VIENNA 2018 TRA
A digital era for transport
solutions for society, economy and environment

Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Analysis of Driving Behaviour Characteristics Based on Smartphone Data

Eleonora Papadimitriou[a*], Dimitris Tselentis[a], George Yannis[a]

[a]Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Heroon Polytechniou str., GR-15773 Athens

Abstract

The objective of this paper is to detect and analyze risky driving behaviour characteristics on the basis of smartphone data, with focus on key risk indicators, namely the number of harsh driving events and the use of mobile phone while driving. Driving behaviour analytics data from a naturalistic driving experiment are exploited in this research recorded by smartphone devices. The driving indicators that are collected include distance travelled, speed, accelerations, brakings, turnings, cornerings, and related 'events' in the form of harsh maneuvers (e.g. harsh acceleration, braking, etc.), as well as mobile phone use. One hundred drivers participated in the designed experiment during a 4-months timeframe and a large database of 18,850 trips was built. The results of this research reveal that distraction originating from smartphone usage has a serious impact on the number of harsh events that occur per kilometer and subsequently on the relative crash risk. Furthermore, mobile phone use while driving may be accurately "detected" by smartphone sensors data in more than 70% of cases.

Keywords: Driving Behaviour; Big Data; Smartphone Data.

1. Introduction

Accurate monitoring of driving behaviour has been proved a difficult task in the past. The rapid technological progress, especially in telematics, and Big Data analytics, along with the increase in the information technologies' penetration and use by drivers (e.g. smartphones), provide new potential for driving behaviour monitoring and analysis. First results from related applications (Tselentis et al. (2017), Theofilatos et al. (2017), Araújo et al. (2012), Enev et al. (2016), Vlahogianni and Barmpounakis (2016)) have confirmed the feasibility, efficiency and usefulness of such big data collection schemes.

A significant number of risk factors that affect the probability of participating in a road traffic accident have been identified in literature. Among others, the most important risk factors recognized in literature (WHO 2015) are human factors such as speeding, distracted driving, number of harsh acceleration and braking events etc. Human factors are considered one of the main causes of road traffic fatalities and injuries every year and therefore it is highly important to study how these factors can affect traffic risk.

Existing studies have shown promising results as regards the identification of some risk factors (i.e. speeding, aggressiveness etc.) through smartphone data collection and processing. However, to the best of the authors' knowledge, there has not been an attempt to detect risks related to the use of the mobile phone while driving, based on data from the smartphone sensors. The most common methodologies applied for the assessment of risks related to mobile phone use while driving are: (i) driving simulator experiments (Papantoniou et al. (2014)), (ii) naturalistic driving experiments (Simmons et al. (2016)), and (iii) contributory factors analysis of actual crash records (Consiglio et al. (2003), Elvik (2011), Backer-Grøndahl and Sagberg (2011)). Each method presents different advantages and limitations; nonetheless, results are fairly consistent regardless of the study method.

Mobile phone use while driving is persistently shown in literature to have significant detrimental effects on driver behavior and safety, due to the higher level of workload involved in such multi-tasking, regardless of conversation difficulty level (Haque and Washington (2015)). Since mobile usage is a standard part of everyday driving process and is expected to increase over the years [Stutts et al. (2001)], its impact on driving behaviour in traffic and road safety is particularly important and should be further investigated. Literature so far has showed that when the driver is using the phone while driving his/her behaviour alters significantly. Therefore, mobile usage is banned in many countries [Carsten and Brookhuis (2005)] as the distraction caused is considered the main risk while driving [Brookhuis et al. (1991), Department for Transport, (2008), Charlton (2009)]. Mobile phone use results in higher speed variation (Haque and Washington (2015)) and difficulties in maintaining vehicle lateral position (Horrey and Wickens (2006)). Lower driving speed is observed, and as a result increased headways, these suggesting possible risk compensatory behavior (Haque and Washington (2015), Horrey and Wickens (2006)). Nevertheless, drivers' reaction times increase significantly when conversing at mobile phone (Consiglio et al. (2003)), and little no benefit of hands-free over handheld mobile phone use has been validated (Caird et al. (2008), Törnros and Bolling (2006)). Overall, there is a statistically significant increase in crash risk, which is almost three times the risk run when a mobile phone is not used (Elvik (2011), Backer-Grøndahl and Sagberg (2011)).

