Keywords: hybrid electric vehicle; driving styles; eco-driving; cluster analysis

Andrzej PUCHALSKI, Iwona KOMORSKA*
Kazimierz Pulaski University of Technology and Humanities in Radom
Malczewskiego 29, 26-600 Radom, Poland
*Corresponding author. E-mail: iwona.komorska@uthrad.pl

DRIVING STYLE ANALYSIS AND DRIVER CLASSIFICATION USING OBD DATA OF A HYBRID ELECTRIC VEHICLE

Summary. Ensuring the effectiveness of adaptive algorithms for advanced driver assistance systems (ADAS) requires online recognition of driving styles. The article discusses studies carried out during real driving cycles based on the GPS parameters and OBD system data of a hybrid vehicle. The work focuses on the search for measures of the speed and acceleration signals of the car and the measures determined on their basis that best describe the driving style responsible for the vehicle traffic safety and ecological safety. Relations between the type of driver, driving dynamics, and fuel consumption were studied. The driver's categorization was based on a statistical analysis of input signals and mean tractive force (MTF) by clustering.

1. INTRODUCTION

Driving style is understood as the way the driver controls the vehicle in the context of external conditions and plays an essential role in the process of managing the operation of all systems responsible for efficiency, ecology, safety, and comfort of driving. However, driving style information cannot be directly measured or detected. To discover and present this information, many efforts have been made from varied perspectives in recent years.

In the work of Sagberg, F. et al [1], a conceptual framework was proposed whereby driving styles are viewed in terms of driving habits established as a result of individual dispositions as well as social norms and cultural values. Moreover, a general scheme for categorizing and operationalizing driving styles was suggested. On this basis, existing literature on driving styles and indicators was reviewed. Links between driving styles and road safety were identified, and individual and sociocultural factors influencing driving style were reviewed. A wide range of both mathematical identification methods and modelling methods of driver behaviour are presented from the control point of view in Wang, W et al [2] based on the driving data, such as the brake/throttle pedal position and the steering wheel angle, among others. Augustynowicz in an article [3] proposed a method of preliminary driver classification, which applies two-criteria-based analysis of the phenomenon of driving style. The most active driver in the sense of the two-criteria analysis is the one who covered a test stretch of the road the fastest while changing the position of the accelerator pedal most intensively. Constantinescu et al. [4] adopted the hierarchical cluster analysis and principal component analysis on the GPS tracking data such as the speed and acceleration of the vehicle and suggested five categories of “aggressiveness” for classification of driving styles. In the study by Sundbom, M. et al [5], a probabilistic ARX model was utilized to predict the human driver’s behaviour and classify the driving styles, i.e., the aggressive and normal driving. Higgs and Abbas [6] segmented and clustered the car following behavior of three drivers who represent high-risk, medium-risk, and low-risk drivers. In the study by Qi, G. et al [7], the topic model Latent Dirichlet Allocation (LDA) was investigated. These
models are used especially in the field of text mining, image classification, social network, stock movement, incident detection, and so forth.

Virtual car design, which is now a standard, requires not only models of control objects and real driving cycles but also classification of driver behavior. Online recognition of driving styles specific to typical manoeuvres, dangerous manoeuvres, a tired driver or driver under the influence of alcohol, is necessary to predict the trajectory of the steered vehicle. It is, therefore, an indispensable function on the road to the gradual autonomy and development of intelligent transport systems.

Increased importance is put on eco-friendly driving style. More and more cars with hybrid and electric drive on the road evidence a change in people's approach to environmental pollution. Energysaving driving techniques (eco-driving) are used by drivers who want to reduce fuel consumption. Green driving is becoming the drivers' operational decision that can maximize fuel economy and thus reduce global greenhouse gas emissions and other air pollutants.

Van Mierlo et al. [8] describe the influence on vehicle emissions and energy consumption of different vehicle parameters and driving style as well as of traffic measures taken to increase transport safety or to reduce traffic jams. Merkisz et al. [9] verify how the driving style influences on the carbon dioxide emissions from vehicle engine and fuel consumption. Lois et al. [10] investigated the short-term effects of eco-driving by developing an analytical model of the key factors that explain fuel consumption and eco-driving, and to examine their relations in greater depth.

