Which driving performance parameters affect speeding? A naturalistic driving experiment

Panagiotis Papantoniou*[1], Christina Gonidi[2], Armira Kontaxi[2], George Yannis[2]

[1]*Department of Surveying and Geoinformatics Engineering, University of West Attica, Agiou Spyridonos St., 12243 Aigaleo, Athens, Greece*
[2]*Department of Transportation Planning and Engineering – National Technical University of Athens, 5 Heroon Polytechniou Str., GR-15773 Athens, Greece*

Email: ppapant@uniwa.gr
Email: cgonidi@mail.ntua.gr
Email: akontaxi@mail.ntua.gr
Email: geyannis@central.ntua.gr

Abstract

While speeding is considered as one of the most important road accident contributory factors, the objective of the present study is to investigate which driving performance parameters affect speeding. In order to achieve this objective a naturalistic driving experiment took place where almost 5.000 trips of 116 drivers were recorded. Then 3 linear regression models were developed aiming to investigate which factors affect speeding which is defined as the percentage of the trip that the driver speed was exceeding the speed limit. More specifically, an overall model is developed for all road types, as well as separate models for urban and rural road type, respectively. Results reveal correlations of speeding percentage with specific driving behaviour and exposure metrics, namely the average speed, the number of harsh acceleration events, the eco-friendly driving performance as well as driving on weekends. The aforementioned analysis results constitute interesting findings and can be very important in policy makers and enforcement authorities in order to deal with the high risk factor of speeding.

Keywords: speeding, mobile data, naturalistic experiment, average speed

[1] * Corresponding author. Tel: +30.6972014255;
E-mail address: ppapant@uniwa.gr

1. Introduction

Driver behaviour has been determined as the critical reason for 65%- 95% of total road accidents [1]. Aggressive driving behaviour parameters, such as harsh accelerations and decelerations, and their correlations with crash risk, have been investigated by the insurance industry [2]. Harsh events have been determined as strongly correlated with driving risk [3].

Speeding is considered as the most important road accident contributory factor as is a key factor in around 30% of road fatal accidents [4]. Some 40-50% of drivers drive faster than the recommended speed limit and 10-20% exceed the limit by more than 10kph. Not only does speeding make you more likely to get involved in an accident, it also increases the likelihood of severe injuries or fatalities [5] Speeding encompasses: excessive speed (driving above the speed limit) and inappropriate speed (driving too fast for the conditions, but within the limits). The main effects of speeding are:

- Excessive or inappropriate speed contributes to about one third of all fatal accidents.
- The level of exhaust emissions, fuel consumption and noise increase with speeding
- Speeding effects the quality of life of urban residents, especially the safe mobility of Vulnerable Road Users (VRU).

The complexity and lack of detailed recorded data as well as the lack of an analysis of the driving conditions under which an accident takes place, do not always allow for the objective estimation of each factor's significance. However, several in-depth accident studies showed that the driver solely or in combination with the other two factors (road network and vehicle) is the main cause of road accidents.

Furthermore, harsh events have a significant impact on energy efficiency as well. The difference, under terms of fuel consumption and gas emissions, between a safe (or calm) driver and an aggressive one is estimated to be greater than 40 % [6]. For this reason, with the intention of reducing the environmental footprint due to the road transport system, in recent years the idea of educating drivers to adopt a more environmentally friendly way of driving has been promoted. Such behaviour could be achieved by reducing harsh accelerations, harsh decelerations and harsh maneuvers [7]. The desired reduction in the emission of gases could be accomplished by finding the factors influencing aggressive behaviour through its analysis.

In [8], researchers were among the first to use an in-vehicle driving data recorder (DDR). DDR is small and light and can be used for 9.000 driving hours or 100.000 kilometres. The data are initially stored in a memory stick and can be later analysed using a personal computer. The data are initially recorded through a time-based system, in which they are temporarily stored. Afterwards, data are stored in a frequency-based system that is based either on the frequency of use or on the frequency of appearance. DDR is designed to record the driving conditions under ordinary road circumstances and analyse driver behaviour on the road, thus providing quantitative driving behaviour data [8].