Speeding is another very important factor that affects accident probability (e.g. reaction distance reduction, loss of control) and crash impact. According to (OECD, 2006) speeding has been a contributory factor in 10% of the total accidents and more than 30% in fatal accidents. According to (Nilsson 1982) the probability of a crash involving an injury is proportional to the square of the speed, the probability of a serious crash is proportional to the cube of the speed and the probability of a fatal crash is related to the fourth power of the speed. Finally, harsh events such as acceleration, breaking and cornering are three significant indicators for driving risk assessment (Bonsall et al. 2005) especially for evaluating driving aggressiveness. These attributes are strongly correlated with unsafe distance from adjacent vehicles, possible near miss accidents, lack of concentration, increased reaction time, poor driving judgement or low level of experience and involvement in situations of high risk. The correlation between harsh acceleration (HA) and harsh braking (HB) events with driving risk has been highlighted in the scientific papers published by (Tselentis et al., 2017, Bonsall et al. 2005) and it has been widely recognized by the insurance and telematics industry.

The objective of this paper is to detect and analyze driving behaviour characteristics on the basis of smartphone data, with focus on key risk indicators, namely the number of harsh driving events and the use of mobile phone while driving. More specifically, by continuously collecting data from smartphone devices while driving, this study aims to examine the way that driving metrics recorded such as harsh events (braking, acceleration, cornering) are influenced by driving distraction in the form of mobile phone usage and therefore predicting the number of harsh events that take place. The driving measures that are collected include distance travelled, mobile usage (dialing, talking, texting etc.), speed, accelerations, braking, steering, cornering, as well as related 'events', in the form of harsh maneuvers (e.g. harsh accelerations, harsh braking, harsh cornering, etc.). Furthermore, the predictability of mobile phone usage while driving through recording of driving related metrics is also examined herein.

The remaining of this paper is structured as follows: Section 2 describes the data collection methodology through smartphone sensors that was used to monitor a sample of one hundred drivers who participated in the related naturalistic driving experiment. The statistical analysis carried out on the basis of mixed effects binary logistic regression techniques, for modelling the use of mobile phone while driving on the basis of other driving behavior indicators is presented in detail in section 3. Final results of the statistical analysis conducted are demonstrated and discussed in section 4. Finally, section 5 includes a final discussion of the findings of this research.

2. Data collection and processing procedure

An innovative data collection scheme using a Smartphone Application that has been developed by OSeven was exploited for the purpose of this research. Driving behaviour analytics is recorded in real time, using smartphone device sensors. One hundred drivers participated in the designed experiment during a 4-months timeframe and a large database of several thousand trips is created. All the data were received in anonymized format without any personal information. The solid integration platform for collecting, transferring raw data and recognizing the driving behaviour metrics via ML algorithms is also developed by OSeven. This ensured a smooth transition from the data collection to the data analysis procedure. Recorded data come with zero user involvement (the app starts and stops recording automatically) every time that a driver is using the vehicle. Data are derived from various smartphone sensors and data fusion algorithms provided by Android (Google) and iOS (Apple) and transmitted to the central database via a Wi-Fi or cellular network. The steps of the standard procedure developed that is followed every time a new trip is recorded by the App, are clearly shown in Figure 1.

