Time-series clustering for pattern recognition of speed and heart rate while driving
A magnifying lens on the seconds around harsh events

Tselentis, Dimitrios I.; Papadimitriou, Eleonora

DOI
10.1016/j.trf.2023.09.010

Publication date
2023

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Citation (APA)
Tselentis, D. I., & Papadimitriou, E. (2023). Time-series clustering for pattern recognition of speed and heart rate while driving: A magnifying lens on the seconds around harsh events. *Transportation Research Part F: Traffic Psychology and Behaviour*, *98*, 254-268. https://doi.org/10.1016/j.trf.2023.09.010

Transportation Research Part F: Psychology and Behaviour 98 (2023) 254–268

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

Time-series clustering for pattern recognition of speed and heart rate while driving: A magnifying lens on the seconds around harsh events

Dimitrios I. Tselentis *, Eleonora Papadimitriou

Faculty of Technology, Policy, and Management, Safety and Security Science Section
Delft University of Technology, Jaffalaan 5, Delft 2628 BX, the Netherlands

A R T I C L E I N F O

Keywords:
Driving Pattern Recognition
Speed Patterns
Heart-Rate patterns
Driving pulse
Time-series Clustering
Unsupervised Learning

A B S T R A C T

Driving pattern recognition has been applied for the purposes of driving styles identification and harsh driving events detection. However, the evolution of driving behavior around and especially before such events has not been investigated at a microscopic level. The objective of this research is to reveal existing driving patterns around harsh events at the driving ‘pulse’ level i.e. a few seconds before and after the event. For that purpose, a time-series clustering approach is applied on speed and heart rate metrics of individual drivers using data collected from a large naturalistic driving study. Results show that there are distinct speed patterns before harsh braking, harsh acceleration, and harsh cornering events. A deceleration is identified shortly before most harsh acceleration and cornering events, which possibly indicates reckless behavior, i.e. drivers not dedicating enough time to smoothly brake before cornering, or of a brief ‘decision-making’ moment before the harsh manoeuvre. On the contrary, speed seems to be steady before harsh braking events. Regarding heart rate, the analysis revealed certain patterns only after raw data were cleansed and filtered. These patterns may show increasing, decreasing or variable heart rate trends, which may correspond to different stress patterns of drivers around harsh events. Finally, we introduce the concept of driving pattern consistency, which can reveal the share of individual drivers that follow the same harsh event pattern. It is indicated that more than half of the drivers are not consistent, suggesting that driving patterns around harsh events may be more context-related than driver personality-related.

1. Introduction

Driving behavior analysis has been studied on a macro and micro level in the past (Karlaftis et al., 2013, Papadimitriou et al, 2019, Mantouka et al., 2019, Eboli et al., 2016). More specifically, driver behavior analytics is a research field has emerged during the past decades with several important applications. While entering the Big Data era, new data collection schemes and advanced modelling techniques related to Machine Learning and Artificial Intelligence became available. There are nowadays considerable opportunities for large-scale collection of new fine-grained data such as driver speeding profiles, longitudinal and lateral movement, physiological indicators, traffic conditions, road surface and environment conditions, detailed weather and spatial information (Weidner et al, 2017,

* Corresponding author.
E-mail address: D.Tselentis@tudelft.nl (D.I. Tselentis).

https://doi.org/10.1016/j.trf.2023.09.010
Received 29 April 2023; Received in revised form 2 August 2023; Accepted 18 September 2023
Available online 29 September 2023
1369-8478/

Ellison et al., 2015) – which can be used for the analysis of driving behavior and its impact on safety, efficiency etc.

In terms of driving style, it is found that gender, personality traits and executive functions should be taken into account when studying safety among older drivers (Adrian et al., 2011). Personality and safety attitudes are also found to be directly related with risky driving behaviors (Chen, 2009). Personality traits, attitudes and risk perception have also been used as predictors of risky driving behaviour of young drivers (Ulleberg & Rundmo, 2003). These measures were collected through questionnaires and included risk perception, attitudes towards traffic safety and self-reported risk-taking in traffic as well as the personality measures of aggression, altruism, anxiety and normlessness. Other research findings of the past literature show a relationship between the physical and emotional conditions and the driving style, e.g. a driver being aggressive or cautious, using questionnaire data (Eboli et al., 2017)). There are also researchers that have investigated driving speed as a risk factor of driving behaviour that can be used to classify behaviour as safe, unsafe, and safe but potentially dangerous (Eboli et al., 2017).

The feasibility and benefits of identifying driver profiles and driving styles based on inertial sensor data (speed, acceleration, braking, steering etc.) across time and space is shown in several studies (Papadimitriou et al, 2019, Mantouka et al., 2019, Weidner et al, 2017, Ellison et al., 2015, Tselentis et al., 2019, Tselentis et al., 2021). Nonetheless, changes in driving behavior may sometimes happen quickly in a very short time scale and therefore, drivers should be continuously monitored at a high resolution to ensure that such behavioral shifts are captured over time and that their safety implications are adequately understood (Tselentis et al., 2021).

More specifically, literature has focused on the detection of harsh events such as harsh acceleration and braking, and their classification based on the intensity of the event, but the microscopic characteristics of driving behavior around and especially before each event have not been investigated. Research thus far has shown that studying behavioral patterns on a driving "pulse" level, defined as the time period that a vehicle is in motion, bounded by two adjacent stops, is a much more promising microscopic level of analysis in order to yield significant insights in driving pattern recognition (Liaw et al., 2002). The definition of the pulse's length certainly depends on the characteristics of the behavioral pattern investigated as well as on the scope of the analysis. In order to study driving behavior at this level, metrics such as speed should be examined on a microscopic level i.e. a time-series recorded on a medium or high frequency. At the same time, it has been shown that a certain driving pattern may vary among drivers and driving events (Lee et al., 2018), i.e. it may be observed in different drivers and in different conditions. Therefore, driving behavior should be investigated in depth to understand whether or not there are underlying groups of behavioral patterns that are more prominent under certain conditions or groups of drivers.

The complete review on driving patterns studies by Tselentis & Papadimitriou (2023) highlights that driving pattern recognition has been investigated in the past for the purposes of recognition of driving styles of specific vehicles e.g. electric vehicles, for driver identification or driving maneuver recognition. It was highlighted that most studies use a more mesoscopic or macroscopic level of analysis, such as on a trip or driver level (Tselentis et al., 2019). Nonetheless, literature has not yet addressed the following three (research) questions, which will be the main focus of this paper:

1) Are there common driving patterns shortly before and after events, e.g. at the driving pulse level?
2) Are there indications that some driving patterns could potentially be related to safety critical behaviors in terms of human factors e.g. aggressiveness, stress and cautiousness?
3) Are these driving patterns also 'driver' patterns, i.e. are they consistent within drivers, or anybody could exhibit a certain pattern under different circumstances?

The objective of this paper is to identify and analyze the existing driving patterns at the microscopic time window of a driving 'pulse', in terms of speed and heart rate (beats per minute) of the driver. A novel time-series clustering approach that is used in this study is combining a rule-based and a K-means algorithm to cluster microscopic driving data shortly before and after common harsh events, i.e. harsh acceleration (HA), harsh braking (HB) and harsh cornering (HC). Heart rate data are cleansed using the moving average of a 5-seconds time-window. The method is applied on recent high resolution European naturalistic driving data, including time series of speed and physiological measurements (heart rate) of drivers.

The remainder of the paper is organized as follows: Section 2 includes a summary of the literature on driving pattern recognition, leading to the specific research questions of this study. Then the data collection and handling procedures are described in section 3, followed by a presentation of the analysis methods used in section 4. Finally, the results are presented in terms of the driving patterns recognized per type of harsh event and driver metric in section 5, followed by a pattern consistency analysis in section 6. The paper concludes with a discussion of findings (section 7).

2. Literature review

Driving patterns have been studied in the past with the purpose of driving style recognition e.g. aggressive, normal, defensive, mild or gentle, using classification or clustering techniques based on support vector machines or neural networks (Cura et al., 2020, Shi et al., 2015). Pattern recognition has also been investigated from the perspective of driving maneuver identification such as normal or HA, HB, HC, lane change, parking and driver foot pedal behavior (Sarker et al., 2021, Schwarz, et al., 2017, Takano et al., 2008), mostly based on neural networks and Hidden Markov models. Researchers have developed models to predict normal or harsh behaviors and maneuvers (Obuhuma et al., 2018), whereas others have focused on the recognition of the type of vehicle for instance, private car, waste collection vehicle, truck or sweeper vehicle (Spyrou et al., 2020). Finally, driving patterns have been analyzed to understand the level of driving risk, which was determined by the crashes and near-misses recorded during simulator experiments (Wang et al., 2010).

Driving pattern recognition studies, usually result into the discovery of the aggressive driving pattern (Sarker et al., 2021*)*. In other words, this is referred to the discretization among different driving patterns such as normal, non-aggressive, defensive, stable, mild and gentle driving, or other manoeuvers such as normal acceleration and braking (events), turning, lane changing and parking (Feng et al., 2018; Garcia-Constantino et al., 2014; Takano et al., 2008; Schwarz, 2017). It was found though that many studies focused on the methodological contributions of their work and did not discuss the specific patterns discovered (Tselentis and Papadimitriou, 2023).