Toledo et al. used the DriveDiagnostics, which is an in-vehicle data recorder with dimensions of 11x6x3cm and is charged using vehicle's battery. The system collects data such as vehicle's acceleration (in x, y, z coordinates), speed, position estimated through GPS, fuel consumption, total trip time, etc. [9]. In addition, Zaldivar et al. used a recording system, called On Board Diagnostics (OBD-II), the development of an application for Android smartphones, which collects vehicle information and detects car accidents. Using the Bluetooth sensor, the mobile phone is connected to the OBD-II device and retrieves information on the condition of the vehicle. In addition, it pinpoints the vehicle's speed and exact position via the GPS sensor [10].

Some of the previous methods of collecting driving behaviour data require high costs to be applied and may not yield objective results. The proliferation of smartphones and the various types of integrated sensors in them created a method of collection: a cheap and easy to install platform for detecting driver behaviour in naturalistic conditions, offering a low cost alternative to driving data collection. The evaluation of driver behaviour through experiments using smartphone data was found to be a very promising method, enabling the acquisition of a wealth of real-life data on driving behaviour and related risks such as distraction and speeding [11].

The exploitation of new technological advancements allows for driver monitoring through smartphone applications and the respective data collection and processing [12]. Smartphones have the advantage of being programmable and a wide array of sensors has now become standard equipment that can be utilized for transport studies (such as accelerometer, digital compass, gyroscope, GPS, microphone and camera) and enable sensing applications, even without user engagement [13]. Therefore, due to the high cost of installing and operating a system unit in the vehicle for data collection, the creation of an application for smartphones which provides data by exploiting their sensors is strongly supported Several studies have been carried out focusing on the spatial analysis of recorded road crashes, road characteristics and census variables, and ultimately on the export of road crash models to improve road safety. Due to the progress of the geographic information systems it is possible the analysis of road crashes can be conducted across different geographic units. However, there is no clear guideline on which geographic unit should be selected for the spatial analysis of road crashes [14].

The above studies and methodologies have greatly contributed to examining a series of factors related to driving speed. However, at the same time, they exploit a rather limited number of drivers, driving situations or variables. Thus, further research is necessary, using a larger number of drivers driving for longer periods in real time conditions, to further investigate which factors are significant predictors of driving speed.

Based on the above, the objective of the present research is to investigate which driving performance parameters affect speeding. In order to achieve this objective, a naturalistic driving experiment took place where almost 5.000 trips of 116 drivers were recorded and linear regression models were developed in order to estimate which performance parameters are correlated with speeding. The paper is structured as follow: In the next chapter the methodology is presented including data collection through the mobile application as well as the theoretical background of the analysis. Then the results include both descriptive statistics as well as the implementation of three linear regression models. Finally, conclusions are stated and some proposal for further research are presented.

2. Methodology

2.1 Data Collection

In order to achieve the research objective, an innovative smartphone application developed by OSeven Telematics (www.oseven.io) was exploited aiming to record driver behaviour using the hardware sensors of the smartphone device. One hundred and sixteen drivers participated in the related experiment during a 6-months timeframe and a large database of several thousand trips is created. The solid integration platform for collecting, transferring raw data and recognizing the driving behaviour metrics via machine learning (ML) algorithms is also developed by OSeven Telematics.

The standard procedure that is followed every time a new trip is recorded by the application is clearly presented in Figure 1. The data collected are highly disaggregated in terms of space and time. Once stored in the backend cloud server, they are converted into meaningful driving behaviour and safety indicators, using signal processing, ML algorithms, Data fusion and Big Data algorithms. Machine learning methods (filtering, clustering and classification methods) are mainly used to clean the data from noise and errors, and to identify repeated patterns within the data.

