Real-time evaluation of driver's alertness on highways

G. S. Larue, A. Rakotonirainy & A. N. Pettitt
Centre for Accident Research and Road Safety, Queensland, Australia

Abstract

Suburbanisation has been internationally a major phenomenon in the last decades. Suburb-to-suburb routes are nowadays the most widespread road journeys; and this resulted in an increment of distances travelled, particularly on faster suburban highways. The design of highways tends to over-simplify the driving task and this can result in decreased alertness. Driving behaviour is consequently impaired and drivers are then more likely to be involved in road crashes. This is particularly dangerous on highways where the speed limit is high. While effective countermeasures to this decrement in alertness do not currently exist, the development of in-vehicle sensors opens avenues for monitoring driving behaviour in real-time. The aim of this study is to evaluate in real-time the level of alertness of the driver through surrogate measures that can be collected from in-vehicle sensors. Slow EEG activity is used as a reference to evaluate driver's alertness. Data are collected in a driving simulator instrumented with an eye tracking system, a heart rate monitor and an electrodermal activity device ($N = 25$ participants). Four different types of highways (driving scenario of 40 minutes each) are implemented through the variation of the road design (amount of curves and hills) and the roadside environment (amount of buildings and traffic). We show with Neural Networks that reduced alertness can be detected in real-time with an accuracy of 92% using lane positioning, steering wheel movement, head rotation, blink frequency, heart rate variability and skin conductance level. Such results show that it is possible to assess driver's alertness with surrogate measures. Such methodology could be used to warn drivers of their alertness level through the development of an in-vehicle device monitoring in real-time drivers' behaviour on highways, and therefore it could result in improved road safety.

1 Introduction

Most metropolitan cities in the world have experienced rapid suburbanisation over the past few decades. This phenomenon has not brought people and jobs closer to each other due to diffuse commuting patterns [1]. Often people are not able to live close to where they work because of a lack of affordable housing in many metropolitan areas. Internationally, average trip lengths have increased dramatically in the last 20 years and this increase has been linked to the growing use of the car [2]. This led to transport networks developments and increments in commuting distances. For instance work trip distance has sharply increased while the mean daily commuting time per capita has shown only a slight growth [3]. More and more people residing in the suburban areas are using faster modes, and the suburbanisation of houses and workplaces is the outcome of the increasing use of faster suburban routes instead of slow urban routes [4]. The suburb-to-suburb journey to work is by far the most widespread. Such journey is made from an outlying residential area to a nearby suburban employment centre, crossing or not some towns [2].

Australian cities, with the exception of central Sydney, have been planned according to the good urban planning practices of the 19th and 20th centuries. To a large extent endemic congestion has not been a major factor as it is in other developed countries. Australian cities continue to grow, space being not a problem, and spatial separation increases at metropolitan, regional and national levels [5]. Disjunction between home and work location increases and communities are increasingly fragmented and connected by highways [6].

The design of highways tends to over-simplify the driving task. On such roads, driving is mainly a lane-keeping task. A lane keeping task is not cognitively stimulating and can cause the driver to suffer from an alertness decrement after less than 20 minutes [7]. Driving behaviour is consequently impaired and drivers are then more likely to be involved in road crashes due to potentially slow reaction times or lack of reaction to unpredictable events. This is particularly dangerous on highways where the speed limit is high. This draws inattention to the most important contributor (27%) to fatal and hospitalisation crashes in 2003 in Queensland, Australia (Queensland Transport, 2005). While effective countermeasures to this decrement in alertness do not currently exist, the development of in-vehicle sensors opens avenues for monitoring driving behaviour and reduction in alertness in real-time.

