade-off of the much lower costs and greater usability of commercial sensors as compared to expensive clinical sensors, high accuracy (above 95%) of the EEG sensor encourage future research and make it affordable to both professional drivers and private car users.

REFERENCES

[1] Sahayadhas, A., Sundaraj, K., Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. Sensors (Basel). 12(12), 16937-53.

[2] Duchowski A T. Eye tracking methodology[J]. Theory and practice, 2007, 328.

[3] Picot A, Charbonnier S, Caplier A. Drowsiness detection based on visual signs: blinking analysis based on high frame rate video[C]//Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE. IEEE, 2010: 801-804.

[4] Santamaria J, Chiappa K H. The EEG of drowsiness[M]. Demos Publications, 1987.

[5] Li G, Chung W Y. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier[J]. Sensors, 2013, 13(12): 16494-16511.

[6] Park H, Oh S, Hahn M. Drowsy driving detection based on human pulse wave by photoplethysmography signal processing[C]//Proceedings of the 3rd International Universal Communication Symposium. ACM, 2009: 89-92.

[7] SmartCap Tech. Retrieved from http://www.smartcaptech.com/

[8] Optalert. Retrieved from Optalert: http://www.optalert.com/

[9] GuardVant. Retrieved from http://www.guardvant.com/

[10] Jacques, P. T. (2003). Monotony of road environment and driver fatigue: a simulator study. Accid Anal Prev, 35(3), 381-91.

[11] Shahid A, Wilkinson K, Marcu S, et al. Karolinska sleepiness scale (KSS)[M]//STOP, THAT and One Hundred Other Sleep Scales. Springer New York, 2011: 209-210.

[12] Posner M I. Measuring alertness[J]. Annals of the New York Academy of Sciences, 2008, 1129(1): 193-199.

[13] Matthias E, Schandry R, Duschek S, et al. On the relationship between interoceptive awareness and the attentional processing of visual stimuli[J]. International Journal of Psychophysiology, 2009, 72(2): 154-159.

[14] Byeon M K, Han S W, Min H K, et al. A study of HRV analysis to detect drowsiness states of drivers[J]. balance, 2006, 5: 6.

[15] Breiman, L. (2006) Machine Learning (2001) 45(1), 5-32

[16] Refaeilzadeh P, Tang L, Liu H. Cross-validation[M]//Encyclopedia of database systems. Springer US, 2009: 532-538.

[17] Wang, D. Y. D., & Entsminger, S. Age and attentional capacity. Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

[18] Philip P, Sagaspe P, Taillard J, et al. Fatigue, sleepiness, and performance in simulated versus real driving conditions[J]. Sleep, 2005, 28(12): 1511-1516.

[19] Hallvig D, Anund A, Fors C, et al. Sleepy driving on the real road and in the simulator—A comparison[J]. Accident Analysis & Prevention, 2013, 50: 44-50.