The objective of vehicle autonomy requires more accurate participation of the driver's driving style in the algorithms of controlling fuel consumption (air pollutants and CO₂ emissions), driving dynamics, and energy management. The ecological indicator must be considered in parallel with driving safety. In hybrid vehicles, the optimum use of electricity has become an additional factor.

An adequate energy management strategy is the key to optimizing hybrid electric vehicle fuel efficiency. Jiang et al. propose a comparison between three promising real-time methods: adaptive equivalent consumption minimization strategy (A-ECMS), optimal control law (OCL), and stochastic dynamic programming (SDP) [11]. Granovskii et al. compared the economic and environmental aspects of conventional, hybrid, electric and hydrogen fuel cell vehicles [12]. The comparative study of real-world driving cycles, energy consumption, and CO₂ emissions of electric and gasoline motorcycles driving in a congested urban corridor was described by Koossalaperom et al. [13]. Burdzik et al. presented the results of simulation studies, the aim of which was to compare the comparative energy consumption of short-distance and long-distance city buses in terms of assessing the needs for e-mobility [14].

Online identification and classification of driving style along with on-board, sensor, and wireless-based image of vehicle movement have become a necessity when designing cyber-physical vehicle models.

Adaptive machine learning algorithms enable online forecasting of motion parameters and generate feedback. They can be passive, taking into account the possibility of accepting or rejecting the proposed decision by the driver or actively implemented by the on-board computer system.

Currently, available driving cycles cannot be used to estimate actual fuel consumption or emissions from vehicles in the selected region because they do not describe the local driving style. Two approaches are used to obtain the large data sets needed for machine learning. One of them is designing driving cycles based on Markov chains and transition probability matrix [15, 16], and the other, used in this article, is creating sets from micro-trips. Large samples of the time series of recorded data are divided into micro-segments [17, 18]. To the best of our knowledge after reviewing the existing literature, previous works have focused on the comparative studies of conventional drives (with combustion engines), HEV, and EV.

The present work focuses on the search for measures of the speed and acceleration signals of the car and the measures determined on their basis, which best describe the driving style: vehicle traffic safety and ecological safety. Section 2 of the article presents the methods used to measure and analyze data. The tests were conducted by two drivers driving the car in completely different styles. Section 3 analyzes the performed measurements in three aspects. The first analysis (section 3.1) aims to find measures of signals allowing for an unambiguous classification of aggressive - not aggressive driving style. In section 3.2, measures describing ecological safety are sought. Here, data are clustered based
Driving style analysis and driver classification using OBD data of a hybrid…

on the fuel consumption measurement (eco-driving, eco-neutral and not eco-driving). The selected parameter MTF is a measure commonly used to evaluate both fuel consumption and total energy consumption. This parameter is related to the emission of toxic components of exhaust gases and CO₂. Section 3.3 presents the influence of driving style on energy management in the tested drive unit. The conclusions are included in section 4.

2. METHODS

To achieve the objectives of the paper, this section includes three aspects: data collection, data processing, and data clustering.

2.1. Data collection

Eight driving cycles were recorded covering a passage through the city of Radom. The first part of the route led through the Radom bypass, and the second through the city center. The test car was Toyota Yaris with a full hybrid drive and the total power of 74 kW, equipped with a combustion engine with a maximum power of 55 kW. Two drivers, a student and a taxi driver, drove the vehicle four rounds each. The route is shown on the map in Fig. 1.

Fig. 1. Location of the field trial in Radom

Car speed, acceleration, and GPS parameters, including terrain elevation, were recorded with Columbus GPS V-900 Data Logger. Other parameters available via the car's OBD connector were also recorded, such as instant gasoline consumption, battery charge level, engine speed, and other available parameters. The ELM 327 OBD2 scanner was used for this purpose.