The review conducted on driving pattern studies revealed that microscopic driving pattern recognition involves mainly analyses of time-series data. The majority of the methodologies used are based on Neural Network (NN) models. Several different NN approaches were found to be used, mainly from the family of Recurrent Neural Network (RNN), such as Long-Short Term Memory (LSTM) model, standard RNN, Convolutional Neural Network (CNN) and fuzzy NN (Sarker et al., 2021; Mendoza and Vermelin, 2019). It was found that, depending on the type of the dataset, these models were employed for supervised and unsupervised tasks, i.e. for classification or clustering respectively. Finally, classification methodologies such as SVM, k-NN, and decision trees were found to be exploited as standalone approaches for pattern identification more frequently than clustering methodologies, such as support vector clustering (Tselentis and Papadimitriou, 2023).

In several studies, patterns are detected through the application of time-series clustering techniques based on distance measurement methods such as dynamic time warping (DTW) and Euclidean distance between time-series, using sensor recordings of vehicle's speed and driver's heart rate (Goffinet et al., 2020, Garcia-Constantino et al., 2014). When it comes to time-series clustering and pattern detection using DTW, it is usually complemented by other clustering methodologies e.g. Hierarchical and K-Medoid clustering (Garcia-Constantino et al., 2014, Baca Mendoza & Söderkvist Vermelin, 2019, Goffinet et al., 2020).

It is also noted that there are also some cases where time-series segmentation, which is the process of segmenting the time-series into smaller parts of consecutive observations, may precede clustering (Goffinet et al., 2020). This is beneficial in cases when the scope of the analysis or the data structure does not facilitate or allow the segmentation of the time-series into parts of the same length. A relatively new concept incorporated in driving behavior analysis studies, which captures the temporal evolution of driving behavior is the driving "pulse". Driving pulse can be defined as the time period that a vehicle is in motion, bounded by two adjacent stops. Some studies found this to be a much more promising microscopic level of analysis (Liaw et al., 2002, Obuhuma et al., 2018) that yields significant results in driving pattern recognition. However, the characteristics, insights and added value of different methods and analysis scales for driver profile and driving pattern recognition have not been systematically explored. Also, variations of the driving pulse definition, such as driving time-series segments of a pre-determined length, were not found to have been used in past literature for driving pattern recognition purposes.

Regarding the driving metrics used, the most commonly used driving metrics in driving pattern studies are speed and acceleration. This demonstrates the high importance that these two metrics play in the safety assessment of individual driving risk. In general, speed is the most popular time-series metric used for driving pattern analysis when time-series clustering is applied (Goffinet et al., 2020). This is because speeding is recognized as one of the most important factors in driving risk since it influences the accident probability (e.g. decreased reaction distance, risk of loss of control) as well as the crash severity (Mesken et al., 2002).

Apart from these two, pedal position and pressure are also strongly preferred followed by braking, RPM, angular velocities and steering. Another finding is that the majority of researchers exploited naturalistic driving data collected from OBD devices or smartphone sensors (Sarker et al., 2021*, Baca* Mendoza and Vermelin, 2019) and that only a few of them used data from driving simulators (Lee et al., 2018, Garcia-Constantino et al., 2014). The driver sample in the studies reviewed ranged between 4 and 34 participants and the data collection frequency was up to 100 Hz. The relatively high collection frequency reveals that a higher granularity of information is required in order to capture and analyze microscopic behavior, and perform microscopic pattern identification. If the frequency level is significantly lower, this may be considered inadequate and thus lead to the acquisition of inadequate information for microscopic analysis (Tselentis and Papadimitriou, 2023).

The number of harsh events that occur during a trip, including HA, HB and HC, are also three significant indicators for driving risk assessment, especially when evaluating driving aggressiveness (Tselentis et al., 2017)Johnson and Trivedi, 2011*,* Bonsall et al., 2005). The correlation between harsh events and crash risk is highlighted in literature and has been recognized by the insurance and telematics industry (Tselentis et al., 2017, Bonsall et al., 2005). Harsh events are strongly correlated with unsafe distance from adjacent vehicles, possible near-misses, decreased reaction time, poor driving judgement and involvement in high risk situations. Harsh event detection has also been the focus of many studies in the past using sensor data (Vlahogianni & Barmpounakis, 2017, Saiprasert et al., 2017).

Physiological indicators such as heart rate and eye movement were used in the past to measure driving performance (Mehler et al., 2008) and detect drivers' states such as drowsiness, stress level and vigilance level (Bergasa et al., 2006; Healey & Picard, 2005; Afghari et al., 2022). The impact of other tasks such as mobile phone usage, or traffic conditions, on heart rate is also investigated (Reimer et al., 2011, Tozman et al., 2015). It is found that heart rate and skin conductance metrics provided the highest overall correlations with continuous driver stress levels (Healey and Picard, 2005). In the latter study, the skin conductance was measured in two locations, the palm of the left hand using electrodes placed on the first and middle finger and on the sole of the left foot using electrodes placed at each end of the arch of the foot.

In other applications, heart rate has been studied with the scope to distinguish between single-task driving and distracted driving (Mehler et al., 2011, June), to identify different driving styles (Meseguer et al., 2018) or to distinguish between alert versus sleep-deprived drivers (Persson et al., 2020). Nonetheless, such indicators have been mainly used for macroscopic analyses, e.g. using heart rate variability data aggregated on 1 or 2 min (Sato et al., 2001), and not with the aim to better understand their relationship with driving patterns and generally with microscopic characteristics of driving behavior. Heart rate is found to be used together with speed (Feng et al., 2018). In some studies, heart rate variability is used instead of the actual heart rate recordings (Shakouri et al., 2018, Mahachandra et al., 2012), when the scope is to identify the transition between different driving phases e.g. from normal driving to driving under drowsiness, fatigue or stress.

3. Data collection and handling

The data used in this study were collected from a database developed by the i-Dreams Horizon 2020 European project ("i-Dreams: Safety Tolerance zone calculation and interventions for diver – vehicle – environment interactions under challenging conditions" of the Horizon, 2020)) using an API service developed by the project partners. The data stored in this database were recorded from a naturalistic driving experiment in which private car drivers from Belgium, UK, Germany and Greece participated. A total of more than 26 thousand trips from 124 drivers were collected from the i-Dreams database between September 2021 and June 2022. Based on past similar studies using naturalistic driving data (Spyrou et al., 2020, Wang et al., 2010, Takano et al., 2008), we notice that most researchers use a smaller sample compared to the one used in this study and therefore, we can consider the sample to be sufficient for this analysis.

During data cleansing, trips with total distance less than 300 m as well as those with duration less than 90 s were eliminated. Outliers of speed and heart rate were also cleaned and removed. The final sample consisted of 19,305 trips from 91 drivers, who performed at least 1 harsh event each. Similarly to (Garcia-Constantino et al., 2014), the analysis was performed using time-series sensor recordings of vehicle's speed and driver's heart rate. The high number of eliminated drivers is attributed to the fact that i) there were many very short trips in terms of duration or distance, ii) the threshold used for the elimination of outliers was 250 for the heart rate and 180 for speed and iii) some of the drivers had just joined the experiment so not many trips and driving events were recorded for them at that point of time.

The time-series data of speed were collected from a GPS sensor with a 1 Hz frequency. The heart rate time-series data were received in the form of inter-beat intervals, indicating the time interval between successive heart beats (about one per second), using a wristband sensor with a 1 Hz frequency. At the end of each trip session, sensor data were validated by removing any repeating sample points and ensuring temporal order of the data points. Regarding GPS data, incorrect latitude and longitude during momentary loss of GPS signal were filtered out to remove position jumps. Further data collection, cleansing, pre-processing and analysis were implemented using Python 3.7 and the Python packages of requests, pandas, numpy, sklearn and tslearn. A total of more than 11 GB of time-series data were processed and modeled.

During the iDreams experiment, harsh events were detected using data related to vehicle trajectory, speed, and accelerometer recorded in real-time conditions, and they are detected and recorded on an event basis. Harsh events may have a varying intensity, which is categorized as low, medium and high by the data providers of this study. High intensity events were only analyzed in this study. In general, harsh events during the experiment were defined as those harsh driving tasks performed by drivers, such as HB, which indicate a sudden change of one or more driving metrics, e.g. speed and direction. For instance, a HB was defined as a sudden reduction of speed. It is highlighted at this point that the data provider used a classified algorithm to classify events as harsh or not. The exact details of this algorithm cannot be disclosed due to confidentiality reasons; in general, it uses data from several sensors such as the GPS and the accelerometer to detect the events taking place and their intensity level. This algorithm is trained using Machine Learning techniques and calibrated through annotated field experiments.