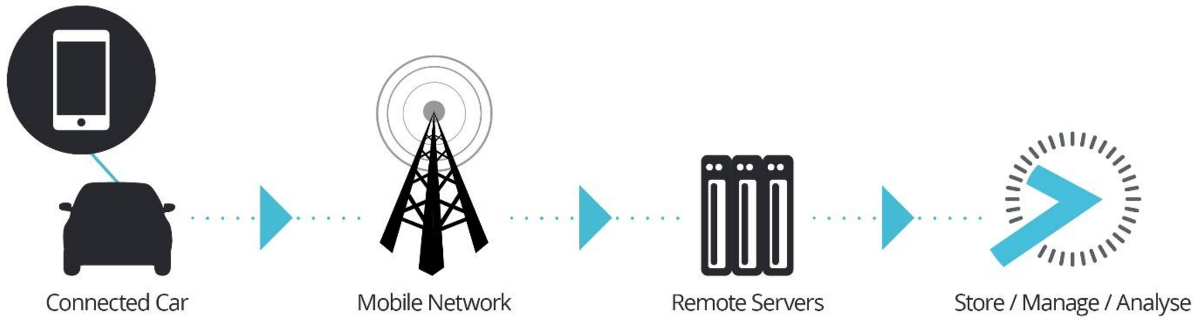


Fig. 1 OSeven data handling chart

The frequency of the data recording varies depending on the type of the sensor with a minimum value of 1Hz. After data is stored in the cloud server machine learning (ML) methods and Big Data mining techniques are applied. Data are filtered and smoothened when necessary, cleaned from existing noise and finally errors and outliers are recognized. Within each trip, speeding regions, harsh acceleration/braking/cornering events, mobile usage and risky hours driving are also detected and finally, driver/passenger recognition and transportation mode detection is carried out. Subsequently, the critical risk exposure and driving behavior indicators arising are considered in the driving risk analysis and more specifically into the driving scoring model developed.

A variety of different indicators are calculated after ML process that are useful to the user or/and the evaluation of the travel behavior, such as:

- Total distance (mileage)
- Driving duration

- Type(s) of the road network used (given by GPS position and integration with map providers e.g. Google, OSM)
- Time of the day driving (Rush hours, Risky hours)
- Weather conditions
- Trip purpose (set by the driver by using the smartphone app and estimated by the app after the driving pattern has been identified)

The driving behavior indicators that are also calculated from the data include indicatively:

- Speeding (duration of speeding, speed limit exceedance etc.)
- Number and severity of harsh events:
  - Harsh braking (longitudinal acceleration)
  - Harsh acceleration (longitudinal acceleration)
  - Harsh cornering (angular speed, lateral acceleration, course)
- Driving aggressiveness (e.g. braking, acceleration)
- Distraction from mobile phone use (the determination of the mobile use, e.g. talking, texting is based on the movements recorded by the mobile sensors and not on the use of several apps)

These indicators along with other data (e.g. data from maps) are subsequently exploited to implement individual driver's statistics, on all road networks (urban, rural, highway, etc.) and under various driving conditions, enabling the creation of a large disaggregate database of driving characteristics.

This study is investigating the macroscopic driving characteristics within a trip and as a result all indicators that were taken into consideration such as harsh events and mobile usage might have not been recorded simultaneously and therefore are not investigated as such.

3. Methodological Approach

As stated above, driving measures collected include indicatively distance travelled, speed, accelerations, braking, steering, cornering and smartphone usage (dialing, talking, texting etc.) in different driving environments (urban, rural, highway). During data processing, new variables were created in order to define the time of the day driving (daylight, morning rush, afternoon rush).

In the present dataset, there are repeated measurements (trips) made over the same units (drivers). These repeated measurements make that the observations are no longer independent, as require the assumptions of regression, and unless accounted for, this dependency may affect the accuracy of the modelling results. In fact, it is necessary to account for random heterogeneity due to differences between drivers, to make sure that the effects identified in the models are true effects of the independent variables on the dependent, and do not reflect unobserved differences between drivers. A mixed model (or random effects model, or multi-level model) is a standard technique in this context, i.e. a statistical model containing both fixed effects and random effects. The formulation of the linear mixed effects model, assuming a random intercept reflecting the repeated measurements (i) over drivers (j), is as follows:

$$y_{ij} = \beta_{0j} + \Sigma\beta_i x_{ij} + e_{0ij} \tag{1a}$$

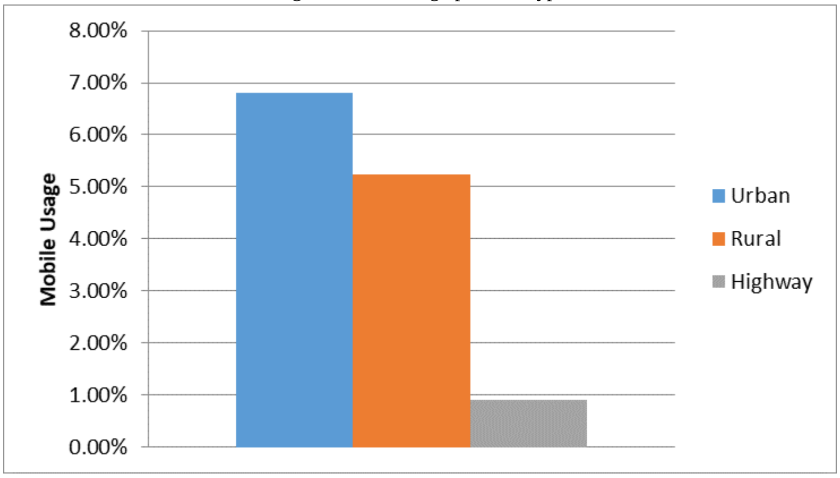
$$\beta_{0j} = \beta_0 + u_{0j} \tag{1b}$$

It is noted that the intercept in the outcome equation consists of two terms: a fixed component β_0 and a driver-specific component, i.e. the random effect u_{0j} assumed to be normally distributed.

A mixed linear model is developed to model the driving characteristics that influence the number of harsh events per distance travelled including mobile phone usage that is one of the components of driver distraction. The influence that each variable has on the number of harsh events occurring in each trip separately are quantified and consequently, the relative risk for each trip can be identified.

Figure 2 shows the average mobile phone usage percentage (duration of mobile usage / driving time) per road type demonstrating again a lower percentage of mobile usage for highways.

Fig. 2 Mobile Usage per road type



A mixed binary logistic model is also utilized to predict the situation of using or not the mobile phone while driving through the observation of different driving measures. The variable of interest in the present analysis is the use of mobile phone while driving. This was available either as a share of trip time during which mobile phone was used, or as a binary variable (yes / no). The latter case was selected for modelling in the present research. Typically, a binary logistic regression estimates the probability that a characteristic is present (e.g. estimated probability of "success") given the values of explanatory variables; $\pi = Pr\ (y = 1|X = x)$. It leads to the development of a mathematical model that gives the odds of this event occurring, depending on some factors that affect it. The odds are expressed by the logit link function as follows:

$$logit(\pi_i) = logit\,\frac{\pi_i}{1-\pi_i} = \beta_0 + \Sigma\beta_i x_i \quad (2a)$$

And the related outcome (event occurrence):

$$y_i = \beta_0 + \Sigma\beta_i x_i + e_{0i} \quad (2b)$$

The trip specific error term e_{0ij} is assumed to follow a logistic distribution.

4. Results

Overall, a change in driving behaviour and more specifically in the number of harsh events occurred such as harsh braking, acceleration and cornering is proved to be predictable using mobile phone usage while driving as indicator. The mixed linear regression model that originated from the above analysis, illustrates a significant dependence between driver's aggressiveness (total harsh events / total distance) and percentage of mobile usage, the average speed, the average exceedance of the speed limit as a percentage of the speed limit, the driving period during a day (morning, afternoon rush). It is noted that more frequent use of mobile phone is associated with fewer harsh events, suggesting smoother driving as a compensatory behaviour of mobile phone use. Among all models tested the most explanatory and statistically significant is presented in Table 1.

Table 1. Mixed linear Regression model output for the estimation of Harsh Events per trip.

Parameter estimates	B	p-value
Constant	.517	<0.001
Driving during morning rush hour	.018	.010
Average Speed	-.005	.007
Average Speed limit exceedance	.294	<0.001

Mobile Usage % time	$-3.36*10^{-5}$	.010
Random effect (variance of random intercept)	0.021	<0.001

In the mixed logistic model, the categorical dependent variable is the use or no use of mobile phone while driving, whereas the explanatory variables are the driving behavior and exposure metrics described previously. The subject of which there are repeated measures (trips) is each driver. Final model is presented in table 2.