The route of each cycle was 14.3-km long and included sections limited to 70, 50, and 40 km/h, some along the city ring road, and some in the city centre.

- Up to 1488 m: city driving limited to 40 and 50 km/h
- From 1488 m to 6134 m: driving along the city ring road with limited Up to 70 km/h
- From 6134 m to 7825 m: driving along the ring road with limited Up to 50 km/h
- From 7825 m to 14300 m: driving through the city centre with a limit of 50 km/h and numerous intersections with traffic lights.
It is illustrated in Fig. 2, which presents the course of the speed and acceleration of the car as a function of the distance travelled. For better visibility, only two cycles are shown in which it is clearly visible where the car has to stop due to traffic lights or to approach a roundabout.

The test was conducted off-peak hours from 9.00-14.00 to minimize the effect of increased traffic, traffic jams, etc. on the driving style of drivers.

Fig. 2a and Fig. 2b show the speed and acceleration of the car, respectively, as a function of distance for two drivers traveling the same route: a dynamic student and a fuel-saving taxi driver.

Fig. 2. Mileage speed of the car (a) and acceleration of the car (b) as a function of distance traveled

The driving style of the driver affects the way the engine is controlled, i.e. the battery charge and the number of starts of the internal combustion engine. A detailed analysis of this issue is described in section 3.3. Fig. 3 presents time waveforms of car speed $V_c$, internal combustion engine rotational speed $n_{ICE}$, battery state-of-charge (SOC), and instantaneous fuel consumption during the test carried out by the taxi driver.

The powertrain energy management system controls the starts and stops of the internal combustion engine depending on the battery charge. The degree of battery charge on the city beltway, where there are long sections of driving at a constant speed, allows the battery charge to fluctuate from 40-65%. With dynamic driving in the city center, the engine controller strives to maintain an almost constant battery charge level of approx. 55-60%. The more dynamic the driving style, the more frequently the combustion engine is run for a short time.

2.2. Data processing

The registered driving cycles were divided into segments of different lengths and duration of driving from 3 to 20 minutes, so-called micro-trips. A set of 397 cycles was obtained for which the following parameters were determined:
- mean value Mean (V) and standard deviation Std (V) of the car speed,
- mean value Mean (Acc +) and standard deviation Std (Acc +) of the car positive acceleration,
- mean value Mean (Acc -) and standard deviation Std (Acc -) of the car negative acceleration (deceleration),
- root mean square of the car acceleration RMS (Acc),
- mean tractive force MTF (Eq.1),
- parameters related to the operation of the internal combustion engine: average number of starts of the ICE over a distance of 1 km and time of the ICE operation after switching on, and
- average fuel consumption per 100 km.
Driving style analysis and driver classification using OBD data of a hybrid...

The Mean Tractive Force (MTF) focuses on the estimation of the traction energy of the vehicle transmitted by the wheels during the driving cycle [16]:

$$\bar{F}_{\text{trac}} = \frac{1}{\tau_{\text{trac}}} \int_{\tau_{\text{trac}}} F(t) v(t) \, dt$$

(1)

where: $\tau_{\text{trac}} = \{ t \in \tau : F(t) > 0 \}$

The total tractive force, $F(t)$, on the wheels consists of aerodynamic resistance, rolling resistance, and inertia resistance.

$$F(t) = F_{\text{air}} + F_{\text{roll}} + ma(t)$$

(2)

$$F_{\text{air}} = 0.5 \rho_a c_d A_{f} v^2(t)$$

(3)

$$F_{\text{roll}} = mgc_r$$

(4)

where: $F(t)$ - Total tractive force, $F_{\text{air}}$ - Aerodynamic resistance, $F_{\text{roll}}$ - Rolling resistance, $v$ - Velocity of a vehicle, $a$ - Acceleration of a vehicle, $\rho_a$ - Air density, $c_d$ - Drag coefficient, $A_{f}$ - Frontal area, $m$ - Vehicle mass, $c_r$ - Rolling friction coefficient, $g$ - Gravitational constant.