The aim of this study is to evaluate in real-time the level of alertness of the driver on different types of Australian highways through surrogate measures that can be collected from in-vehicle sensors. This study focuses on two factors that decrease alertness: (*i*) the road design and (*ii*) the roadside environment. Slow EEG activity is used as a reference to evaluate driver's alertness. Alertness decrement is detected in real-time using surrogate measures related to the driving performance. Data are

collected in a driving simulator instrumented with an eye tracking system, a heart rate monitor and an electrodermal activity device ($N = 25$ participants). Four different types of highways (driving scenario of 40 minutes each) are implemented through the variation of the road design (amount of curves and hills) and the roadside environment (amount of buildings and traffic).

This paper will first introduce the background of the research, and then provide a description of the experimental design. Then the results of the evaluation of the driver's alertness on the different highways will be presented. The last part of this paper will discuss the implications of the results of the this analysis.

2 Background

2.1 Measuring alertness

The most reliable and reproducible way to measure the alertness of a driver driving is to use an EEG [8–10]. EEG signals are analysed in the frequency domain, and four different bands contain the information: α, β, θ and δ. The most reliable method to measure alertness variation is to use the following algorithm: $\frac{\theta+\alpha}{\beta}$. When increasing, this ratio between slow and fast wave activities indicate a decrement of alertness [11, 12]. Bursts are also of interest to detect increments in bands occurring relatively sparsely. It can particularly be used to detect microsleep following alpha and theta activities [13].

An EEG device cannot be used in a vehicle for at least three reasons: (*i*) the inconvenience for the driver, (*ii*) the prohibitive cost, and (*iii*) the noise introduced due to electromagnetic field interferences. Nevertheless, such a device can be used in a laboratory-based experiment so that correlation with driving performance (observed variables from the driver the car and the environment) can be isolated and investigated.

2.2 Surrogate measures

Surrogate measures correlated to the alertness level during this study have been investigated in [14]. They are summarised here:
- ECG (heart rate, inter-beat intervals, heart rate variability)
- Eye activity (blink frequency, eye closure)
- Head rotation (toward the side)
- Electrodermal activity (Skin conductance Level SCL, Non specific fluctuation rate, rise-time and half-recovery time)
- Simulator data (lane lateral shift variability, speed variability, steering wheel variability, time to lane crossing)

2.3 Evaluation of alertness using neural networks

Neural networks are composed of a network of simple processing elements called neurons, which can exhibit complex global behaviour, determined by the

connections between the processing elements and element parameters. An elementary neuron computes an output from the inputs as follows: first each input (here the observations such as reaction times) is weighted appropriately, then the sum of the weighted inputs and a bias form the input to a transfer function f which provides the output (here the accuracy) [15].

NNs are capable of approximating any function with a finite number of discontinuities. Training the neural network consists of adjusting the values of the connections (weights) between elements. This permits the modelling of the complex relationship between the relevant measurements Obs_t (inputs) and the alertness $Alert_t$ (output) at time t. Particularly, multilayer feedforward networks can model such complex, non-linear relationships and when properly trained, provide reasonable answers when presented with totally new inputs. In such a network, inputs are processed through successive layers of neurons, the information moving in only one direction, forward, from the input nodes (taking the variables of the problem as inputs), through the hidden nodes (if any) and to the output node (whose outputs should be as close as possible to the observed outputs, using the MSE as a cost function), without any cycles or loops in the network. These networks are trained by backpropagation, which is a gradient descent algorithm in which the network weights are moved along the negative of the gradient of the performance function [16].

3 Methods

3.1 Participants

A stratified random sampling approach was used to obtain a representative population of licensed drivers (for at least two years), regular drivers from different age groups (as per categories used in road safety i.e. 18–24, 25–59 and 60+). The 60+ category was not targeted in this study due to vision impairments and possible circadian and cognitive functioning changes related to ageing [17].

Twenty-five subjects aged between 18 and 49 (mean age = 29.1 years, SD = 8.3) volunteered for this study. Thirty participants were expected to drive in this experiment but five subjects were removed from the sample due to motion sickness which occurred during training on the driving simulator.