Table 2. Parameter estimates of the mixed binary logistic models for all road types

Parameter estimates		
Fixed effects	B	P-value
Constant	1.094	<0.001
Morning rush	0.130	0.006
Afternoon rush	-0.262	<0.001
Average percentage of speed over the speed limit	-0.334	0.027
Average speed	-0.004	0.001
Average angular speed	-0.058	<0.001
Total Harsh events	-0.064	<0.001
Random effect (variance of random intercept)	1.261	<0.001

The modelling results reveal the following:

- Driving during morning rush hours increases the probability of mobile phone use during the trip; this is not surprising, as drivers may make / receive more calls during morning rush hours (e.g. work related).
- Accordingly, driving during afternoon rush hours reduces the probability of mobile phone use during the trip.
- The length of the trip (time driving) was not found to affect the probability of mobile phone use.
- Average speed per trip was found to be negatively associated with the probability of mobile phone use, confirming existing studies.
- However, the effect of exceeding the speed limit was found to be more sensitive than that of average speed. The average percentage exceedance of speed limits reduces the probability of mobile phone use; in general, drivers who are speeding more, are less likely to use their mobile phone during the trip.
- The higher the number of harsh events per trip distance, the lower the probability of mobile phone use; this is also intuitive, as the literature suggests that drivers reduce speed while distracted, and therefore are less prone to harsh events. It is also in accordance with the linear regression results.
- The variable average angular speed (measured in °/s) reflects the amount of smooth cornering during the trip. This particular type of harsh event was found statistically significant, suggesting that the higher the angular speed, the lower the probability of mobile phone use.

Table 3 presents the classification of outcomes as per the final model for all road types. It can be seen that more than 70% of actual cases where mobile phone was used during a trip are correctly classified by the model. "False positives", i.e. cases falsely classified as mobile phone used are 28.6%.

Table 3. Outcomes Classification Table for all road types

Mobile Phone use	**Predicted (%)**	
Observed	No	Yes
No	71.4	28.6
Yes	29.9	70.1

It is noted that the "false positives" are very minor share of the classified cases, suggesting that the driving metrics can very accurately identify "not talking on mobile phone" conditions, but not so accurately the "talking on mobile phone" conditions.

5. Discussion

This paper primarily aimed to investigate the possibility of detecting risky driving behaviour on the basis of driver exposure and behavior metrics collected by smartphone sensors, focusing on harsh driving events and the use of mobile phone while driving. The latter question is particularly important as mobile phone use while driving is known as a major and persistent risk factor (alongside speeding, alcohol, etc.). Given that the penetration and use of mobile phones is expected to further increase, together with their numerous emerging functionalities and apps, driving risks may also increase. However, this increase in the use of modern smartphones, and the rapid digitalization of so many every day activities, may be also seen as an opportunity to exploit the wealth of data that can be made available by smartphone sensors, and the new possibilities for data transmission and processing, to identify new ways of mitigating the risk factors and improve road safety.

The results of this research substantiate that distraction originating from smartphone usage has a serious impact on the number of harsh events that occur per kilometer and subsequently on the relative crash risk. Further analysis of the data collected is implemented through statistical methods and led to the quantification of this influence. The quantification of change in driving behaviour can constitute a measure of evaluating driver's performance and thus, it can potentially lead to the implementation of a driving risk model based on driving behaviour and degree of exposure. It may also contribute towards the practice of evaluating driver's traffic and safety behaviour as well as to classify drivers in different safety categories depending on their relative level of risk. Moreover, results of this research reveal that the data collection and handling scheme tested is a feasible and in several ways advantageous approach, providing a wealth of real-life data on driving behaviour and related risks (e.g. distraction, speeding), by means of 'portable' smartphone sensors and related applications. Further analysis of the data by means of statistical techniques is a promising field for further research.