2.3. Data clustering

Cluster analysis or Lloyd’s algorithm [19] is the task of grouping a set of objects in such a way that items in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). Clustering is not an algorithm, and rather it is a way of solving classification problems. There are multiple algorithms that solve classification problems by using the clustering method. These algorithms differ in their efficiency, their approach to sorting objects into the various clusters, and even their definition of a cluster. $K$-means clustering is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of $k$ groups (i.e. $k$ clusters), where $k$ represents the number of groups pre-specified by the analyst. It classifies objects in multiple groups (i.e., clusters), such that objects within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas objects from different clusters are as dissimilar as possible (i.e., low inter-class similarity). In $k$-means clustering, each cluster is represented by its center (i.e., centroid), which corresponds to the mean of points assigned to the cluster.
The basic idea behind k-means clustering consists of defining clusters so that the total intra-cluster variation (known as a total within-cluster variation) is minimized [20]. There are several k-means algorithms available. The standard algorithm is the Hartigan-Wong algorithm (1979), which defines the total within-cluster variation as the sum of squared distances Euclidean distances between items and the corresponding centroid:

$$W(C_k) = \sum x_i \in C_k (x_i - \mu_k)^2$$  

(5)

Each observation \(x_i\) is assigned to a given cluster such that the sum of squares (SS) distance of the observation to their assigned cluster centers \(\mu_k\) is minimized.

3. RESULTS AND DISCUSSION

Statistical analysis of driving cycles was carried out owing to two criteria:
- driver characteristics and preferences in terms of driving dynamics (aggressive and not aggressive), and
- ecological safety (eco-driving, eco-neutral, and not eco-driving).

The recognized characteristics and proposed measures were used to perform the analysis of the energy management algorithm in a hybrid electric vehicle.

3.1. Statistical analysis due to driver characteristics and preferences

Driving safety is conditioned by a number of external factors and driver characteristics. The study was conducted in circumstances limiting the effect of environmental conditions, such as heavy traffic of vehicles and pedestrians, bad weather, insufficient lighting conditions, and road works. Monitoring and identification of these factors can be carried out using other features of the vehicle's on-board system. To determine the individual characteristics and preferences of the driver, indicators related to the recorded signals of speed of a car, its acceleration, and deceleration were adopted.

For the eight selected parameters characterizing the driving style of the driver, their relationships were analyzed by determining the correlation coefficient (Tab. 1).

<table>
<thead>
<tr>
<th></th>
<th>Mean (Acc+)</th>
<th>Std (Acc+)</th>
<th>Mean (Acc-)</th>
<th>Std (Acc-)</th>
<th>RMS (Acc)</th>
<th>Mean (V)</th>
<th>Std (V)</th>
<th>MTF (N)</th>
<th>Fuel (l/100km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Acc+</td>
<td>1</td>
<td>0.300</td>
<td>-0.806</td>
<td>0.077</td>
<td>0.136</td>
<td>-0.248</td>
<td>-0.383</td>
<td>0.018</td>
<td>0.074</td>
</tr>
<tr>
<td>Std (Acc+)</td>
<td>0.300</td>
<td>1</td>
<td>-0.003</td>
<td>0.603</td>
<td>0.811</td>
<td>-0.112</td>
<td>0.097</td>
<td>0.683</td>
<td>0.530</td>
</tr>
<tr>
<td>Mean Acc-</td>
<td>-0.806</td>
<td>0.003</td>
<td>1</td>
<td>-0.125</td>
<td>0.052</td>
<td>0.092</td>
<td>0.174</td>
<td>0.157</td>
<td>0.017</td>
</tr>
<tr>
<td>Std (Acc-)</td>
<td>0.077</td>
<td>0.603</td>
<td>-0.125</td>
<td>1</td>
<td>0.857</td>
<td>-0.110</td>
<td>0.049</td>
<td>0.690</td>
<td>0.569</td>
</tr>
<tr>
<td>RMS (Acc)</td>
<td>0.136</td>
<td>0.811</td>
<td>0.052</td>
<td>0.857</td>
<td>1</td>
<td>-0.138</td>
<td>-0.056</td>
<td>0.796</td>
<td>0.662</td>
</tr>
<tr>
<td>Mean (V)</td>
<td>-0.248</td>
<td>-0.112</td>
<td>0.092</td>
<td>-0.110</td>
<td>-0.138</td>
<td>1</td>
<td>0.411</td>
<td>-0.314</td>
<td>-0.247</td>
</tr>
<tr>
<td>Std (V)</td>
<td>-0.383</td>
<td>0.097</td>
<td>0.174</td>
<td>0.049</td>
<td>-0.056</td>
<td>0.411</td>
<td>1</td>
<td>-0.012</td>
<td>0.033</td>
</tr>
<tr>
<td>MTF (N)</td>
<td>0.018</td>
<td>0.683</td>
<td>0.157</td>
<td>0.690</td>
<td>0.796</td>
<td>-0.314</td>
<td>-0.012</td>
<td>0.804</td>
<td>1</td>
</tr>
<tr>
<td>Fuel (l/100km)</td>
<td>0.074</td>
<td>0.530</td>
<td>0.017</td>
<td>0.569</td>
<td>0.662</td>
<td>-0.247</td>
<td>0.033</td>
<td>0.804</td>
<td>1</td>
</tr>
</tbody>
</table>