Table 1 provides some descriptive statistics on the harsh event data collected during the 10 months of the experiment. For instance, it is shown that the median number of high-intensity HA events per driver is 19 whereas the respective metric for HC events is 26. It should be highlighted at this point that the sensors recording heart beats were installed at a later stage during the field experiments and consequently, heart rate data are not available during all harsh events, as opposed to speed data. Table 1 shows that speed data are available in 4 to 14 times more events than heart rate data.

All data coming from different sensors were synchronized before using them in the analysis. These data were recorded during the 10-month experiment. It is noted though that the heart-rate sensor was installed at a later stage of the experiment and this is why there are significantly less recordings.

The heart rate data were found to be extremely 'noisy' and therefore, a different cleansing procedure was followed. In order to smoothen the time-series of heart rates, a moving average was used, using a 5-seconds time window. The length of the time window was chosen after several trial and errors considering both the number of time-series peaks that were smoothened and the insights produced by the end results. To this end, 3 different levels of differences between consecutive observations of the smoothened heart rate per minute were taken into account, namely 10, 20 and 30 heart beats/ minute. In other words, when the smoothened heart beats/ minute difference of 10 was tested between 2 consecutive seconds, the data cleansing process was filtering out all time-series that have one or more cases of 2 consecutive seconds within the time-series having a difference of 10 or higher smoothened heart beats/ minute.

Table 1
Descriptive statistics on the harsh event data collected for this study during the 10-month experiment.

	Per driver				Total	
Harsh event	Min	Max	Mean	Median	Total # events with available speed data	Total # events with available heart beat data
Harsh braking	0	48	7	2	642	129
Harsh acceleration	0	1,689	108	19	9,794	722
Harsh cornering	0	971	141	26	12,862	3,084

4. Methods

4.1. Methodological approach

This study follows a time-series clustering approach to identify the underlying driving patterns in terms of speed and heart rate within a driving pulse in which a harsh event of high intensity takes place. The harsh events considered were HB, acceleration and cornering. To this end, driving pulses of 14 s around harsh events, consisting of a time-series that covers 10 s before and 3 s after the harsh event, including the second of the event, are modelled. Given the lack of similar previous research, the length of the pulse / time-series was specified after testing all pulse lengths from 5 to 20 s, by evaluating the interpretability and stability of the patterns.

The time-series of speed and heart rate were normalized before clustering. Among other reasons, normalization is important to reveal real patterns without being influenced by the order of magnitude of the time-series values. This means that the same pattern could be detected in different road types showing e.g. without focusing on whether speed is reducing from 100 to 80 km/h or from 50 to 30 km/h.

This research applies a combination of Euclidean distance measurement between all time-series and a K-Means clustering based on these distances. Since all time-series have the same length, the use of DTW that is shown in other studies (Goffinet et al., 2020, Garcia-Constantino et al., 2014) was not examined herein. Taking into account the 3 harsh event categories and the 2 different metrics mentioned above, this leads to the application of a total number of 6 clustering applications groups.

4.2. Time series k-means clustering

Clustering allows finding and analyzing groups that are formed naturally, instead of defining groups prior to looking at the data. The K-Means algorithm is one of the most popular clustering algorithms that belongs to the unsupervised learning techniques. It aims to find the optimum way to group given data, based on the feature similarity of the observations, with the number of groups represented by the variable K that is given as input. The centroid of each cluster is a collection of feature values that defines resulting groups based on which, the average behavior of the resulting groups is interpreted *(*Hartigan & Wong, 1979*)*. K-Means is an iterative algorithm that starts with randomly selecting K points as the initial centroids and assigning each observation to each cluster based on their distance to each centroid. Once the full sample is assigned to K clusters for the first time, the centroids are recalculated. The process is repeated several times until either the centroids are not changing or the maximum number of algorithmic iterations is reached.

The detailed algorithmic steps of the methodology were presented by (Hartigan and Wong, 1979). The algorithm requires as input a matrix of M points in N dimensions and a matrix of K initial cluster centers in N dimensions. The general procedure is to search for a K-partition with locally optimal within-cluster sum of squares by moving points from one cluster to another.

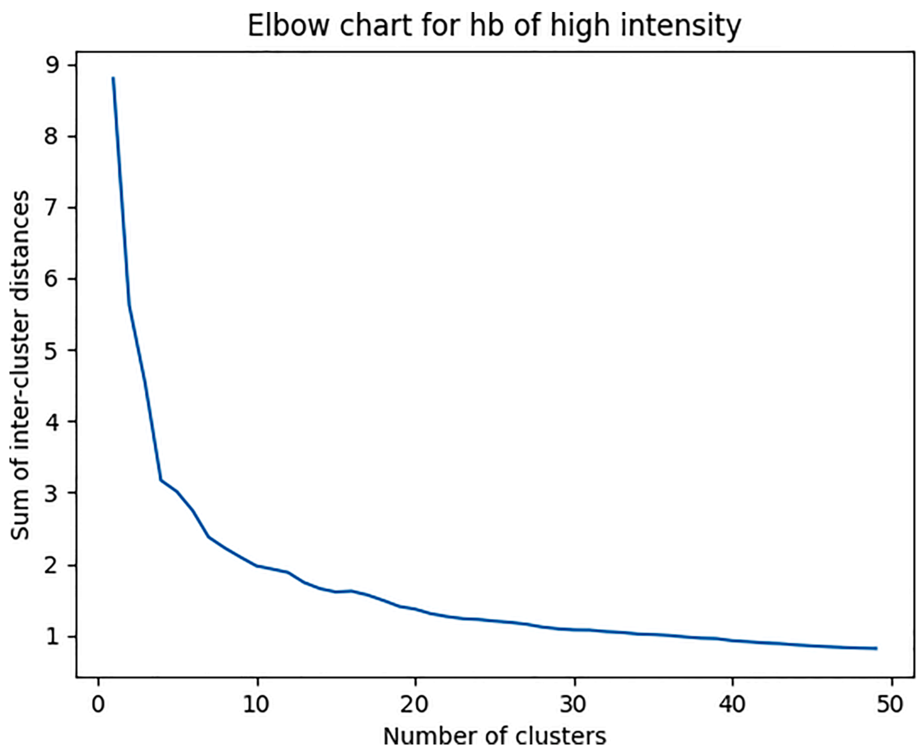


Fig. 1. Elbow chart for speed pattern clusters during harsh braking events.

In road safety, the method has been used to cluster several subjects into groups with similar behavior, e.g. to discover profiles and patterns of all road users including pedestrians and cyclists (Vogel et al., 2014, Kim & Yamashita, 2007) or drivers (Mantouka et al., 2019, Tselentis et al., 2021, Warren et al., 2019). It is also successfully used for time series clustering in road safety, e.g. for the identification of lane-changing profiles (Chen et al., 2021), event detection (Chetouane et al., 2021), driving styles identification (de Zepeda et al., 2021) and driving cycle development (Fotouhi & Montazeri-Gh, 2013). To the best of the authors' knowledge, this is the first time a time series clustering approach is implemented for the identification for microscopic driving patterns before harsh events on the basis of both driving metrics and physiological indicators.

4.3. Elbow method

A way to determine the number of clusters beforehand is by running the K-Means algorithm several times and comparing the results. To compare the results across different values of K, a popular metric is the mean distance between cluster centroids and the data points assigned to each one of them, which is decreased while the number of clusters is increased. When this metric is plotted as a function of the number of clusters K, the "elbow point", where the rate of this metric's decrease sharply shifts, is revealed together with the corresponding number of clusters K. It is noted that there is no methodology to estimate the optimal number of clusters beforehand and therefore, the elbow method is used to get an indication of the area of solutions.