Young drivers were recruited from Queensland University of Technology (QUT). Other participants were selected from staff at QUT and the general community. Participants had their licence for a minimum of two years and drove a minimum of three days per week similar to previous research [18]. All subjects provided written consent for this study which was approved by QUT ethics committee. Participants were paid AUS $80 for completing the four driving sessions; students undertaking the first year psychology subject received course credit for their participation.

3.2 Experimental design

Both road design and roadside environment of the highway were varied in this experiment. The combination of low and high variability of both parameters

Table 1: The four experiment scenarios.

		Roadside variability	
		low	high
Road design variability	low	scenario 1	scenario 2
	high	scenario 3	scenario 4

resulted in four different scenarios (see Table 1). Scenario 1 is characterised by low road design variability and a low roadside variability. Roadside variability is changed to high for scenario 2, while road design variability is increased for scenario 3. Scenario 4 is done with both road design and roadside variability high.

In each experiment, the participants were asked to drive and follow road rules for approximately 40 minutes. Each participant is tested on each scenario (repeated measures design) in the following conditions:

- driving consists of following a lane (no itinerary involved) at constant speed (60 kilometres per hour), without having to stop the car (no red traffic lights or other stops) or to press the brakes frequently (no T intersections or perpendicular turns)
- no manual gear changes were required
- no use of indicators was required
- low traffic conditions (no congestion).

Road geometry is varied through the curvature of the road as well as its altitude. In the road design with low variability, the road is essentially straight or with little curves and flat. In the road design with high variability, the road is a sequence of small straight sections, significant curves and hills.

The roadside environment is varied in terms of scenery (low versus high variability). Low roadside variability is composed of a desert-like scenery with bushes along the road with periodic traffic occurring in the oncoming lane only. This models an Australian rural highway. High roadside variability is composed of various buildings, roadside barriers, overhanging lights, overpasses and trees/foliage with traffic surrounding the vehicle (no congestion and no requirements to overtake). This represents Australian urban highways.

3.3 Experimental conditions

Participants were tested individually in a quiet room in four sessions lasting approximately one hour per session. Each participant drives in one of the four scenarios (randomly assigned) in the simulator once a week for four weeks at a fixed testing time. Testing times are scheduled at 9am, 11am, 1pm and 3pm. Each participant chooses a testing time for which they feel they are the most alert.

A short practice is performed to familiarise participants with the driving task on the simulator permitting the setting up of sensors for the experiment at the same
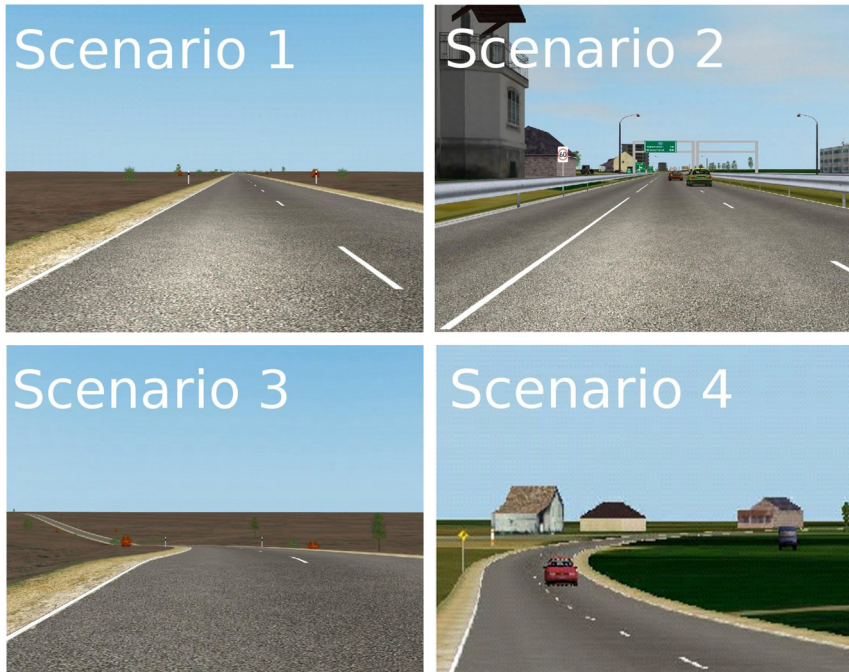
Figure 1: Screenshots of the 4 scenarios.

time. Next, participants performed their scenario and at the end of the experiment they answer questions about their level of alertness to check for the absence of fatigue effect. Participants are also asked not to consume alcohol 12 hours prior to the experiment.