Fig. 1. **The OSeven data flow system**

A variety of different metadata are eventually calculated, including indicatively the following exposure indicators:

- 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)

The driving behaviour indicators that are also identified 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)
- Distraction from mobile phone use (mobile phone use is considered any type of phone use by the driver e.g. talking, texting etc.).

Finally, it must be noted that since privacy and security consist two critical principles in the field of telematics, the OSeven platform has very clear privacy policy statements for the end users covering the type of data collected, the reason data is collected for, the time that data is stored and the procedures for data security based on encryption standards for data in transit and at rest. All this is done using state-of-the-art technologies and procedures in compliance with standing Greek and European personal data protection legislation (GDPR).

2.2 Theoretical Background

As mentioned above, models of the linear regression family are developed herein, the theoretical background of which is given below. Simple linear regression aims to model the relationship between two quantitative variables, x and y, by fitting a linear equation to observed data. One variable is considered an explanatory variable, and the other is considered a dependent variable. The linear regression line has an equation of the form Y=α+b_i*X, where X is the explanatory variable and Y is the dependent variable. Parameter b is the slope of the line, and α is the intercept (the value of y when x is equal to 0).

Before attempting to fit a linear model to observed data, one should first determine whether there is a relationship between the variables of interest. This does not necessarily imply a causation between the two variables, but that there is some significant correlation between the two variables. A scatter plot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no correlation between the proposed explanatory and dependent variables (i.e., the scatter plot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model. A valuable numerical measure of two variables' association is the correlation coefficient, which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables [15 ,16].

In all linear regression models, the following steps are necessary. First of all, the values and the sign of the regression coefficients b must be explainable. Coefficient b indicates how much the dependent variable changes as the independent variable changes. Second, the constant coefficient a of the equation must be the lowest possible. The constant considers the parameters that have not been taken into account. Third, the correlation coefficient R^2 needs to be as high as possible. The coefficient R^2 indicates the percentage of the dependent variable that is explained by the independent variables. Fourth, the t-statistic must be higher than 1.7 for significance level of 5%. The t-statistic measures shows whether or not the initial hypothesis is rejected. Fifth, the relative influence ei*, used for quantifying the influence of each individual variable, should be checked, which allows for the comparison between the influence of different variables in a single model. In other words, ei is the elasticity value and ei* is the elasticity value normalized.

All these steps are checked for the three mathematical models developed in this research. After an appropriate number of data processing tests performed, this study concluded to the linear regression models presented in the next subchapters.

3. Results

3.1 Descriptive Statistics

Overall, during the 6-months timeframe of the experiment 49.018 trips from a sample of 116 car drivers have been recorded. Due to the General Data Protection Regulation (GDPR), drivers characteristics were not correlated with the present trips, thus they are not further included in the analysis.

Before proceeding to the core analysis of the research, explanatory descriptive analysis of the data is implemented, allowing for an overview of the examined risk factor that may affect speeding behaviour. It should be also highlighted that the key variable in this paper is the "speeding percentage" which is defined as the percentage of travel time that the driver exceeded the speed limit (% of travel time).

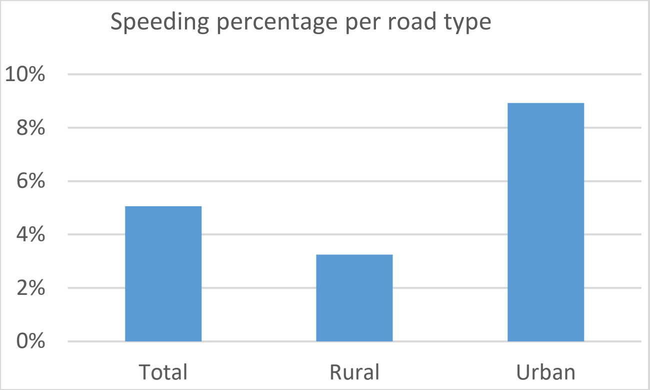

Fig. 2. **Speeding pertentage per road type**

In particular, it is found that the speeding percentage is higher in urban areas than in rural areas, which seems logical, given the pattern of driving on each type of road and the respective speed limit which is lower in urban areas that in rural ones.