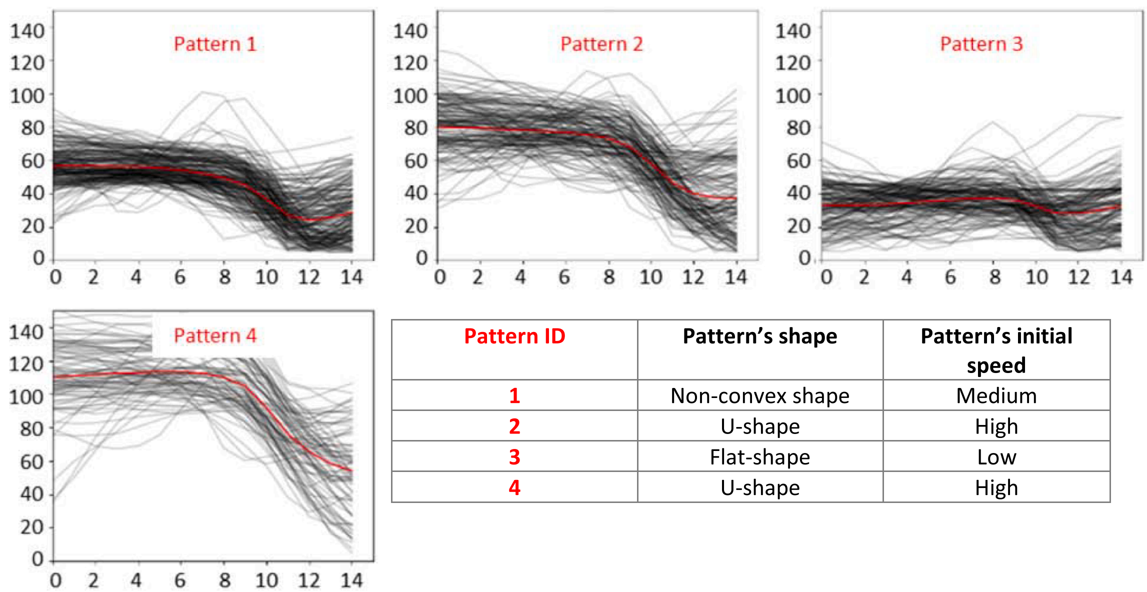
5. Results

5.1. Speed patterns

This section presents the results of the speed patterns before and after high intensity harsh events, followed by the heart rate patterns. The speed pattern results are illustrated in Figs. 2, 3 and 4, where the Y axis represents speed (km/h) and X axis time (sec). It is pointed out that the red line of each figure shows, at each of the 14 s recorded, the average speed in km/h of all the time-series that belong to each cluster/ pattern. Underneath each figure (also in heart rate patterns), a table is added describing the shape of the pattern and the initial speed at the beginning of the driving pulse. This subjective, qualitative description of what each cluster represents is done for readability purposes.

The elbow chart of HB events clustering is indicatively shown in Fig. 1, showing that there may exist 4 different speed patterns since the reduction rate of the sum of inter-cluster distances (Y axis) is significantly lower for more than 4 clusters. As aforementioned, the results of 3 to 6 clusters were also investigated in this case, with the 4-cluster solution eventually providing the most meaningful results in terms of interpretability. Similarly to HB, the elbow charts were created for the rest of the events but will not be presented for the economy of space.

Fig. 2 presents the 4 speed patterns revealed for the HB events. The similarity between patterns 1 and 3 is that brakings are followed by an acceleration of the vehicle after second 11, which is the second of the event, whereas the U-shaped patterns 2 and 4 are followed mainly by further deceleration after the event. The difference between patterns 2 and 4 is that the speed difference before and after the event is higher in pattern 4, which mainly represents patterns taking place with higher starting speed. Pattern 4 is represented by the



Pattern ID	Pattern's shape	Pattern's initial speed
1	Non-convex shape	Medium
2	U-shape	High
3	Flat-shape	Low
4	U-shape	High

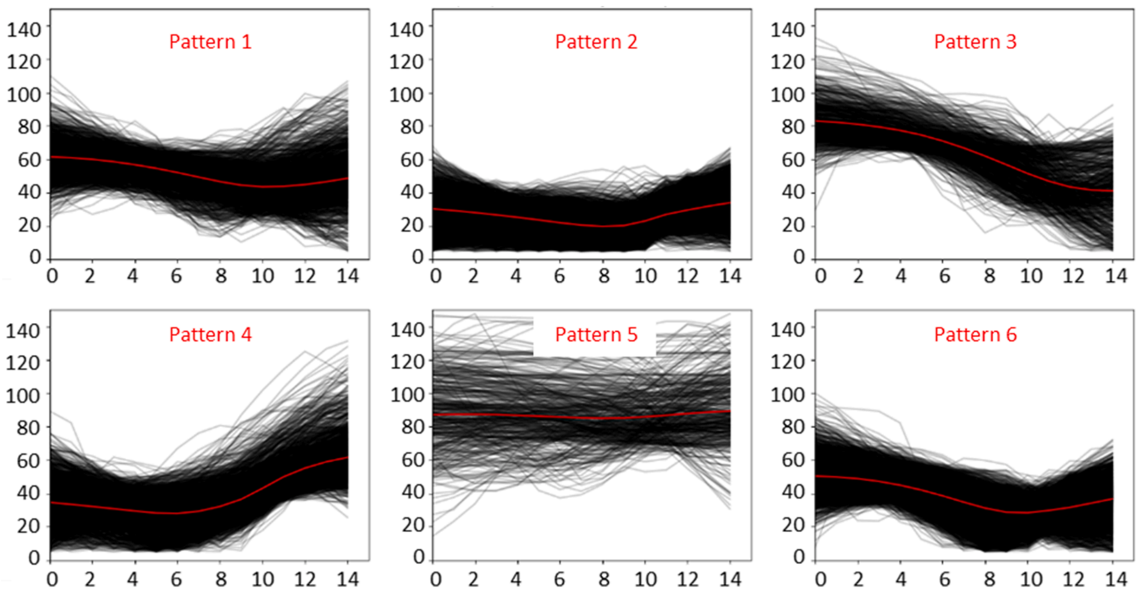
Fig. 2. Results of the 4 different speed patterns identified during harsh braking events (Y axis: Speed, X axis: time, harsh event occurs at the 11th second of the X-series).

least number of observations within the HB sample, that is 13.5 % of total. On the other hand, the difference between patterns 1 and 3 is that pattern 1 has a non-convex shape and mainly consists of 'harsher' braking events, as can be inferred from the higher reduction of speed. It appears that during all patterns, a constant speed is largely maintained before the harsh event.

Regarding HA events, the elbow chart indicated that six different speed patterns exist. In Fig. 3, it is apparent that the most intense acceleration pattern is #4, which has the largest speed difference before and after the event, a low starting speed and a mild braking before the event. Pattern 2 is the pattern with the mildest deceleration before, and acceleration after the event. U-shaped patterns 1 and 6 are the most similar among the others, with the difference that pattern 6 presents a more harsh deceleration before the event. It also appears that on average, pattern 6 starts from a lower speed compared to pattern 1. Both these patterns have more significant deceleration before the event compared to patterns 2 and 4. It is also worth mentioning that the deceleration observed before the HA event in patterns 1, 2 and 6, is a possible indication of an aggressive behavior since i) speed is approximately the same before and after the event and ii) drivers could have probably avoided the event in cases of events that could have been foreseen, such as HB before stopping at a traffic light or stop sign, if they maintained a lower average speed before the event.

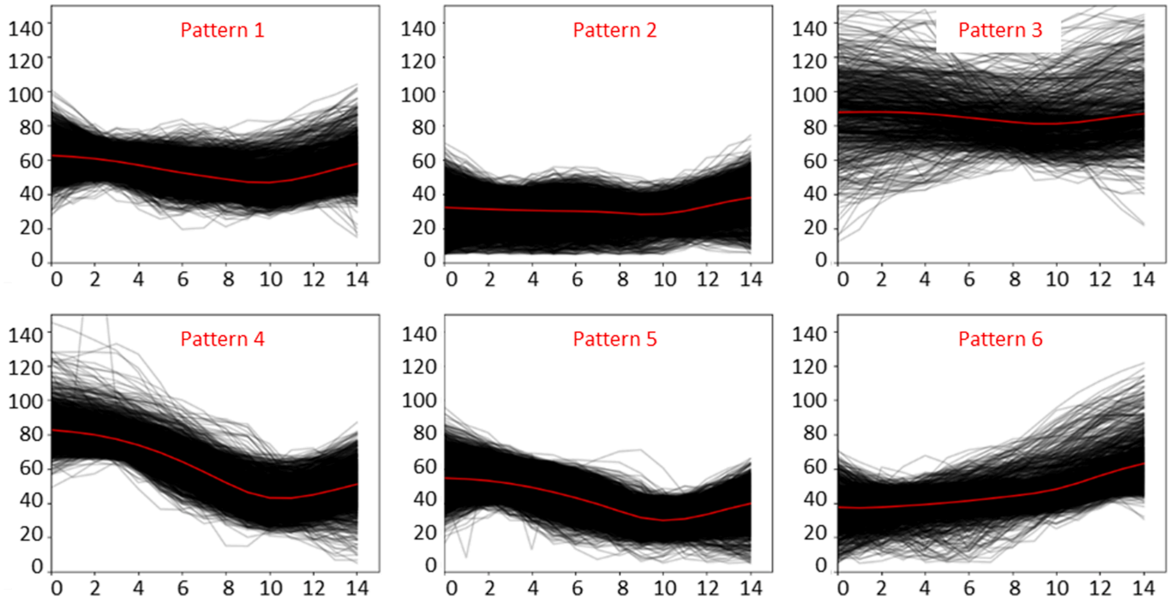
Pattern 5 is the least represented pattern, consisting of 4.8 % of the total observations most of which are high-speed patterns. It appears that this cluster consists of the patterns that were the most dissimilar to the rest and therefore, they were classified together since they could not be assigned to any other pattern. These observations should be further investigated. As for pattern 3, it is not a representative example of speed pattern for HA events, but it is probably identified as a distinct one because it displays a similar behavior for the period before the event, but a different post-event speeding behavior. It is not a well-represented high-speed pattern, with only 7.4 % of the observations belonging to it. Again, this means that the algorithm recognized the similarity of these observations and most importantly, the dissimilarity of these observations compared to the rest of the patterns.