3.4 Materials

3.4.1 Driving interface
Experimentation was conducted on the driving simulator Scaner from OKTAL. The participant sits in front of a screen where the Scaner simulator is played by a RGB video projector. The simulator displays a view from the inside of the vehicle with a speedometer. The participant drives the simulator using a modified computer steering wheel which provides force feedback and a two pedal set (brake and accelerator only). Five speakers reproduce the acoustics environment of inside a car.

3.4.2 Sensors
Data related to the vehicles dynamics or the environment is collected by the driving simulator software. Data related to the driver are collected with:
- Bioradio which provides data related to EEG and ECG
- Biopac which provides data related to skin conductance

- Facelab which provides data related to the driver's eyes (eye movements, blinks, etc.).

3.4.3 Synchronisation interface

Data collected from the simulator and the different sensors are synchronised using RTmaps. This software records and time stamps data from different devices and different computers.

3.5 Data analysis

3.5.1 Alertness measure

Driver alertness is assessed through analysis of data collected with the EEG. EEG data is collected at 7 different positions on the scalp (O1, O2, T5, T6, P3, P4 and F3) following the *International 10–20 Electrode Placement System* at 80 Hz and are divided into 1 second epochs.

Epochs with too high/low values (threshold $\pm 75\,\mu V$), linear trends, improbable data and/or abnormally distributed data are rejected. A 4-term Blackman-Harris window and a 0.5 Hz cut-off high-pass filter were also used to reduce low frequency artefacts. Fast Fourier Transform (FFT) is performed and this provided α, β, θ and δ band activities. $\frac{\alpha+\theta}{\beta}$ is computed for each selected epoch. Bursts for α and θ activities are also computed (threshold 1.5 above the mean). The five first minutes are used as a reference of normal driving behaviour for comparison with the alertness impairment throughout the driving experiment. Epoch values above two standard deviations of the mean of the first five minutes are categorised as indicating a reduced alertness ($\frac{\alpha+\theta}{\beta}$ ratio above average).

The proportion ρ of epochs indicating low alertness or microsleeps is calculated over a time window of one minute. This is used as a measure of alertness by calculating the complementary proportion $Alert(t) = 1 - \rho(t)$. Alertness is then separated into discrete levels (as suggested by the data collected in this experiment) as follows:

- Alert: 0.75 to 1%
- Reduced alertness: 0.55% to 0.75
- Low alertness: 0 to 0.55

3.5.2 Surrogate measures

ECG data are recorded at 80 Hz and are used to automatically extract the heart rate and inter-beat-interval (IBI). Thresholds to detect peaks are manually adapted in each session for each participant. Unrealistic values obtained for IBIs are filtered using 500 and 1300 ms as lower and upper limits respectively.

Eye activity data are collected at 60 Hz. Blink frequency and eye closure are extracted by Facelab. This device also furnishes data about the driver head movements, in particular the rotation of the head to the side.

Electrodermal activity is collected at 1 Hz. Skin conductance level and non-specific fluctuation rates, rise-time and half-recovery time are extracted. A threshold of $0.02\,\mu S$ is used to find non-specific responses.

Car and environment variables are obtained from the simulator. Data collected from the simulator is sampled at 20 Hz. Lane lateral shift variability, speed variability, steering wheel movement variability and Time to Lane Crossing (TLC) were used in this analysis. Only straight sections of road were used to compute these metrics.