Another interesting parameter that is found to affect the examined variable refers to the effect of the day of the week. For this purpose, two values were developed, week days referring to Monday to Friday and weekend referring to Saturday and Sunday (Figure 3).

Figure 3 indicates the speeding percentage is higher on weekend comparing the rest days of the week indicating the different driving behaviour (more aggressive) that occurs in the weekend.

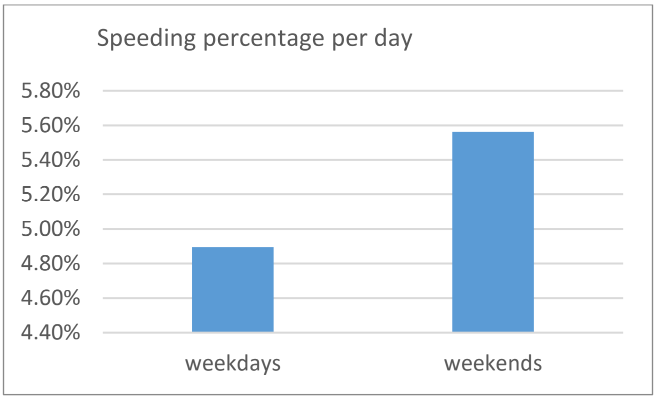

Fig. 3. **Speeding pertentage per day of week.**

Furthermore, one more observation is the effect of the hour of the trip on speeding. For this purpose, three values were investigated

- peak hours in the morning (06:00-10:00)
- peak hours in the afternoon (16: 00-20: 00)
- off peak hours

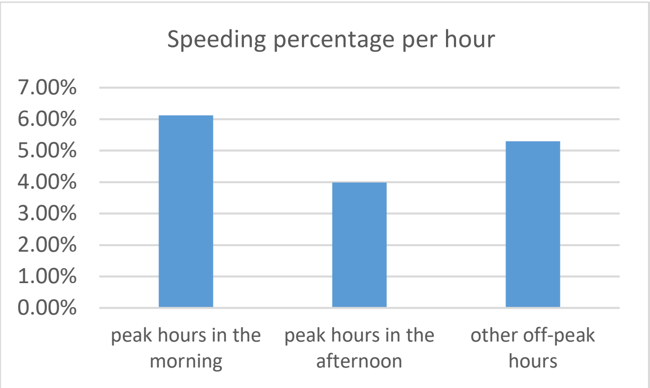

Fig. 4. **Speeding pertentage per hour**

This is probably due to the fact that peak hours in the morning, the drivers are in a hurry to go to work as a result of which speeding percentage. On the contrary peak hours in the afternoon, which is the time when people return from work are tired, want to relax from the stress of the day and speeding percentage less.

As presented above in the morning peak hours the percentage is higher. This is probably due to the fact that in the peak hours in the morning, the drivers are in a hurry to go to work. On the contrary, peak hours in the afternoon, which is the time when people return from work are tired, want to relax from the stress of the day resulting in a lower speeding percentage.

Finally, figure 5 illustrates the average percentage of speeding per all trips. It is evident that the vast majority of trips show a speed exceedance in the trip for less than 20%. However, it should be noted that there are a few trips where speeding reaches a maximum value equal to 60% of the total trip duration.