As for the HC events, the indication of the elbow chart is that there exist 6 different speed patterns. The common attribute of all patterns is that they are characterized by a mild to intense acceleration after the cornering event that seems to be more intense when (the normalised) starting speed is higher. This can be noticed also in Fig. 4. Another common characteristic of all patterns except pattern 6, is that they display a mild to intense deceleration before the event that also seems to be more intense when starting speed is higher. Patterns 1 and 5 are the most similar among them, having a medium initial speed and with the difference between them being that the deceleration in #5 before the event is harsher. Pattern 2 shows the mildest speed change both before and after the event, whereas pattern 4 shows exactly the opposite. Patterns 4 and 5 present a similar non-convex shape and harsh deceleration before cornering, which is likely to indicate driving recklessness or aggressiveness since drivers could have probably dedicated more time to



Pattern ID	Shape	Initial speed
1	U-shape	Medium
2	Flat-shape	Low
3	Non-convex	High
4	U-shape	Low
5	Flat-shape	High
6	U-shape	Medium

Fig. 3. Results of the 6 different speed patterns identified during harsh acceleration events (Y axis: Speed, X axis: time, harsh event occurs at the 11th second of the X-series).



Pattern ID	Shape	Initial speed
1	U-shape	Medium
2	Flat-shape	Low
3	Flat-shape	High
4	Non-convex	High
5	Non-convex	Medium
6	Flat-shape	Low

Fig. 4. Results of the 6 different speed patterns identified during harsh cornering events (Y axis: Speed, X axis: time, harsh event occurs at the 11th second of the X-series).

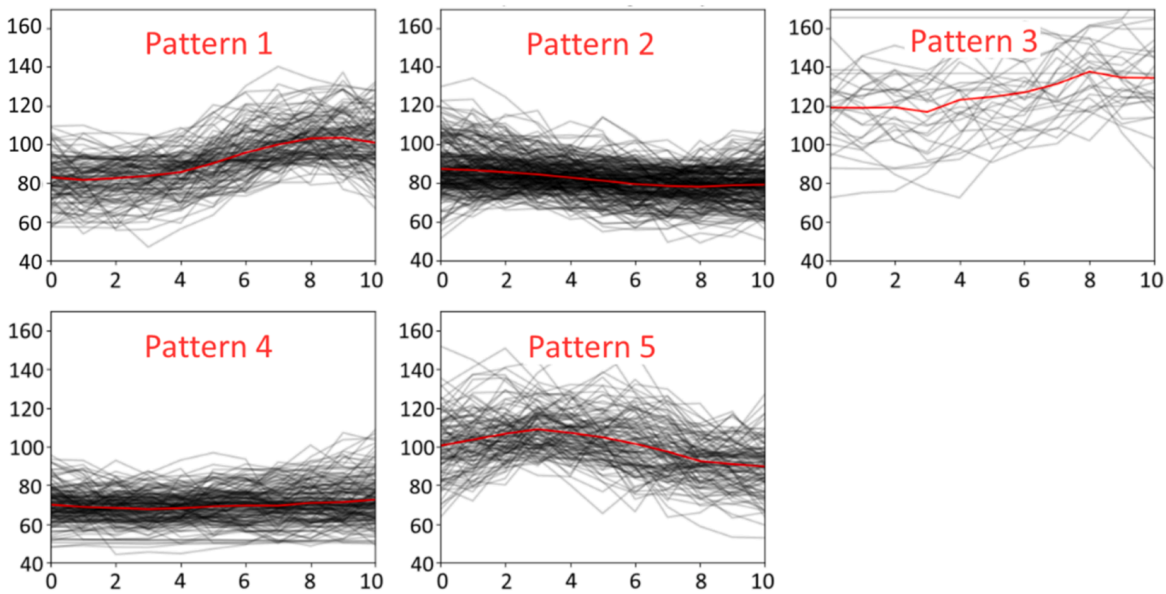
brake before cornering in some of these cases. Pattern 3 is the least represented pattern with 5.9 % of total observations, having a flat shape and high initial speed, and concentrating the most dissimilar observations compared to the rest of the patterns.

5.2. *Heart rate patterns*

The initial investigation of heart rate patterns around harsh events showed that the results produced are not robust when raw heart rate data are used. This was recognized initially by the elbow chart, where no sharp shift is identified at any point of the graph. A wide range of results for all harsh events were examined and results were similarly poor in all cases. The results of heart rate patterns indicated no strong patterns were noticed for the whole pulse of the time-series, except from some common peak areas based on which time-series were grouped together. This was an indication that raw data should be smoothened and cleaned before heart rate patterns were investigated.

Having cleansed the heart rate data (as described in section 3), time-series clustering was tested separately for all different levels of cleansing i.e. for heart rate inter-second difference of 10, 20 and 30 heart beats/ minute. It was noticed that all results produced from data cleansed using the 10 heart beats/ minute difference were not insightful since i) significant part of the data were filtered out and ii) mostly flat patterns were discovered because of i). It was found that data cleansed using the 30 heart beats/ minute difference were mainly insightful achieving the best balance between over-filtering and necessary noise removal. These results that will be presented below. Results arising from data cleansed using the 20 heart beats/ minute difference will also be indicatively presented only for HA. It is also noted that the X axis in the figures below show 10 s instead of 14 since the moving average of a 5-seconds time windows was applied to the 14 s pulses around harsh events that were investigated. Therefore, seconds from 7 to 10 include the second at which the event occurred.

Fig. 5 illustrates the smoothened heart rate patterns around HA events where the 30 heart beats/ minute difference threshold was used for cleansing. The comparison between the results arising when 20 and 30 heart beats/ minute difference threshold are used, showed that results are more insightful when the threshold used is 30, since a significantly larger sample is used and heart rate differences between consecutive seconds are also larger. In this case, the elbow method indicated 5 different heart rate patterns as optimum. Compared to the results of the previous run, the results of this run indicate only one pattern with increasing heart rate before the events (pattern 1) and one pattern with a slight increasing heart rate before the event (pattern 2). Both these patterns have a



Pattern ID	Shape	Initial heart rate
1	Non-convex	Medium
2	Flat-shape	Medium
3	Non-convex	High
4	Flat-shape	Low
5	Non-convex	High

Fig. 5. Results of the 5 smoothened heart beat patterns identified during harsh acceleration events, using 30 heart beats/minute difference threshold (Y axis: Heart rate/ minute, X axis: time, harsh event occurrence is included in all seconds between 7th to 10th of the X-series).

medium initial heart rate. Two new patterns compared to the previous run were discovered here, one with a low and flat heart rate (pattern 4) and one that is eventually decreased while initially increased (pattern 5). Finally, non-convex shaped pattern 3 is similar to pattern 3 of the previous results discussed that concentrates the outliers of the sample.

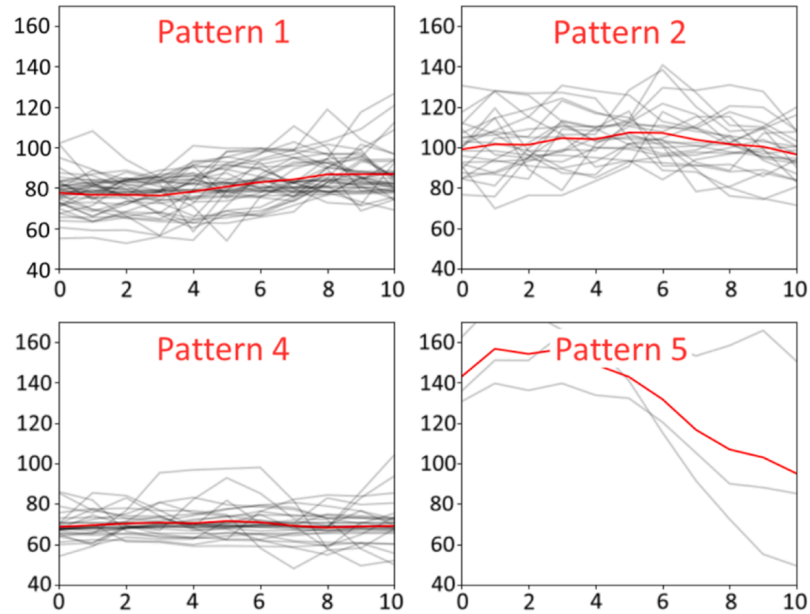
Regarding smoothened heart rate patterns around HB events, the elbow method indicated 5 different patterns. Fig. 6 illustrates the heart rate patterns around HB events where the 30 heart beats/ minute difference threshold was used for cleansing. Two flat-shaped patterns, pattern 4 and pattern 1 are observed, having a low and medium initial heart rate. Patterns 2 and 3 present a decreasing trend around the event with pattern 2 showing also an initial increasing trend. Again, pattern U-shaped 5 is concentrating those patterns that could be considered outliers (the average heart rate is higher than the rest of the patterns) or are not fitting well in any other pattern.