Based on the matrix of correlation coefficients, the following can be stated:
- the root mean square acceleration value of the car is correlated with the standard deviation of acceleration (0.811) and deceleration (0.857) and the mean tractive force (0.796),
- there is a significant correlation between the average acceleration and deceleration values of the car (-0.806), and
- the average value and standard deviation of the car speed do not correlate with other driving parameters.
In the mean tractive force (MTF), acceleration is the most important element. The root mean square value is mainly influenced by the high values of both positive and negative rapid accelerations due to their second power. Considering the aforementioned conclusions, it was assumed that the driver's individual characteristics and preferences in the field of dynamics are well described by the root mean square acceleration value of the car [RMS (Acc)].

Fig. 4 visualizes two driving styles: aggressive (student) and not aggressive (taxi driver) presented on the set RMS (Acc) vs Mean (Acc +). Aggressive driving is characterized by higher RMS (Acc) values for equal Mean (Acc +).

![Fig. 4. Cluster analysis of driving styles based on the car acceleration](image)

Two drivers represented extremely different driving styles, so it can be assumed that the data obtained for other drivers will either coincide with those obtained or cover the area between them. Using these two coordinates, it is possible to separate the clusters. The slope of the line on the Fig. 4 (tangent of the slope angle) can be interpreted as the shape factor of the acceleration signal defined as the ratio of the RMS value of this signal to the mean value calculated for positive accelerations.

### 3.2. Statistical analysis for ecological safety

This criterion is related to the emission of air pollutants and greenhouse gases and is assessed based on the gasoline consumption. Instantaneous fuel consumption is determined in the car based on the signal from the air flow meter in the intake manifold and the signal from the oxygen sensor. Information is currently provided to the on-board diagnostic system (OBD).

Fuel consumption is best related to mean tractive force (MTF). Correlation coefficient between the MTF and average fuel consumption is 0.804, whereas analogously between RMS (Acc) and average fuel consumption, it is 0.662. MTF is a parameter depending on the speed of the car, as well as its acceleration and deceleration (See Eq.1).

Fig. 5 shows the relationship of average fuel consumption on MTF.

There is a noticeable large spread of average fuel consumption for the same MTF values. This is related to the hybrid drive modes, especially for low values of MTFs, where there is a large share of
purely electric drive. For the largest MTFs, the spread is low, because both machines work: internal combustion engine and electric motor.

During the driving cycle of each driver, there are sections with different fuel consumption, for example, forced by road conditions, speed limit, traffic jams, and traffic lights. Drivers driving aggressively also have sections of ecological driving. The division into three categories provides for a narrow transition zone between ecological and non-ecological driving. This is an arbitrary division based on an objective fuel consumption index. Fig. 5 also shows information about the driver to whom this part of the cycle relates.