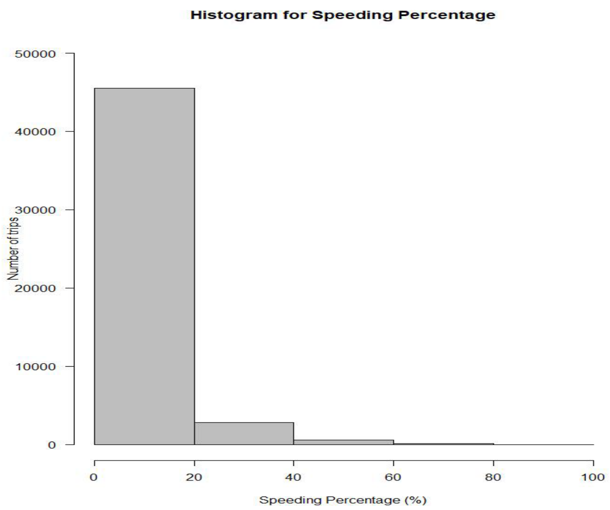

Fig. 5. **Speeding pertentage per trip.**

3.2 Linear Regression Models

In order to model the speeding percentage, models in a linear Regression framework were calibrated, as previously explained. An overall model is developed for all road types, as well as separate models for urban road type and rural road type.

The selected variables were chosen after taking into account the following: high statistical significance, low correlation among them, and final rational interpretation of their impact on the dependent variable.

Table 1 provides a description of the parameters that were found to be significant in the linear regression models.

Table 1: Description of the parameters used in the models

Independent Variables	Description
speed_avg	average speed (km/h)
harsh _acc	number of harsh acceleration events occurred
work_weekend	Workday (0=yes, 1=no)
smooth_eco	Eco-friendly driving

The three mathematical models developed are summarized below in Table 2. Results indicate that in all models, average speed and weekend exist as significant variables and that in all models, the very same variables have been set. All results are reasonably explained and confirmed by the findings of the existing literature.

The final models are presented in Table 2. The sign of "–" in the table indicates that the specific variable was not used in the particular model.

Table 2. Summary table of the three model developed

Independent Variables	Models					
	Overall		*Urban*		*Rural*	
	B	t	B	t	B	t
Constant	-0.119	-71.819	-0.215	-183.421	-0.053	-71.287
speed_avg	0.002	110.299	-	-	-	-
harsh _acc	0.006	33.368	-	-	-	-
work_weekend	0.009	10.574	-0.002	-2.013	-0.003	-3.354
smooth_eco	0.189	58.158	-	-	-	-
speed_urban_avg	-	-	0.010	275.838	-	-
harsh _acc_urban	-	-	0.003	10.748	-	-
speed_rural_avg	-	-	-	-	0.002	136.085
harsh _acc _rural	-	-	-	-	0.007	20.873
Adjusted R^2	0.316		0.630		0.310	

Modelling results regarding the speeding percentage indicate that average speed, number of harsh acceleration events occurred, have all been determined as statistically significant and positively correlated with the speeding percentage in all different models. In other words, as the average speed and the harsh acceleration events increase, the higher the speeding percentage in a trip. These two factors both indicate a stressful driving behaviour which is strongly correlated to speed exceedance while driving.

On the other hand, the eco-friendly driving variable is included only in the overall model indicating that the different road environment does not affect eco-driving performance in terms of speeding. More specifically, the lowest the value of this variable, the more eco-friendly the driving behavior. Therefore, the fact that that the present variable is positively correlated to the dependent variable, means that the worse the driving performance in terms of eco-driving, the higher the speeding percentage.

Furthermore, another parameter that is statistically significant in all models refers to the day of the week (weekdays referring to Monday to Friday and weekend referring to Saturday and Sunday). It is interesting that although in the overall model it is the weekdays that have a positive correlation to speeding, both in the urban and rural models, weekends indicates increase of speeding percentage.

4. Conclusions

This present research aimed to investigate which driving performance parameters affect speeding based on data collected by smartphone sensors. In order to achieve that objective, a naturalistic driving experiment was carried out in order to examine driving behaviour as expressed by speeding percentage.

Results reveal correlations of speeding percentage with specific driving behaviour and exposure metrics, namely the average speed, the number of harsh acceleration events occurred, the eco-friendly driving as well as driving on workdays.

One of the conclusions drawn here is that drivers who tend to accelerate harshly and frequently, also tend to exceed the speed limit in a greater amount of travel time. This is probably because drivers that accelerate more harshly, they are often the ones that have a higher average speed and therefore exceed the speed limits for a longer period of time. In addition, drivers who have higher average speeds, also have a higher speeding percentage.