Regarding the smoothened heart rate patterns around HC events, a higher number of patterns is noticed most probably because a larger sample was available for the analysis. Fig. 7 illustrates the heart rate patterns around HC events where the 30 heart beats/ minute difference threshold was used for cleansing. Three out of the five flat-shaped patterns observed (patterns 2, 6 and 8) differ in terms of the average heart rate and the slight decreasing or increasing trends shown around the event. Patterns 5 and 7 show a decreasing trend of different rate, with pattern 7 showing an overall decrease of at least 30 heart beats/ minute. Patterns 1 and 4 on the other hand are non-convex shaped and illustrate an increasing trend of approximately 20 heart beats/ minute. Their difference is that heart rate in pattern 4 is gradually decreasing right before the event occurs, which may be an indication of the driver becoming calm when realizing that a potential incident is avoided. Finally, pattern 3 has a high initial heart rate and concentrates the outliers of the sample.

6. Discussion

6.1. Driving behavior interpretation

The analysis of the results above revealed that there are distinct speed patterns before and after HA, HB and HC events of high intensity within a driving pulse. In other words, there is a systematic speed behavior that can be grouped into 4 to 6 groups of behaviors depending on the event. Regarding HC events, it was found that all patterns are characterized by a mild to intense acceleration after the event. A deceleration is also observed in most speed patterns shortly before HA events. This can be attributed to the fact that HB event may be occur more 'urgently', due to a rapid change in the environment, a lead-vehicle or pedestrian unexpected movement or other



Pattern ID	Shape	Initial heart rate
1	Flat-shape	Low
2	U-shape	Medium
3	Non-convex shape	High
4	Flat-shape	Medium
5	U-shape	High

Fig. 6. Results of the 5 smoothened heart beat patterns identified during harsh braking events, using 30 heart beats/minute difference threshold (Y axis: Heart rate/ minute, X axis: time, harsh event occurrence is included in all seconds between 7th to 10th of the X-series).

type of conflict. On the other hand, HA and HC may be associated with more ‘conscious’ decision making, in which the driver has a couple of seconds to decelerate before making the swift decision that leads to the harsh maneuver. Alternatively, it can be considered that HA and HC events may be associated with HB, occurring shortly after them.

This behavior should be further analyzed to understand the cases where this deceleration could have been avoided, as these may indicate driver recklessness or aggressiveness. Some patterns are represented by a higher percentage of observations than others. In order to draw safety-related conclusions from this analysis, these patterns should be correlated with crash risk in the future.

Regarding heart rate patterns, a variety of profiles, including upward, downward and convex trends were identified before the examined events, depending of course on the type of the event One significant difference found was that the number of patterns for HC events is much higher than those for HA and HB events.

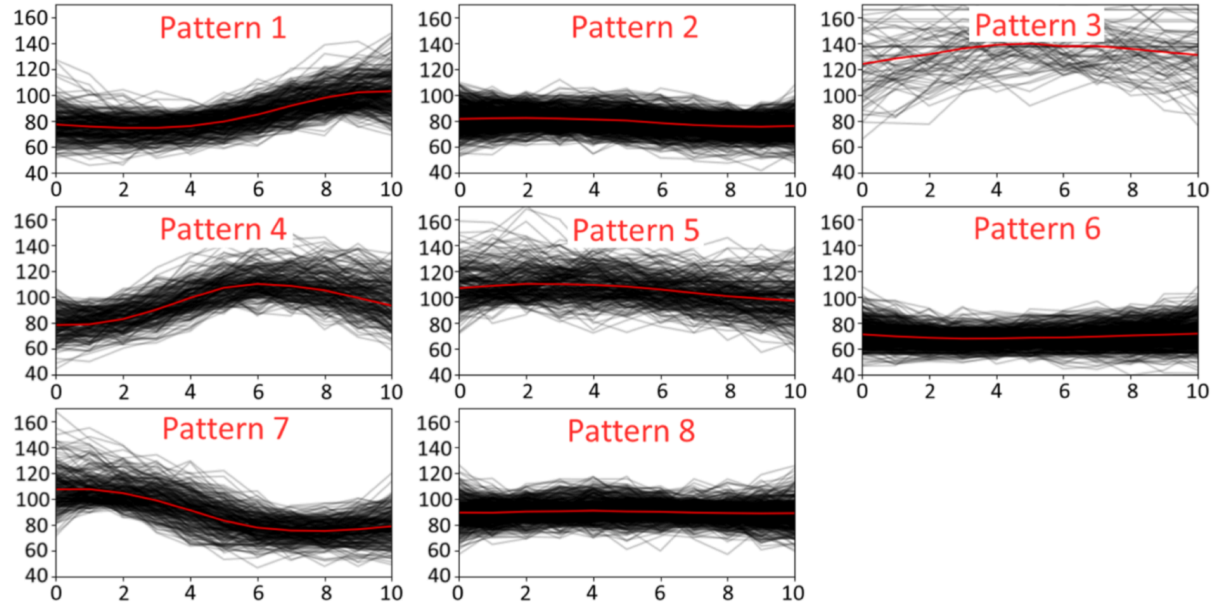
More specifically, the upward trend patterns may correspond to situations at which the driver becomes more stressed shortly before the event, while the downward trend may correspond to situation at which the stress and alertness level of the driver were raised before, or at the beginning of the examined driving pulse, and the harsh event decision taken resulted in gradual decrease of the heart rate until the event moment. Some of the patterns discovered were found to have an initial increase followed by a decrease during the examined pulse, which probably indicates that the driver is becoming calmer by realizing that a potential incident is avoided.

Nevertheless, it should be highlighted that such physiological indicators are known to be associated more with driver stress or sleepiness (Mårtensson, et al., 2018). The association of heart rate with specific driving outcomes such as harsh events may be more indirect and possibly weaker, compared to speed which is well known to have clear association with harsh events.

6.2. Drivers’ pattern consistency

In order to understand whether there is a systematic repetition of these patterns on a driver level, the concept of pattern consistency is introduced at this point. A driver can be considered consistent in terms of driving behavior when performing the same pattern for the majority of his/her events of a certain type (e.g. acceleration, braking or cornering). Consistency could be defined as the share of drivers in the sample exhibiting the same pattern by 50 %, 75 % or 90 % of their total events performed.

Table 2 provides the share of drivers having a consistent speed pattern in the 3 types of harsh events as per the above definition. It appears that depending on the type of harsh event, 45 to 68 % drivers are consistently following the same speed pattern for more than 50 % of the times. These shares are significantly dropping when the consistency threshold is increased to 75 % and especially for HA



Pattern ID	Shape	Initial heart rate
1	Non-convex	Low
2	Flat-shape	Medium
3	Flat-shape	High
4	Non-convex	Low
5	Flat-shape	High
6	Flat-shape	Low
7	Non-convex	High
8	Flat-shape	Medium

Fig. 7. Results of the 8 smoothened heart beat patterns identified during harsh cornering events, using 30 heart beats/minute difference threshold (Y axis: Heart rate/ minute, X axis: time, harsh event occurrence is included in all seconds between 7th to 10th of the X-series).

Table 2
Share of drivers showing consistency in speed patterns per type of event.

	Share of drivers following a consistent pattern for more than		
Harsh event type	50 % of events	75 % of events	90 % of events
Harsh acceleration	57 %	22 %	13 %
Harsh braking	68 %	31 %	29 %
Harsh cornering	45 %	12 %	8 %

and HC events. On the other hand, the reduction of the share of consistent drivers is not significantly reduced when moving from a 75 to a 90 % threshold and especially for HB events. This, together with the fact that the share of drivers having a consistent pattern in more than 50 % of cases is the highest for HB events, indicates that HB events is the type of event that is the most consistently followed by individual drivers.

It is also important to further discover which are those specific patterns that drivers are mainly consistent to. This is shown in Table 3 that presents the distribution of drivers with consistent pattern (using the threshold of 75 % of total events mentioned above) across different harsh event patterns. For instance, the first row of this table is interpreted as that 79 % of drivers with consistent speed pattern before HA events are in fact following the HA pattern #2 (see Fig. 3). This appears to be the strongest preference for consistent drivers among all types of harsh events. HA patterns 1, 3 and 5 are not followed by any consistent driver, which indicates that they may be more dependent on external conditions rather than driver characteristics. Regarding HB events, all patterns are followed to some extent by consistent drivers, without noticing any strong tendency towards any of the patterns. Pattern 3 on the other hand is the only one not followed by consistent drivers performing HC events. The strongest trend for cornering events is to follow patterns 2 and 5 (See Fig. 4).

A similar analysis was performed for the heart rate patterns with the scope to understand whether there is a systematic repetition of

Table 3
Distribution of drivers with consistent pattern behavior for more than 75% of events occurred across different harsh event patterns.

Harsh event type	Pattern ID	% of consistent drivers
	2	79 %
	4	16 %
	6	5 %
Harsh braking	1	32 %
	2	23 %
	3	32 %
	4	13 %
Harsh cornering	1	10 %
	2	30 %
	4	10 %
	5	40 %
	6	10 %

these patterns on a driver level and whether some of the patterns are consistently performed by certain drivers. Having excluded those drivers participating in only one event, it was found that the rest do not present consistency in their heart rate patterns. It should be noted though that the sample of drivers with heart rate data was 36, significantly less than those participating in the experiment in total.