The extreme samples on the scatter plot (Fig. 5) belong to short cycles (3 minutes). Those with very low fuel consumption were determined for fragments with a high proportion of electric propulsion, whereas those with the highest fuel consumption for fragments with internal combustion engines with dynamic speed changes. For 10-minute and longer cycles, average fuel consumption is in the range 4.7 - 8.2 l/100 km.

Fig. 6 shows the effect of driver acceleration and deceleration, expressed by RMS value, on fuel consumption along with the classification according to the mean tractive force (MTF). Besides, information about the driver to which the part of the cycle relates is included.

Eco and non-eco driving areas are distinguished, whereas the eco-neutral area mixes with the others.

3.3. Statistical analysis for energy management

The combustion engine is switched on when there is a greater power demand than the electric motor can provide, or when the battery needs recharging. The battery state is maintained within 40-80%, but such fluctuations are allowed by the control system when driving at a set speed, without rapid changes. The more dynamic the ride, the battery state-of-charge (SOC) parameter is maintained at a constant level close to 55% (see Fig. 3c). Therefore, during dynamic driving in the urban cycle, the combustion drive is frequently switched on for a few seconds. This results in increased fuel
consumption. The histograms of the number of starts of the combustion engine were compared as a function of engine operation time during one start for the driving cycle implemented by the student and the taxi driver on the same route (see Fig. 7).

The graphs show that during not aggressive driving, the internal combustion engine starts 27 times for a period of 2 to 20 seconds, with the distribution being quite even, whereas with the same initial battery charge, the aggressive driving engine starts 56 times, with the most short actions up to 8 seconds (up to 15 times for about 2 seconds).

In Fig. 8., the dependence of the combustion engine operation time over a 1-kilometre section is presented depending on the average traction force with the MTF classification.

The correlation coefficient between the data shown in Fig. 8 is 0.713. The large dispersion of the combustion engine operating time value for the same average traction forces results from different modes of propulsion system operation. The outlying highest values are recorded for the battery charging mode. For the largest values of traction forces, the ICE time spread is low, because there is one possibility - both engines work.

4. CONCLUSIONS

The study analyzes the driving styles of a hybrid vehicle, taking into account ecological safety and individual driver preferences in terms of dynamics. The driving style is not measured directly, and the quoted literature shows a variety of approaches to this topic. This work uses the speed and acceleration signals and the measures determined on their basis.

Based on the mean tractive force, it was proposed to identify and classify registered driving styles for fuel consumption as eco-driving, eco-neutral, and not eco-driving. It has been shown that the driver's individual characteristics and preferences in terms of dynamics for aggressive and not aggressive models are best described by the mean square value of car acceleration (RMS Acc) with mean positive acceleration values (Mean Acc+). The recognized characteristics and measures of
classification were verified during testing of energy management algorithms. Owing to the limited number of measurements, the cycles were divided into micro-trips, obtaining larger numbers of measurements for randomly selected sections of the route. Two drivers represented extremely different driving styles, so it can be assumed that the data obtained for other drivers will either coincide with those obtained or cover the area between them. This will become the subject of further research.

Fig. 7. Histogram of switching on the combustion engine during an urban cycle a) not aggressive, b) aggressive

Fig. 8. Impact of driving style to time of internal combustion engine operation per one kilometre

The tested measures of the speed and acceleration signals of the car, including MTF, will be used to verify simulated real-world driving cycles. The simulation analysis performed shows the credibility of this verification for a car with a hybrid drive.

The proposed method of online assessment of the driving style, leading to obtaining a driver model, can be used to support and improve the quality of the decision-making process. Correction of driving style by advanced on-board advanced driver-assistance systems (ADAS) can be implemented in passively in the feedback mode adapted to individual driver preferences or in an active manner, taking into account other parameters monitored on an ongoing basis.
Driving style analysis and driver classification using OBD data of a hybrid…

References


Received 13.05.2019; accepted in revised form 27.11.2020