These variables are normalised for each participant using the five first minutes of driving for each of the four sessions. Due to the high number of surrogate measurements, a Principal Component Analysis (PCA) is performed to reduce the number of inputs to the mathematical model. PCA extracts relevant information from complex datasets and reduces possible correlated variables into a smaller number of uncorrelated variables called principal components. Principal components are ordered as a function of their explanation of data variability [19]. Only components explaining more than 5% of the variance in the data are selected as predictors for the neural networks.

3.5.3 Neural networks

Neural network are trained on part of the data (training set), the remainder (test set) being used for testing the accuracy of the model in a robust way. Participants driving only once on a specific scenario, have their data used both for training and testing for the model accuracy. For each time t ($t > 2$ min) data up to time $t - 1$ is used to train the model and then data at time t is used to assess alertness at that time. This result is compared to the alertness at time t obtained from the EEG.

Neural networks were fitted for this experiment using the tan-sigmoid transfer function and the Levenberg-Marquardt backpropagation algorithm.

NNs were fitted in two different ways. First NNs were fitted with a discrete target output. Target output was the alertness level obtained with the EEG (three categories) and was coded as a vector in three dimensions. The alertness level inferred from the model is then the category (dimension) with the highest value. Next, NNs were fitted to model the frequency of alertness lapses during the one minute time window. These results were then categorised into alertness levels as detailed in section 3.5.1. Modelling with neural networks was optimised for each participant in order to obtain the best results. This optimisation is done through the variation of the number of time slices observed and the number of layers and neurons.

Assessment of the model's accuracy was performed as follows. The best model is the one with (*i*) the largest mean - between sensitivity and specificity in detecting the state of low alertness - in the case of a discrete output and (*ii*) the smallest MSE in the case of continuous output. Once the best model is selected for each method, continuous outputs are categorised so that both methods can be compared in their ability to detect reduced alertness.

4 Results

Principal Component Analysis resulted in the identification of six components explaining more than 5% of the variance. PCA_1 is mainly composed of the

Table 2: Comparison of the two methods to train NNs.

Method	Scenario	MSE (%)	Sensitivity (%)	Specificity (%)	Mean (%)
Discrete	1	-	87	88	88
	2	-	89	93	91
	3	-	96	95	96
	4	-	95	93	94
Continuous	1	1.33	88	38	63
	2	0.81	84	43	64
	3	1.05	95	41	68
	4	1.27	94	65	80

skin conductance level. PCA_2 can be reduced to blink frequency. PCA_3 is a combination of skin conductance measures (SCR rise-time and SCR half-recovery time). PCA_4 is the result of a selection of ECG metrics while PCA_5 combines driving performance (lane keeping, steering wheel movement) to skin conductance measures (NSF rate and SCR rise-time). Finally PCA_6 is largely composed of the head rotation.

NNs were fitted as described in the method section. A comparison of means shows that the first method (discrete) provided better estimates of the state of alertness (see Table 2). Actually, both methods are approximately reliable when their result is a low alertness state (around 90%). The former method is as good in detecting lapses in alertness (around 90%) while the latter misses half of them (detection around 50%). Neural networks were able to detect reduced alertness with an 88% (96% respectively) accuracy on a straight (curvy respectively) rural highways. NNs were 91% (94% respectively) accurate for straight (curvy respectively) urban highways.

5 Discussion

This study have shown that Neural networks suit the driving task complexity. It provides estimates of driver alertness in the range of what is required for the development of an in-vehicle device predicting driver lapses in alertness. Such results are obtained after optimisation for each participant, highlighting the need to adapt to each driver behaviour. Using personality traits groups (sensation-seeking level, extroversion) in the model is not sufficient in obtaining a reliable model.

Alertness decrement was observed for all different highways modelled in this study. Such alertness decrement resulted in driving impairment which has been

accurately used as a surrogate measurement to follow driver alertness. Detection is accurate on various highways, as underlined by the variation used in this experiment for both road design and roadside environment.