As identified in the literature review conducted, this is the first study to focus on the driving pulse level to identify patterns of driving behaviour around harsh events. Most studies in the past focused on driving style recognition using already classified events or maneuvers with the purpose of identifying the driver, the type of maneuvers or their intensity (Martinelli et al., 2021, Sarker et al., 2021, Schwarz 2017). The difference between this and these past studies is that our study uses an unsupervised learning approach, and not a supervised learning approach that assumes prior knowledge on what each pattern represents and in which cluster it belongs to. The advantage of our approach is that it can be applied in any new driver sample even when we have little or no information on the sample beforehand.

7. Conclusions

This paper investigated the microscopic characteristics of driving behavior around driving pulses of 3 types of harsh events of high intensity, namely braking, acceleration and cornering. This is important because it can serve as a useful basis for real time detection and prediction of events at a fine time resolution, possibly based on knowledge of only a small number of key metrics. To this end, a time-series clustering approach was proposed for driving pattern recognition using naturalistic driving data of speed and heart rate profiles of the driver. This study addressed three research questions in the way described below.

1) Are there common driving patterns shortly before and after events, e.g. At the driving pulse level?

Results show that there are 4, 6 and 6 repetitive speed patterns recognized during HB, HA and HC events respectively. The investigation of patterns during harsh braking revealed a steady speed before events, whereas before HA and HC events, a mild to intense deceleration is noticed in almost all patterns. The distribution of observations across patterns of harsh events is not uniform, with some patterns being significantly under-represented. Moreover, some patterns show higher variability of behavior within the cluster than others. It is suggested for future research to focus on better understanding the relationship between such patterns and risky behaviors, and eventually incidents and accidents.

Heart rate patterns around harsh events also exist and present upward and downward trends before these events, depending of course on the type of the event. It would be valuable to further investigate this question, in order to draw stronger conclusions. For instance, it would be worth examining a larger sample or a different length of time-series (pulse). Moreover, the variability that is noticed in the raw IBI data leads to the conclusion that the 5-seconds moving-average heart rate is a good method to use for the data smoothening. Finally, it would be interesting to jointly examine the speed and heart rate patterns around events, and eventually contribute to understanding the correlation between physiological and traffic behavior indicators.

2) Are there indications that some driving patterns could potentially be related to safety critical behaviors in terms of human factors e.g. aggressiveness, stress and cautiousness?

This study discovered the existence of three speed patterns around HA events that have significantly higher deceleration before the event compared to the rest of the patterns. Moreover, two speed patterns before HC events also presented a similar harsh deceleration before cornering. The deceleration observed before these harsh events is a possible indication of an aggressive or reckless behavior since in some cases speed is approximately the same before and after the HB event. In other cases, drivers may have probably been able to avoid the harsh event if a lower average speed was maintained before the event. Finally, some HC events could be avoided if more time was dedicated to brake before cornering.

3) Are these driving patterns also 'driver' patterns, i.e. Are they consistent within drivers, or anybody could exhibit a certain pattern under different circumstances?

A new concept of driving pattern consistency was defined in this research, as 'the extent of repetition of the same pattern in the majority of the harsh events performed by the same individual' – with 3 different thresholds of 50, 75 or 90 % of the total events performed with the same pattern used as consistency thresholds. It was found that HB events are the type of event in which drivers are most consistent to their driving patterns. Consistent drivers were evenly distributed across patterns of HB, whereas in HA and HC events, a stronger tendency to follow a certain specific pattern was found. Consistency can be assumed to provide a good basis for forecasting events and the individual driver level, and therefore provide more timely and personalized warnings to 'consistent' drivers in short time close to the event. It is recommended for future research to also investigate driving pattern consistency based on driver age, gender and road type; this data was not available in the dataset at the time that this research was conducted.

On the other hand, a large share of drivers were found to be inconsistent, i.e. a certain individual might exhibit different HB, HA or HC patterns. It can be concluded that driving patterns, although clearly identifiable, are to a large extent driver-agnostic. This suggest that external factors, e.g. traffic conditions, road and environment features, may be more important determinants of the harsh event pattern than underlying driver personality traits. This creates a challenge, as the continuous monitoring and 'interpretation' of the environment is not feasible in current low-automation vehicles; therefore, new solutions are needed to provide not only individual-specific interventions, but also context-specific interventions.

In terms of the contributions and innovations of this research, it develops a methodological framework for microscopic driving pattern recognition that does not currently exist in literature. It also discovers and analyzes the existing driving patterns shortly before and after harsh acceleration, braking and cornering events with regards to speed and driver's heart rate. Finally, it provides insights on the existence of consistent driving patterns for some drivers as well as on which are these patterns mostly followed by drivers.

Our study has some limitations; because of the data collection sensors used in the experiment, the driving sample used for heart rate pattern recognition was relatively small compared to the sample used for speed pattern recognition. Likewise, the sample collected around harsh braking events was relatively smaller compared to the driving samples collected around harsh acceleration and braking events. Moreover, crash data and crash history data were not available and therefore, the direct or indirect relationship between patterns and crash probability could not be detected. Finally, due to the lack of data related to the rest of the driving environment, this study could not dive deeper to understand whether some microscopic patterns can be attributed to aggressiveness or recklessness or reflect random variation in driver behavior.

An area of future research would be to also consider medium and low intensity harsh events and understand if the patterns identified here are differentiated during events of lower intensity. As for the length of the time-series pulse, its precise identification is also important and should be further explored. As mentioned above, future research should also investigate whether some patterns are related to aggressiveness or recklessness. It was also aforementioned that to draw safety-related conclusions from this analysis, patterns detected should be correlated with crash risk in the future. Multivariate clustering is a recommended next step since it could combine the parallel evolution of several metrics such as speed, heart beats and others that could be available in the future e.g. headways, eye-tracking behavior. Finally, prediction models should be developed to see how predictable these behavioral patterns can be based on other recorded metrics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101027222.

The authors would like to thank the iDreams consortium, for providing all necessary data to accomplish this study, and the Technical University of Munich in particular for their valuable support in the data processing. The iDreams research project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 814761.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: D. Tselentis, E. Papadimitriou; data collection: D. Tselentis; analysis and interpretation of results: D. Tselentis, E. Papadimitriou; draft manuscript preparation: D. Tselentis; manuscript review & editing: D.Tselentis, E.Papadimitriou. All authors reviewed the results and approved the final version of the manuscript.

References

"i-Dreams: Safety Tolerance zone calculation and interventions for diver – vehicle – environment interactions under challenging conditions" of the Horizon 2020 framework programme on transport research of the European Commission (2019-2022).

Adrian, J., Postal, V., Moessinger, M., Rascle, N., & Charles, A. (2011). Personality traits and executive functions related to on-road driving performance among older drivers. *Accident Analysis and Prevention, 43*(5), 1652–1659.

Afghari A.P., Papadimitriou E., Pilkington-Cheney F., Filtness A., Brijs T., Brijs K., Cuenen A., De Vos B., Dirix H., Ross V., Wets G., Lourenço A., Rodrigues L. (2022). Investigating the effects of sleepiness in truck drivers on their headway: An instrumental variable model with grouped random parameters and heterogeneity in their means. Analytic Methods in Accident Research, Volume 36, December 2022, 100241.

Baca Mendoza, K., & Söderkvist Vermelin, W. (2019). Discovering Patterns in Driving Data.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems, 7*(1), 63–77.

Bonsall, P., Liu, R., & Young, W. (2005). Modelling safety-related driving behaviour—impact of parameter values. *Transportation Research Part A: Policy and Practice, 39*(5), 425–444. https://doi.org/10.1016/j.tra.2005.02.002

Chen, C.-F. (2009). Personality, safety attitudes and risky driving behaviors-Evidence from young Taiwanese motorcyclists. *Accident Analysis and Prevention, 41*(5), 963–968.

Chen, T., Shi, X., & Wong, Y. D. (2021). A lane-changing risk profile analysis method based on time-series clustering. *Physica A: Statistical Mechanics and its Applications, 565*, Article 125567.

Chetouane, N., Klampfl, L., & Wotawa, F. (2021). Extracting information from driving data using k-means clustering. In Proceedings of the 33rd International Conference on Software Engineering & Knowledge Engineering (SEKE), KSIR Virtual Conference Center, Pittsburgh, USA.

Cura, A., Küçük, H., Ergen, E., & Öksüzoğlu, İ. B. (2020). Driver profiling using long short term memory (LSTM) and convolutional neural network (CNN) methods. *IEEE Transactions on Intelligent Transportation Systems*.

de Zepeda, M. V. N., Meng, F., Su, J., Zeng, X. J., & Wang, Q. (2021). Dynamic clustering analysis for driving styles identification. *Engineering applications of artificial intelligence, 97*, Article 104096.

Eboli, L., Guido, G., Mazzulla, G., Pungillo, G., & Pungillo, R. (2017). Investigating car users' driving behaviour through speed analysis. *Promet - Traffic, 29*(2), 193–202.