The outcomes of this research will benefit industry and particularly the road and vehicle industry. By identifying the key risk factors in driver distraction accidents, vehicle manufacturers will be able to develop new systems that will directly improve the safety of vehicle occupants and other road users through primary and secondary safety features. One specific application area for the industry relates to the development of targeted advanced automatic driver distraction preventing systems. Furthermore, the opportunity of forecasting the use or no use of mobile phones according to the observed driving measures facilitates the detection of distracted drivers. It is found that it is feasible to predict mobile usage solely by some driving characteristics such as duration, time of the day driving, speeding etc. Finally, it is expected that considerable gains for the society can be achieved, since the stakeholders including policy makers and industry could rely on the results and recommendations regarding risk factors that appear to be critical for safe driving.

Acknowledgements

This research has exploited data provided by OSeven Telematics, London, UK.

6. References

Araújo, R., Igreja, Â., de Castro, R., & Araujo, R. E. (2012). Driving coach: A smartphone application to evaluate driving efficient patterns. In Intelligent Vehicles Symposium (IV), 2012 IEEE (pp. 1005-1010). IEEE.

Backer-Grøndahl A., Sagberg F. Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science Vol. 49, Issue 2, February 2011, Pages 324-330.

Brookhuis, K.A., de Vries, G., de Waard, D. (1991). The effects of mobile telephoning on driving performance, Accident Analysis & Prevention 23(4), 309-316.

Caird, J. K., Willness, C. R., Steel, P. & Scialfa, C. A meta-analysis of the effects of cell phones on driver performance. Accident Analysis & Prevention, Vol. 40, 2008, pages 1282-1293.

Carsten, O.M.J. and Brookhuis, K. (2005). The relationship between distraction and driving performance: towards a test regime for in-vehicle information systems, Transportation Research Part F, n.8(2), pp.75–77.

Charlton, S. (2009). Driving while conversing: Cell phones that distract and passengers who react. Accident Analysis and Prevention, n.41, pp.160–173

Consiglio, W., Driscoll, P., Witte, M. & Berg, W.P. Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. Accident Analysis & Prevention, Vol. 35, 2003, Pages 495-500.

Department for Transport, (2008). Reported road casualties, Great Britain 2008: Annual Report.
Elvik, R. Effects of mobile phone use on accident risk: Problems of meta-analysis when studies are few and bad. Transportation Research Record No 2236, 2011, Pages 20-26.
Enev M., Takakuwa A., Koscher K, and Kohno T. (2016) Automobile Driver Fingerprinting. Proceedings on Privacy Enhancing Technologies (1):34-51
Haque M.M., Washington S. The impact of mobile phone distraction on the braking behavior of young drivers: A hazard-based duration model. Transportation Research Part C: Emerging Technologies, Vol. 50, January 2015, Pages 13-27.
Horrey, W. J. & Wickens, C. D. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Human Factors Vol. 48, 2006, pages 196-205.
Papantoniou P., Papadimitriou E., Yannis G. Assessment of Driving Simulator Studies on Driver Distraction. Advances in Transportation Studies Vol. 35, 2014, Pages 129-144.
Papantoniou, P., Papadimitriou, E., Yannis, G. (2015). "Assessment of Driving Simulator Studies on Driver Distraction", Advances in Transportation Studies, Issue 35, pp. 129-144.
Simmons, S. M., Hicks, A. & Caird, J. K. Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis. Accident Analysis & Prevention Vol. 87, 2016, Pages 161-169.
Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A., (2001). The role of driver distraction in traffic crashes. Report Prepared for AAA Foundation for Traffic Safety.
Theofilatos A., Tselentis. I.D., Yannis. G., Konstantinopoulos M. (2017) ”Willingness – to - Pay for Usage-Based Motor Insurance”, Proceedings of the 96th Annual meeting of the Transportation Research Board, Washington, D.C, January 8-12, 2017.
Törnros J., Bolling A. Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 9, Issue 4, July 2006, Pages 298-306.
Tselentis, D.I., Yannis, G., & Vlahogianni, E.I. (2017). Innovative motor insurance schemes: a review of current practices and emerging challenges. Accident Analysis & Prevention, 98, 139-148.
Vlahogianni, E. I., & Barmpounakis, E. N. (2017). Driving analytics using smartphones: Algorithms, comparisons and challenges. Transportation Research Part C: Emerging Technologies, 79, 196-206.