This experiment also shows similar results in terms of sensitivity and specificity around 90%. This underlines the high level of accuracy of the detection of episodes of low alertness, as well as the low level of false alarms when using this methodology.

During this simulated experiment, it was possible to assess in real-time driver's alertness decrement accurately through surrogate measures. Such modelling supports the idea of designing a countermeasure against crashes due to alertness impairments on urban and rural highways using ITS devices detecting in real-time such alertness decrements.

Acknowledgements

References

[1] Cervero, R., *The transit metropolis : a global inquiry*. Island Press: Washington, D.C., 1998.

[2] Buchanan, N., Barnett, R., Kingham, S. & Johnston, D., The effect of urban growth on commuting patterns in christchurch, New Zealand. *Journal of Transport Geography*, **14(5)**, pp. 342–354, 2006.

[3] Ma, K.R. & Kang, E.T., Time-space convergence and urban decentralisation. *Journal of Transport Geography*, 2010. Doi: 10.1016/j.jtrangeo.2010.06.016.

[4] Levinson, D. & Wu, Y., The rational locator reexamined: Are travel times still stable? *Transportation*, **32(2)**, pp. 187–202, 2005.

[5] Taylor, M.A.P. & Ampt, E.S., Travelling smarter down under: policies for voluntary travel behaviour change in Australia. *Transport Policy*, **10(3)**, pp. 165–177, 2003.

[6] Goudie, D., Zonal method for urban travel surveys: sustainability and sample distance from the cbd. *Journal of Transport Geography*, **10(4)**, pp. 287–301, 2002.

[7] Thiffault, P. & Bergeron, J., Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, **35(3)**, pp. 381–391, 2003.

[8] Lal, S.K.L. & Craig, A., Reproducibility of the spectral components of the electroencephalogram during driver fatigue. *International Journal of Psychophysiology*, **55(2)**, pp. 137–143, 2005.

[9] Pollock, V.E., Schneider, L.S. & Lyness, S.A., Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. *Electroencephalography and Clinical Neurophysiology*, **79(1)**, pp. 20–26, 1991.

[10] Tomarken, A.J., Davidson, R.J., Wheeler, R.E. & Kinney, L., Psychometric properties of resting anterior EEG asymmetry: Temporal stability and internal consistency. *Psychophysiology*, **29(5)**, pp. 576–592, 1992.

[11] Lal, S.K.L., Craig, A., Boord, P., Kirkup, L. & Nguyen, H., Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research*, **34(3)**, pp. 321–328, 2003.

[12] Bastien, C.H., Ladouceur, C. & Campbell, K.B., EEG characteristics prior to and following the evoked K-Complex. *Canadian Journal of Experimental Psychology*, **54(4)**, pp. 255–265, 2000.

[13] Eoh, H.J., Chung, M.K. & Kim, S.H., Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, **35(4)**, pp. 307–320, 2005.

[14] Larue, G.S., Rakotonirainy, A. & Pettitt, A.N., Driving performance on monotonous roads. *20th Canadian Multidisciplinary Road Safety Conference*, Niagara Falls, Canada, 2010.

[15] Russell, S. & Norvig, P., *Artificial Intelligence - A Modern Approach*. Prentice Hall series in artificial intelligence, Pearson Education, 2nd edition, 2003.

[16] Rumelhart, D.E., Hinton, G.E. & Williams, R.J., Learning representations by back-propagating errors. *Nature*, **323(6088)**, pp. 533–536, 1986.

[17] Blatter, K. & Cajochen, C., Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*, **90(2-3)**, pp. 196–208, 2007.

[18] Campagne, A., Pebayle, T. & Muzet, A., Oculomotor changes due to road events during prolonged monotonous simulated driving. *Biological Psychology*, **68(3)**, pp. 353–368, 2005.

[19] Shaw, P.J.A., *Multivariate Statistics for the Environmental Sciences*, 2003.