Eboli, L., Mazzulla, G., & Pungillo, G. (2017). The influence of physical and emotional factors on driving style of car drivers: A survey design. *Travel Behaviour and Society, 7*, 43–45.

Eboli, L., Mazzulla, G., & Pungillo, G. (2016). Combining speed and acceleration to define car users' safe or unsafe driving behaviour. *Transportation research part C: emerging technologies, 68*, 113–125.

Ellison, A. B., Greaves, S. P., & Bliemer, M. C. (2015). Driver behaviour profiles for road safety analysis. *Accident Analysis & Prevention, 76*, 118–132.

Feng, Y., Pickering, S., Chappell, E., Iravani, P., & Brace, C. (2018). In *September). Driving style analysis by classifying real-world data with support vector clustering* (pp. 264–268). IEEE.

Feng, Z., Yang, M., Zhang, W., Du, Y., & Bai, H. (2018). Effect of longitudinal slope of urban underpass tunnels on drivers' heart rate and speed: A study based on a real vehicle experiment. *Tunnelling and underground space technology, 81*, 525–533.

Fotouhi, A., & Montazeri-Gh, M. J. S. I. (2013). Tehran driving cycle development using the k-means clustering method. *Scientia Iranica, 20*(2), 286–293.

Garcia-Constantino, M., Missier, P., Blythe, P., & Guo, A. W. (2014). In *October). Measuring the impact of cognitive distractions on driving performance using time series analysis* (pp. 2047–2053). IEEE.

Goffinet, E., Lebbah, M., Azzag, H., & Giraldi, L. (2020). Autonomous Driving Validation with Model-Based Dictionary Clustering. *In ECML/PKDD, 4*, 323–338.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), 100-108.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems, 6*(2), 156–166.

Johnson, Derick A., and Mohan M. Trivedi. Driving style recognition using a smartphone as a sensor platform. In Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on, 2011 pp. 1609-1615. https://doi.org/10.1109/ITSC.2011.6083078.

Karlaftis, M. G., Gleason, J. M., & Barnum, D. T. (2013). Bibliography of Urban Transit Data Envelopment Analysis (DEA). *Publications. Available at SSRN*. http://ssrn.com/abstract=1350583,. https://doi.org/10.2139/ssrn.1350583

Kim, K., & Yamashita, E. Y. (2007). Using a k-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu. *Hawaii. Journal of advanced transportation, 41*(1), 69–89.

Lee, M. S., Park, J., Jang, Y. J., Kim, W., & Yoon, D. (2018). In *October). Individual Stable Driving Pattern Analysis for Evaluating Driver Readiness at Autonomous Driving Levels 2 and 3* (pp. 315–319). IEEE.

Liaw, B. Y., Bethune, K. P., & Kim, C. S. (2002). Time-series field trip data analysis using adaptive recognition approach. Analysis on driving patterns and vehicle usage for electric vehicles. In *In Proceedings of the 19th Electric Vehicle Symposium (EVS-19)* (pp. 19–23).

Mahachandra, M., Sutalaksana, I. Z., & Suryadi, K. (2012). In *July). Sensitivity of heart rate variability as indicator of driver sleepiness* (pp. 1–6). IEEE.

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identifying driving safety profiles from smartphone data using unsupervised learning. *Safety Science, 119*, 84–90.

Mårtensson, H., Keelan, O., & Ahlström, C. (2018). Driver sleepiness classification based on physiological data and driving performance from real road driving. *IEEE Transactions on Intelligent Transportation Systems, 20*(2), 421–430.

Martinelli, F., Marulli, F., Mercaldo, F., & Santone, A. (2021). Neural networks for driver behavior analysis. *Electronics, 10*(3), 342.

Mehler, B., Reimer, B., & Wang, Y. (2011, June). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In Driving Assesment Conference (Vol. 6, No. 2011). University of Iowa.

Mehler, B., Reimer, B., Pohlmeyer, A. E., & Coughlin, J. F. (2008). *The association between heart rate reactivity and driving performance under dual task demands in late middle age drivers* (p. 5370). An International Journal, Special Issue: Advances in Transportation Studies.

Meseguer, J. E., Calafate, C. T., & Cano, J. C. (2018). On the correlation between heart rate and driving style in real driving scenarios. *Mobile Networks and Applications, 23*, 128–135.

Mesken, J., Lajunen, T., & Summala, H. (2002). Interpersonal violations, speeding violations and their relation to accident involvement in Finland. Ergonomics, 45(7), 469-483. Nilsson G. The effects of speed limits on traffic accidents in Sweden. Sartryck, Swedish National Road and Transport Research Institute, 1982.

Obuhuma, J. I., Okoyo, H. O., & McOyowo, S. O. (2018). Driver Behaviour Profiling Using Dynamic Bayesian Network. *International Journal of Modern Education & Computer Science, 10*(7).

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.

Persson, A., Jonasson, H., Fredriksson, I., Wiklund, U., & Ahlström, C. (2020). Heart rate variability for classification of alert versus sleep deprived drivers in real road driving conditions. *IEEE Transactions on Intelligent Transportation Systems, 22*(6), 3316–3325.

Reimer, B., Mehler, B., Coughlin, J. F., Roy, N., & Dusek, J. A. (2011). The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. *Transportation research part F: traffic psychology and behaviour, 14*(1), 13–25.

Saiprasert, C., Pholprasit, T., & Thajchayapong, S. (2017). Detection of driving events using sensory data on smartphone. *International journal of intelligent transportation systems research, 15*(1), 17–28.

Sarker, S., Haque, M. M., & Dewan, M. A. A. (2021). *Driving Maneuver Classification Using Domain Specific Knowledge and Transfer Learning*. IEEE Access.

SATO, S., Taoda, K., Kawamura, M., WAKARA, K., Fukuchi, Y., & Nishiyama, K. (2001). Heart rate variability during long truck driving work. Journal of human ergology, 30(1-2), 235-240.

Schwarz, Chris. Time Series Categorization of Driving Maneuvers Using Acceleration Signals. In: Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 26-29, 2017, Manchester Village, Vermont. Iowa City, IA: Public Policy Center, University of Iowa, 2017: 270-276. https://doi.org/10.17077/drivingassessment.1646.

Shakouri, M., Ikuma, L. H., Aghazadeh, F., & Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. *International journal of industrial ergonomics, 66*, 136–145.

Shi, B., Xu, L., Hu, J., Tang, Y., Jiang, H., Meng, W., & Liu, H. (2015). Evaluating driving styles by normalizing driving behavior based on personalized driver modeling. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45*(12), 1502–1508.

Spyrou, E., Vernikos, I., Savelonas, M., & Karkanis, S. (2020, June). An Image-Based Approach for Classification of Driving Behaviour Using CNNs. In Conference on Sustainable Urban Mobility (pp. 263-271). Springer, Cham.

Takano, W., Matsushita, A., Iwao, K., & Nakamura, Y. (2008). Recognition of human driving behaviors based on stochastic symbolization of time series signal. In *In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 167–172). IEEE.

Tozman, T., Magdas, E. S., MacDougall, H. G., & Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. *Computers in Human Behavior, 52*, 408–418.

Tselentis, D., & Papadimitriou, E. (2023). Driver Profile and Driving Pattern Recognition for Road Safety Assessment: Main Challenges and Future Directions. Article in press, IEEE Open. *Journal of Intelligent Transportation Systems.*

Tselentis, D., Vlahogianni, E., & Yannis, G. (2019). Driving safety efficiency benchmarking using smartphone data. *Transportation research part C, 109*, 343–357.

Tselentis, D. I., Vlahogianni, E. I., & Yannis, G. (2021). Temporal analysis of driving efficiency using smartphone data. *Accident Analysis & Prevention, 154*, Article 106081.

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. https://doi.org/10.1016/j.aap.2016.10.006

Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. *Safety Science, 41*(5), 427–443.

Vlahogianni, E. I., & Barmpounakis, E. N. (2017). Driving analytics using smartphones: Algorithms, comparisons and challenges. *Transportation Research Part C: Emerging Technologies, 79*, 196–206.

Vogel, M., Hamon, R., Lozenguez, G., Merchez, L., Abry, P., Barnier, J., … Robardet, C. (2014). From bicycle sharing system movements to users: A typology of Vélo'v cyclists in Lyon based on large-scale behavioural dataset. *Journal of Transport Geography, 41*, 280–291.

Wang, J., Zhu, S., & Gong, Y. (2010). Driving safety monitoring using semisupervised learning on time series data. *IEEE Transactions on Intelligent Transportation Systems, 11*(3), 728–737.

Warren, J., Lipkowitz, J., & Sokolov, V. (2019). Clusters of driving behavior from observational smartphone data. *IEEE Intelligent Transportation Systems Magazine, 11* (3), 171–180.

Weidner, W., Transchel, W., & Weidner, R. (2017). Telematic driving profile classification in car insurance pricing. *Annals of actuarial, science,11(2)*, 213–236.