It is also found that drivers traveling more in terms of distance and time, usually drive at a higher speed indicating that they are at higher exposure, behavioural risk and speeding percentage. For the majority of the drivers, driving behaviour on rural roads is found to be similar or better to that on urban roads. This is because when driving in urban road, users drive in general more abruptly and their average acceleration is higher.

Future research should also focus on the improvement of the accuracy of the models, by exploring more variables and alternative modelling techniques. Furthermore, with respect to speeding percentage, analyses per gender, age, history of accidents, self–assessment, driving experience and more demographic characteristics could be undertaken in order to capture any particular trends found in the categories of these parameters. In addition, it would be useful to exploit information collected from questionnaires on drivers' traits, years of driving experience, driving habits etc., in order to combine and compare it with the actual driving information collected from smartphones. It should be mentioned again that all these variables were not included in the present research due to the General Data Protection Regulation.

Moreover, one more limitation of the present research is that there are no data regarding trip on motorways. Considering that driving characteristics on motorways are different comparing to urban and rural network (different speed limit, road geometry etc.) a future work refers to the investigation of speeding behaviour on motorways.

Acknowledgment

This research is based in the DistrApp project which is co-financed by Greece and European Union (Social European Fund) through the Operational Program "Human Resource Development, Education and Lifelong Learning", within the action "Strengthen Postdoctoral Researchers - Second Phase" (MIS 5033021), implemented by the State Scholarships Foundation (IKY)

The authors would like to thank OSeven Telematics, London, UK for providing all necessary data exploited to accomplish this study.

References

1. D. Singh; M. Singh, “Internet of vehicles for smart and safe driving,” International Conference on Connected Vehicles and Expo (ICCVE). (2015)
2. Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012). Driving behavior analysis with smartphones: insights from a controlled field study. In Proceedings of the 11th International Conference on mobile and ubiquitous multimedia (pp. 1-8).
3. 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.
4. Organisation for Economic Co-operation and Development (2016)
5. SafetyNet, S. Web Text 2009, retrieved 3/3/2018.
6. Alessandrini, A., Cattivera, A., Filippi, F., Ortenzi, F., 2012. Driving style influence on car CO2 emissions. In: Proceedings of the 20th International Emission Inventory Conference - "Emission Inventories - Meeting the Challenges Posed by Emerging Global. National, and Regional and Local Air Quality Issues". Tampa, Florida. Aug.
7. Yamakado, M., Takahashi, J., Saito, S., Abe, M., 2009. G-vectoring: new vehicle dynamics control technology for safe driving. Industrial systems. Hitachi Review 58 (December 7), 347.
8. Ohta, T., & Nakajima, S. (1994). Development of a driving data recorder. JSAE review, 15(3), 255-258.
9. Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers’ behavior. Transportation Research Part C: Emerging Technologies, 16(3), 320-331.
10. Zaldivar, J., Calafate, C. T., Cano, J. C., & Manzoni, P. (2011, October). Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In Local Computer Networks (LCN), 2011 IEEE 36th Conference on (pp. 813-819). IEEE.
11. Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. Safety Science, 119, 91-97.
12. Vlahogianni, E.I., Barmpounakis, E.N., 2017. Driving analytics using smartphones: algorithms, comparisons and challenges. Transport. Res. C: Emerg. Technol 79 (June), 196–206
13. Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2018). Mobile sensing and machine learning for identifying driving safety profiles (No. 18-01416)
14. Ziakopoulos, A., Tselentis, D., Kontaxi, A., & Yannis, G. (2020). A critical overview of driver recording tools. Journal of Safety Research\
15. Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (2016). Understanding regression analysis: An introductory guide (Vol. 57). Sage Publications.
16. Ekstrom, C. T., & Sørensen, H. (2014). Introduction to statistical data analysis for the life sciences. CRC Press.